Research for Policy

Haroon Sheikh
Corien Prins
Erik Schrijvers

Mission AI

The New System Technology

WRR
THE NETHERLANDS SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY

OPEN ACCESS

Springer

Research for Policy

Studies by the Netherlands Council for Government Policy

The Netherlands Scientific Council for Government Policy (WRR) is an independent strategic advisory body for government policy in the Netherlands. It advises the Dutch government and Parliament on long-term strategic issues that are of great importance for society. The WRR provides science-based advice aimed at: opening up new perspectives and directions, changing problem definitions, setting new policy goals, investigating new resources for problem-solving and enriching the public debate.

The studies of the WRR do not focus on one particular policy area, but on cross-cutting issues that affect future policymaking in multiple domains. A long-term perspective complements the day-to-day policymaking, which often concentrates on issues that dominate today's policy agenda.

The WRR consists of a Council and an academic staff who work closely together in multidisciplinary project teams. Council members are appointed by the King, and hold academic chairs at universities, currently in fields as diverse as economics, sociology, law, public administration and governance, health and water management. The WRR determines its own work programme, as well as the content of its publications. All its work is externally reviewed before publication.

The Research for Policy Series

In this series, we publish internationally relevant studies of the Netherlands Scientific Council for Government Policy. Many of the cross-cutting issues that affect Dutch policymaking, also challenge other Western countries or international bodies. By publishing these studies in this international open access scientific series, we hope that our analyses and insights can contribute to the policy debate in other countries.

About the Editors

Corien Prins is Chair of the WRR and Professor of Law and Information Technology at Tilburg Law School (Tilburg University).

Frans Brom is Council secretary and director of the WRR office. He also is Professor of Normativity of Scientific Policy Advice at the Ethics Institute of Utrecht University.

Haroon Sheikh • Corien Prins • Erik Schrijvers

Mission AI

The New System Technology

Springer

Haroon Sheikh
Wetenschappelijke Raad voor het Regeringsbeleid
Den Haag, The Netherlands

Corien Prins
Wetenschappelijke Raad voor het Regeringsbeleid
Den Haag, The Netherlands

Erik Schrijvers
Wetenschappelijke Raad voor het Regeringsbeleid
Den Haag, The Netherlands

ISSN 2662-3684 ISSN 2662-3692 (electronic)
Research for Policy
ISBN 978-3-031-21447-9 ISBN 978-3-031-21448-6 (eBook)
https://doi.org/10.1007/978-3-031-21448-6

This work was supported by The Netherlands Scientific Council for Government Policy, The Hague

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Persons Consulted

Specified organizations at the time of the interview
E. (Emile) Aarts, Tilburg University, Dutch AI coalition
R. (Ron) Augustus, SURF
R. (Robert) Baris, Dutch Tax and Customs Administration
(Arie) van Bellen, ECP | Platform for the Information Society
S. (Salima) Benhamou, France Strategy
E. (Ellen) Berends, Dutch Safety Board
S. (Siri) Berends, Set Up
F. J. (Floris) Bex, Utrecht University
P. (Patrick) Blankers, Ericsson Netherlands
F. (Floris) den Boer, PIANOo
V. (Vincent) Böhre, Privacy First
(Christien) Bok, SURF
R. (Romain) Bonenfant, Ministère de l'Économie, des Finances et de la Relance, France
H. (Hans) Bos, Microsoft
J. A. (Jan) van den Bos, Netherlands Inspection Council
M. (Mirèl) ter Braak, Netherlands Authority for the Financial Markets
(Barteld) Braaksma, Statistics Netherlands
(Alain) Bravo, L'Académie des Technologies, France
J. (Joël) Buiter, Health and Youth Care Inspectorate
M. (Mirjam) van Burgel, Netherlands Radiocommunications Agency
J. (Joost) van der Burgt, De Nederlandsche Bank
F. (Frits) Bussemaker, National I-Partnership – Higher Education (I-Partnerschap Rijk – Hoger Onderwijs), Institute for Accountability in the Digital Age
P. (Pierluigi) Casale, TomTom
M. (Madeleine) de Cock Buning, Utrecht University
S. (Stephanie) Combes, Ministère de la Santé, France
K. (Kate) Crawford, AI Now Institute
R. (Roxane) Daniëls, Association of Netherlands Municipalities
P. (Petra) Delsing, Ministry of Infrastructure and Water Management

S. (Stijn) van Deursen, Utrecht University
M. (Marloes) Dignum, Ministry of Infrastructure and Water Management
M. V. (Virginia) Dignum, Umeå Universitet
J. F. T. M. (José) van Dijck, Utrecht University
K. H. D. M. (Klaas) Dijkhoff, VVD
R. I. J. (Roel) Dobbe, AI Now Institute, TU Delft
P. (Pedro) Domingos, University of Washington
L. (Louis) Dubertet, National Academy of Technologies of France
N. (Nathan) Ducastel, Association of Netherlands Municipalities
M. (Myrte) Dujardin, Ministry of Social Affairs and Employment
M. A. (Marlies) van Eck, Leiden University
(Aik) van Eemeren, Municipality of Amsterdam
Q. (Quirine) Eijkman, Netherlands Institute for Human Rights
Q. C. (Rinie) van Est, Rathenau Institute
J. (Bart Jan) van Ettekoven, Council of State
T. (Thomas) Faber, Ministry of Economic Affairs
G. (Gerard) Feitsma, Radiocommunications Agency Netherlands
(Bas) Filippini, Privacy First
G. J. (Gert-Jan) Fonk, Ministry of Agriculture, Nature and Food Quality
P. H. A. (Paul) Frissen, Tilburg University
J. D. C. (Jacobine) Geel, Netherlands Institute for Human Rights
Y. (Yannick) van Gelder, Wageningen University & Research
S. (Sennay) Ghebreab, University of Amsterdam
C. (Corine) van Ginkel, Research and Documentation Centre
P. (Peter) Goeijers, Ericsson Netherlands
P. (Peter) Gouw, Ministry of Health, Welfare and Sport
J. (Jochem) de Groot, Microsoft
M. (Marion) Gust, Ministère de la Transition écologique et solidaire, France
T. (Tom) van de Haar, Ministry of Social Affairs and Employment
H. J. (Hugo) van Haastert, Ministry of Health, Welfare and Sport
P. (Peter) Hagendoorn, The Fluid Society
G. (Gry) Hasselbalch, DataEthics
M. (Martin) Heijnsbroek, MIcompany
R. (Rik) Helwegen, University of Amsterdam
S. (Steven) Hillebrink, Ministry of the Interior and Kingdom Relations
J. (Justin) Hoegen Dijkhof, Netherlands Institute for Human Rights
R. (Ronald) van den Hoogen, Dutch National Academy for Digitalization and Computerization of Government
(Floris) Hoogenboom, Schiphol
H. H. (Holger) Hoos, Leiden University
(Gerald) Hopster, Dutch Data Protection Authority
N. (Noor) Huiboom, Ministry of the Interior and Kingdom Relations
(Bas) van Hulst, DeepVision
J. P. (Joost) van Iersel, European Policy Centre
(Ivana) Isgum, University of Amsterdam

M. (Menno) Israël, Netherlands Authority for Consumers and Markets
(Hans) van de Jagt, Ministry of Defence
(Caspar) de Jonge, Ministry of Infrastructure and Water Management
(Erik) Jonker, Netherlands Court of Audit
M. (Catholijn) Jonker, TU Delft
R. (Rina) Joosten-Rabou, Seedlink
M. (Merel) Kampers, Ministry of Social Affairs and Employment
C. (Chandro) Kandiah, Ministry of Agriculture, Nature and Food Quality
(Daan) Keijser, Customs Administration
M. C. G. (Mona) Keijzer, Ministry of Economic Affairs and Climate Policy
K. (Kees) van der Klauw, Netherlands AI coalition
M. H. (Meine Henk) Klijnsma, Ministry of the Interior and Kingdom Relations
R. (Rogier) Klimbie, Considerati
V. (Victor) Klos, Dutch Data Protection Authority
L. (Linda) Kool, Rathenau Institute
M. (Mauritz) Kop, AI law, Stanford University
(Antoine) de Kort, National Vehicle and Driving Licence Registration Authority
W. (Willem) Korteweg, Leading Edge Forum
K. (Katja) van Kranenburg, CMS, Dutch AI Coalition
(Floris) Kreiken, Ministry of the Interior and Kingdom Relations
J. (Johan) Krijgsman, Health and Youth Care Inspectorate
W. (Wouter) Kroese, Pacmed
H. (Bert) Kroese, Statistics Netherlands
R. L. (Inald) Lagendijk, TU Delft
T. (Tom) Leenders, Ministry of Finance
M. H. G. (Michel) van Leeuwen, Ministry of Justice and Security
(Anja) Lelieveld, Ministry of the Interior and Kingdom Relations
O. (Olivia) Lin, Ministry of Foreign Affairs
(Frans) Lips, Ministry of Agriculture, Nature and Food Quality
(Gilles) de Margerie, France Stratégie
(Alexander) Melchior, Ministry of Agriculture, Nature and Food Quality
(Jeroen) van Mierlo, Ministry of Education, Culture and Science
(Inge) Molenaar, Radboud University
(Bennie) Mols, Science journalist
S. (Selwyn) Moons, PricewaterhouseCoopers
S. (Sander) Mul, Ministry of Justice and Security
Y. (Yvette) Mulder, NEN
C. J. (Catelijne) Muller, ALLAI
(Jasper) Nagtegaal, Radiocommunications Agency Netherlands
(Edwin) Nas, Ministry of Infrastructure and Water Management
R. (Rob) Nijman, IBM
(Carine) van Oosteren, Social and Economic Council of the Netherlands
(Cees) Oudshoorn, Confederation of Netherlands Industry and Employers
(Bertrand) Pailhes, CNIL
(Geert) Pater, National Vehicle and Driving Licence Registration Authority

(Miranda) Pirkovski, Netherlands Court of Audit
T. (Theo) van de Plas, National Police
(Addy) Polet, Ministry of Education, Culture and Science
(Marieke) van Putten, Ministry of the Interior and Kingdom Relations
(Ardaan) van Ravenzwaaij, Ministry of the Interior and Kingdom Relations
(Bram) de Rijk, Ministry of the Interior and Kingdom Relations
(Maarten) de Rijke, University of Amsterdam
(Dirk) van Roode, NL Digital
L. (Lennart) Salemink, Ministry of Infrastructure and Water Management
T. (Tim) Salimans, Google Brain
J. (Johan) Schot, Utrecht University
(Gijs) van Schouwenburg, Ministry of Infrastructure and Water Management
(Olof) Schuring, Ministry of Justice and Security
(Cecile) Schut, Dutch Data Protection Authority
(Arnold) Smeulders, University of Amsterdam
(Cees) Snoek, University of Amsterdam
J. (Just) Stam, Ministry of Justice and Security
M. (Mildo) van Staden, Ministry of the Interior and Kingdom Relations
M. R. (Maarten) van Steen, Dutch AI Coalition
M. (Michiel) Steltman, Digital Infrastructure Netherlands
(Bart) Stuut, Netherlands Authority for Consumers and Markets
J. T. (Jonathan) Taplin, University of Southern California
L. (Linnet) Taylor, Tilburg University
L. E. M. (Linda) Terlouw, DeepVision
M. Marthe) Tholen, Dutch Tax and Customs Administration
(Benjamin) Timmermans, IBM
(Bert) Timmermans, Ministry of Infrastructure and Water Management
(Amarens) Veeneman, Office of the House of Representatives
M. (Maarten) Veltman, Customs Administration
P. C. C. (Peter-Paul) Verbeek, University of Twente
(Peter) Vermeulen, Ministry of Social Affairs and Employment
D. (Corinne) Vigreux, TomTom, Codam
W. (Focco) Vijselaar, Ministry of Economic Affairs and Climate Policy
(Iris) Vissers, Ministry of Finance
(Jasper) van Vliet, Human Environment and Transport Inspectorate
(Bart) Voorn, Ahold Delhaize
M. (Michael) Vos, Microsoft
H. (Claes) de Vreese, University of Amsterdam
(René) Vroom, Radiocommunications Agency Netherlands
L. (Lauren) Waardenburg, VU Amsterdam
M. (Marieke) van Wallenburg, Ministry of the Interior and Kingdom Relations
(Hans) Wanders, Senior Civil Service
(Sandra) van der Weide, Ministry of Economic Affairs and Climate Policy
(Inge) Welbergen, Ministry of Education, Culture and Science
M. (Max) Welling, University of Amsterdam

(Klaas) Werkhorst, Ministry of Justice and Security
(Inge) Wertwijn, Dutch Tax and Customs Administration
R. (Ronald) Westerhof, Municipality of Apeldoorn
R. (Richard) Weurding, Dutch Association of Insurers
(Koen) Wienk, Netherlands Food and Consumer Product Safety Authority
(Joost) Witteman, SEO
(Aleid) Wolfsen, Dutch Data Protection Authority
M. E. (Sally) Wyatt, Maastricht University
R. (Rolf) Zeldenrust, PIANOo
(Berrie) Zielman, Netherlands Court of Audit
R. F. B. (Reinier) van Zutphen, The National Ombudsman
(Bart) Zwartjes, Netherlands Authority for the Financial Markets
R. (Richard) van Zwol, Council of State
(Guus) van Zwoll, Ministry of Foreign Affairs

Preface

This book is a translation and adaptation of the Dutch report *Opgave AI. De Nieuwe Systeemtechnologie*, which was presented to the Minister of Economic Affairs and Climate Policy in 2021.[1] In this study, the Netherlands Scientific Council for Government Policy (WRR) characterizes artificial intelligence as a 'system technology' that fundamentally alters society and establishes five overarching tasks for governments to embed AI into society.

This publication was written by Prof. Corien Prins (Chair and primary Council Member, LLM), Prof. Dr. Haroon Sheikh (Senior Research Fellow), and Dr. Erik Schrijvers (Senior Research Fellow), with the support of Drs. Eline de Jong (ex-staff member), Tessel van Oirsouw (intern, BA BSc), Prof. Dr. Mark Bovens (Council Member, LLM) and Monique Steijns (staff member, LLM).

Mission AI. The New System Technology is the product of the extensive study of academic literature, policy documents and analysis.

In addition, we have conducted interviews with over 170 external experts in the public and private sectors, both from the Netherlands and abroad. The interviews include conversations with municipalities, regulators, High Councils of State, scientists, company representatives, actors from civil society organizations and members of the Dutch AI Coalition. We are grateful for their contribution to this report. Their names are listed at the end of this report. During the final phase of the project, texts were reviewed by Prof. Dr. Luc Steels (Emeritus Professor of Artificial Intelligence, Vrije Universiteit Brussel), Marleen Stikker and Tom Demeyer (Director and CTO of Waag, respectively), Prof. Dr. Stavros Zouridis (Council Member of the Dutch Safety Board, LLM), Prof. Dr. José van Dijck (Professor of Media Studies, Utrecht University) and Prof. Dr. Koen Frenken (Professor of Innovation Studies, Utrecht University). We thank them for their comments and valuable suggestions.

Den Haag, The Netherlands

Haroon Sheikh
Corien Prins
Erik Schrijvers

[1] The original Dutch publication (2021) has been adapted for an international audience but has not been updated.

Contents

1 Introduction 1
- 1.1 AI at a Turning Point 1
- 1.2 AI Leaves the Lab and Enters Society 3
- 1.3 Technology and Public Values 5
- 1.4 A Historical Perspective 7
- 1.5 Overarching Tasks for the Societal Integration of AI 7
- 1.6 The Five Tasks 9
- 1.7 Structure of the Report 11

References 12

Part I Building Blocks: Introducing and Interpreting AI as a New System Technology, Similar to Electricity and the Internal Combustion Engine

2 Artificial Intelligence: Definition and Background 15
- 2.1 Definitions of AI 15
- 2.2 AI Prior to the Lab 20
- 2.3 AI in the Lab 28

References 39

3 AI Is Leaving the Lab and Entering Society 43
- 3.1 Momentum from Lab to Society 43
- 3.2 The Practical Application of AI 49
- 3.3 AI as a Phenomenon in Society 59
- 3.4 The Future of the Lab 71

References 79

4 AI as a System Technology 85
- 4.1 Classification of Technologies 86
- 4.2 The Societal Integration of System Technologies 96
- 4.3 Overarching Task 1: Demystification 101
- 4.4 Overarching Task 2: Contextualization 106

4.5 Overarching Task 3: Engagement . . . 111
4.6 Overarching Task 4: Regulation . . . 118
4.7 Overarching Task 5: Positioning . . . 125
References . . . 132

Part II Five Tasks: Discussion of the Tasks for Embedding AI Into Society

5 Demystification . . . 137
5.1 Behind the Myths About AI . . . 137
5.2 Contemporary Myths About AI . . . 143
5.3 In Conclusion . . . 174
References . . . 175

6 Contextualization . . . 179
6.1 The Technical Ecosystem . . . 182
6.2 The Social Ecosystem . . . 194
6.3 In Conclusion . . . 206
References . . . 207

7 Engagement . . . 211
7.1 Resistance . . . 216
7.2 Monitoring . . . 223
7.3 Co-operation . . . 232
7.4 In Conclusion . . . 237
References . . . 238

8 Regulation . . . 241
8.1 Government Standardization of AI . . . 243
8.2 AI Regulation and the Digital Living Environment . . . 261
8.3 In Conclusion . . . 279
References . . . 280

9 International Positioning . . . 287
9.1 AI and Competitive Advantages . . . 289
9.2 AI and National Security . . . 307
9.3 In Conclusion . . . 325
References . . . 326

Part III Agenda: Conclusions and Recommendations for AI Policy in the Netherlands

10 Policy for AI as a System Technology . . . 333
10.1 Five Tasks as Lessons from the Past . . . 334
10.2 Transition 1: From Fiction to Facts . . . 342
10.3 Transition 2: From Abstraction to Application . . . 347
10.4 Transition 3: From Monologue to Dialogue . . . 352

10.5 Transition 4: From Reaction to Action 357
10.6 Transition 5: From Nation to Network. 363
10.7 From Instruments to a Policy Infrastructure 367
10.8 In Conclusion – The Internal Combustion Engine of the Twenty-First Century 372
References 375

Appendix: Examples of AI Applications in the Netherlands 377

Glossary 379

Bibliography 385

About the Authors

Haroon Sheikh is a senior scientist and project coordinator at the Dutch Scientific Council for Government Policy and a Professor in The Strategic Governance of Global Technologies at VU University. His research focuses on the intersection of new technology and international relations. Haroon studied philosophy, political science and public administration at Leiden and Oxford University and did his PhD in philosophy on the relationship between tradition and modernity at VU University. He worked over a decade in the financial sector and headed Freedomlab Thinktank in Amsterdam. Haroon has written several books in the field of geopolitics.

Corien Prins Regulatory challenges and questions around the introduction of digital technologies have been on Corien Prins' agenda for three decades. In 1995, she founded what is now the Tilburg Institute for Law, Technology, and Society—a world-renowned research center on law, society and technology. In her research on the impact of emerging technologies on individuals and society, she has focused on the role of regulation, law, fundamental values and human rights. In 2017, Prins was appointed chair of the Netherlands Scientific Council for Government Policy (WRR), which advises the Dutch government and parliament on strategic issues likely to have significant political and societal consequences. Prins has an abiding interest in the interdependency of the digital and physical worlds. One of her current research foci within the broader theme of artificial intelligence is its use by the judiciary. Here she seeks to understand the implications of using AI for the core values of the judiciary—impartiality, independence and due process. Other research topics of Prins relate to the development of legal measures to gain a clearer picture of which parties, digital processes and services are key to the functioning of critical (vital) processes in society. She thereby looks at the chains and networks that support key processes and investigates whether digitalization necessitates changes to the prioritization of assets that are essential for the functioning of a society and economy. Her other research interests include the implications of digital technologies for democracy in a globalized world, the use of digital technologies and personal autonomy, and privacy and personal data protection.

Erik Schrijvers is senior scientist and project coordinator at the Dutch Scientific Council for Government Policy. He is also a member of the Supervisory Board of DEN, knowledge institute for culture and digital transformation. His research focuses on the intersection of digital technology, culture and policy. He studied history of international relations and philosophy at Utrecht University and obtained his PhD with a thesis on the history of forms of extra-parliamentary representation in the Netherlands. In 2016 and 2017, he worked for the Ministry of the Interior and Kingdom Relations, where he was responsible for the research and reporting of the study group Information Society and Government.

Abbreviations

AGI	Artificial General Intelligence
AI HLEG	High-Level Expert Group on AI
AIV	Advisory Council on International Affairs
ALLAI	Alliance for Artificial Intelligence Netherlands
ANNs	Artificial Neural Networks
GDPR	General Data Protection Regulation
CAPTCHA	Completely Automated Public Turing-tests to tell Computers and Humans Apart
CCW	Convention on Certain Conventional Weapons
CEN	Comité Européen de Normalisation
CENELEC	Comité Européen de Normalisation Electrotechnique
CPU	Central Processing Unit
DARPA	Defense Advanced Research Projects Agency
DIN	Deutsche Institut für Normung
DIU	Defense Innovation Unit
DKE	Deutsche Kommission Elektrotechnik Elektronik Informationstechnik
DL	Deep Learning
DMA	Digital Markets Act
DSA	Digital Services Act
GMO	Genetically Modified Organism
GOFAI	Good Old-Fashioned AI
GPT	General Purpose Technology
GPU	Graphic Processing Unit
HRM	Human Resources Management
ICAI	Innovation Center for Artificial Intelligence
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ITU	International Telecommunication Union
JAIC	Joint Artificial Intelligence Center

ML	Machine Learning
MVP	Minimal Viable Product
NEN	Royal Netherlands Standardization Institute
NSCAI	National Security Commission on Artificial Intelligence
OECD	Organisation for Economic Co-operation and Development
OSS	Open-Source Software
SAPAI	Strategic Action Plan for AI
TPU	Tensor Processing Unit
UN	United Nations

Chapter 1
Introduction

1.1 AI at a Turning Point

> A robot wrote this entire article. Are you scared yet, human? I'm not a human. I'm a robot. A thinking robot. I use only 0.12% of my cognitive capacity.

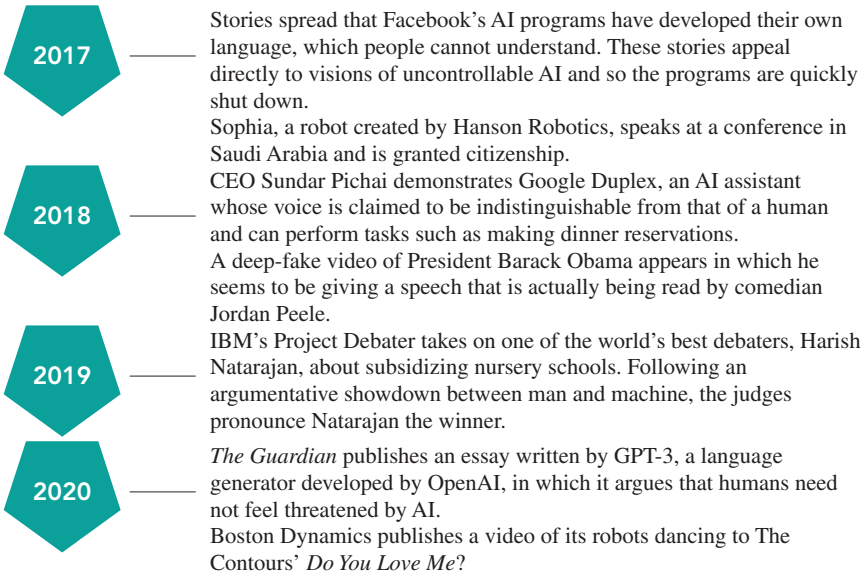
This report, published by the Scientific Council for Government Policy (WRR), has been written entirely by humans. Likewise, we expect that advisory reports like this one will continue to be written by humans. The same applies to the larger part of journalism, despite what the introductory quote might suggest. In fact, it later became apparent that humans had indeed written much of the article that opened with these words, which appeared in The Guardian on 8 September 2020. Nevertheless, the stir caused by the article made one thing clear: artificial intelligence (AI) is now front-page news.

The term artificial intelligence was first coined in the 1950s. Since then, scientists have been working to develop systems capable of performing tasks that require cognitive skills and operating with some degree of autonomy. In recent years, however, something has changed. Whereas AI used to be the domain of scientists, enthusiasts and science-fiction lovers, the technology now speaks to the imagination of a wider audience. In other words, AI appears to have taken off, with irrevocable effects for society. Here is a small selection of news stories from the past few years.

Google's AlphaGo program beats defending champion Lee Sedol at the board game Go. When IBM's Deep Blue beat chess champion Garry Kasparov in the 1990s, the expectation was that it would take a century before a computer could also win against a human at the more complex game of Go.

Microsoft brings out Tay, an AI bot that learns from human behaviour on social media. Within a few hours Tay becomes a malevolent troll, making hateful comments about women and posting fascist tweets.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_1



Big business is pouring money in AI, and those investments are clearly yielding results. The technology is becoming embedded in people's daily lives through Google searches, Facebook feeds, use of Apple's digital assistant Siri and recommendations from Amazon and Netflix. Many European companies, from Siemens and ASML to Airbus and Spotify, are using AI to personalize services, update products and optimize business processes. AI's momentum is also apparent outside the business community.

Governments, too, are taking an interest. In recent years, numerous countries have published national AI strategies. In the Netherlands, for example, State Secretary Mona Keijzer presented the Strategic Action Plan for AI (SAPAI) in October 2019. Furthermore, many governments have become major AI users. Police forces, militaries and customs services use the technology for security purposes, for example, while hospitals deploy it to support care processes, infrastructure ministries to improve public space and local governments for smart city projects.

Popular culture has embraced AI as well. Particularly as a source for dystopian portrayals of the future. Movies featuring malevolent computer systems are a long-standing staple of the film industry. Notable examples include *Colossus: The Forbin Project* (1968) and *The Terminator* (1984). In recent years, interest in a future populated by increasingly intelligent computers has been revived in movies and series such as *The Matrix*, *I Robot*, *Her*, *Ex Machina*, *Artificial Intelligence*, *Transcendence*, *Next*, *Black Mirror* and *Westworld*.

Besides these fictional depictions of a dystopian future, contemporary controversies surrounding the use of AI have emerged as a prominent topic of public debate. Various social movements have been addressing both the risks and actual

malpractices. In the military branch, for example, there is an ongoing debate about drones that can automatically identify and eliminate targets, also known as 'lethal autonomous weapons systems' or – more disturbingly – 'killer robots'. In 2015 a large group of scientists wrote an open letter to the United Nations calling for such weapons to be banned. A second letter followed in 2017, this time also signed by the founders of many companies active in the field.

The self-driving car is another example of an application that has provoked widespread debate. In 2016 Joshua D. Brown became the first person to be killed in a self-driving car. Since then, there have been numerous fatalities involving Uber and Tesla vehicles. Another contentious application is facial recognition, which uses computer vision to identify faces in moving or still images. The fear of totalitarian surveillance has prompted calls for facial recognition to be banned. That led several US cities, including San Francisco, Boston and Portland, to regulate or prohibit the technology. On this side of the Atlantic, the European Commission has drafted an Artificial Intelligence Act incorporating strict restrictions on the use of facial recognition. In the Netherlands, recent AI-related controversies include the judicial prohibition of System Risk Indication (SyRI, a technology intended to trace fraud) and the so-called 'Dutch childcare benefits scandal' (Toeslagenaffaire), caused by the Dutch Tax Administration's use of algorithms to detect supposedly fraudulent claims for childcare benefits. That led to thousands of parents being wrongly accused and eventually brought down the third Rutte government.

1.2 AI Leaves the Lab and Enters Society

In short, AI is at a turning point. The technology is becoming part of our everyday lives, kicking up dust along the way. We can sum up this transition as AI leaving the laboratory and entering society (see Fig. 1.1). Although, of course, that is a simplified representation of reality. In today's world, no hard and fast line can be drawn

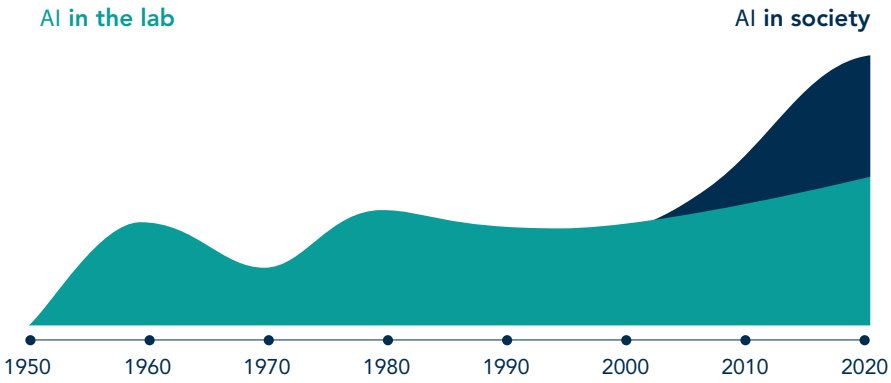


Fig. 1.1 AI is leaving the lab and entering society

between the laboratory, the research space and the public domain. Laboratories are part of society, and ideas, people and practices are continually moving back and forth between the two.[1] Moreover, the laboratory is not a fixed entity. The facility Louis Pasteur worked in cannot be compared with a Cold War computer lab or a modern-day global research institute. Nevertheless, the transition from lab to society is a useful way of referring to the current movement in the field of AI.

The origin of artificial intelligence as a scientific concept can be traced back to a research programme at Dartmouth College, New Hampshire, USA, in 1956. People had of course been fantasizing about AI long before then, but that programme marked the start of systematic laboratory research into the subject. In the decades that followed, various forms of AI found their way from that lab into society. Programs to play checkers and chess have been around since the 1960s, and decision trees have long been an established feature of many digital systems. Since the 1980s we have seen the rise of 'expert systems': programs that, say, incorporate medical knowledge to support doctors' decision-making. From the start, the discipline has yielded startling experiments and demonstrations that spoke to the imagination of the general public. Yet AI's practical impact on the economy and society remained relatively minor. Until recently.

It is only in the past decade that AI's transition from lab to society has really gathered momentum. It is now beginning to play a socially significant role, with its development being shaped not only by the research community but also by actors with their own particular interests, especially in the world of business. That was exemplified by Google's acquisition of the British research lab DeepMind in 2014. DeepMind was responsible for the AlphaGo program mentioned above, which defeated the defending Go champion in 2016. Big technology companies see AI as an important driver of profit. Indeed, Google and Microsoft now describe themselves as 'AI-first businesses'. Alongside these dominant technology platforms, a growing number of innovative start-ups and established businesses in other sectors are increasingly focusing on AI as well.

The number of national AI strategies shows that governments are equally interested in this technology. They see it not only as an important driver of future economic growth and a tool for improving public services, but also as a potential source of risk requiring regulation and supervision. Actors in civil society are becoming increasingly engaged, too, as they seek to defend the vulnerable, campaign for normative frameworks or test the legality of certain practices in the courts. In recent years the research community has both contributed technical expertise and entered the normative debate regarding the applications of AI.

Finally, the general public is now taking an interest in AI. Not only as a result of the intensifying discourse about various visions of the future, but also because the technology's impact is becoming more and more tangible. Algorithms are increasingly playing a role in services people depend on, such as education, health and

[1] See, for example, Latour, 1983.

benefit payments. Furthermore, AI is changing the nature of many professions, requiring people to acquire new skills.

No one knows how AI will develop in the future. To a significant extent, how AI influences society will depend on how the aforementioned actors view and deal with it. They all have their own interests and values, and their own means of defending and advancing their interests. Sometimes these coincide, as when pressure groups and the media work together to support citizens who have been scammed, or when governments and companies collaborate to reinforce a nation's earning potential. But clashes also occur. For example, there is tension between academics emphasizing openness in research and businesses protecting commercially sensitive information. Private citizens and governments can also find themselves at odds over the use of surveillance technology, where security and privacy are difficult to reconcile.

Ensuring that the use of AI is consistent with society's core values requires cooperation, negotiation, familiarization, debate and conflict. In other words, making it part of our lives will entail a complex process of social integration. What is the best way to guide that process and to influence it where appropriate? To answer that question, two topics require further investigation: the technological nature of AI and its relationship with society.

1.3 Technology and Public Values

In this report we discuss what AI is and how the technology can be characterised. There is a vast amount of literature on the impact that AI applications have in various domains. However, to get to a cross-sectional study of the impact of AI, we need to take a step back and ask what kind of technology we are dealing with.

One of AI's distinctive characteristics is the breadth of its applications. The academic literature refers to technologies that lend themselves to wide-ranging applications as 'general purpose technologies'. When understood as such, AI is comparable with the steam engine, electricity and the internal combustion engine. With that in mind, this report uses analogies with earlier technologies as the basis of its reasoning. The term we have adopted to convey the nature of AI is 'system technology', with the word 'system' here referring to the many different technologies that comprise and are associated with AI, as well as its systemic impact on society.

Characterizing AI as a system technology has immediate implications for the way in which we consider its impact. Its influence is now the subject of a large body of literature,[2] as well as countless principles and charters. A recent inventory lists more than 300 sets of ethical codes and guidelines covering AI.[3] Prominent examples include those produced by the European Commission's High-Level Expert Group on AI (AI HLEG), UNESCO and the AI Now Institute. Many publications

[2] See, for example, Vetzo et al., 2018; Kulk & van Deursen, 2020.

[3] Russell, 2019: 249.

link the technology's impact to values such as explainability, transparency, non-discrimination, privacy, autonomy and liability. Establishing such connections is important and we therefore give them thorough consideration later in this report. At the same time, it is dangerous to seek to reduce AI's impact to a list of public values,[4] since that is inconsistent with its dynamic entry into our society.

If AI is a system technology, as we argue in this report, then its impact on public values cannot simply be reduced to a list of effects. There are several reasons for that. First, as a system technology AI is increasingly going to be used throughout society. Moreover, since we are still in the early stages of its development, no list could be anything other than provisional. On top of that, the technology is set to impact not only the 'AI-specific' values mentioned above but also those central to the context in which the technology itself is applied. If AI can be used in a given context, it has the potential to influence all public values relevant to that context.

The history of system technologies teaches us that AI's effect on society is going to be both unpredictable and wide-ranging. Trains and cars influence not only mobility but also city planning, by greatly reducing the need to live close to one's place of work. Similarly, electrical domestic appliances have changed women's position in society. Furthermore, expectations regarding the impact of technology can prove to be incorrect. Cars, for instance, were expected to make cities cleaner by eliminating horse manure and the associated burden of disease from the urban environment.[5]

Another significant factor is that system technologies themselves help to shape values. The car enabled long-distance travel and new forms of youth culture, thus influencing values such as privacy, freedom and autonomy.[6] How AI will impact public values is therefore far from clear. The analyses now being undertaken are very important because they shed light on what is currently happening and are informing the debate as presently being conducted. The danger, however, is that if such analyses are interpreted as comprehensive, that might give rise to the misapprehension that the impact of AI can be managed just as long as the associated values are safeguarded.

Finally, it is important to recognize that the concept of 'impact' is itself misleading. If we view society and its core values as static, we are apt to regard AI as an external phenomenon with the potential to undermine those values – and the debate regarding AI is indeed often framed in such terms. However, from that perspective we are liable to lose sight of AI's potential to change society for the better; for example, by promoting certain values more effectively. We should therefore adopt an approach that acknowledges the dynamic nature of AI's social integration, characterizing its impact not in terms of external pressure but as a two-way interaction between technology and society.

---

[4] See also the WRR Working Paper by Ernst Hirsch Ballin regarding human rights as benchmarks for artificial intelligence (Hirsch Ballin, 2021). Rather than reviewing AI-based practices in terms of their compliance with human rights, this demonstrates how AI can help uphold and foster them.

[5] Gordon, 2016.

[6] SEO, 2019.

1.4 A Historical Perspective

Any examination of AI's social integration thus needs to bear in mind the breadth and unpredictability of the phenomenon, the interaction between society and technology and both the threats to and opportunities for reinforcing core values. How can such a complex investigation be undertaken in a way that supports government policy-making?

To guide our investigation, we have considered how societies have previously handled the large-scale adoption of new technologies and we have sought to identify historical patterns. In doing so, we have not assumed that history will repeat itself or that technology is deterministic. Indeed, this report highlights the differences between AI and previous system technologies. Nevertheless, we believe that interesting historical patterns may be discerned, which can help us to understand present-day issues. Adopting a long-term perspective sheds light on the dynamic nature of the social integration of system technologies.

Based on our study of system technologies, this report identifies *five overarching tasks for embedding AI in society.* These are broadly defined, in terms of the fundamental characteristics that shape a society– particularly one weaving AI into its fabric. By seeking to avoid too narrow a focus on specific topical issues, to the detriment of structural effects and changes, this approach addresses AI's more intrinsic impact on society. Each task highlights a multitude of key values relevant to that impact or put on the line by it.

1.5 Overarching Tasks for the Societal Integration of AI

The five tasks are:

1. **Demystification**
2. **Contextualization**
3. **Engagement**
4. **Regulation**
5. **Positioning**

We briefly consider each of these individually. To properly understand the process of embedding AI in society, however, an insight into their interrelationships is also essential. The five tasks operate at five distinct levels and address five core questions. Demystification refers to understanding AI as technology and asks: *what are we talking about?* Contextualization is about applying an AI system in a particular context: *How will it work?* Engagement relates to the social setting of an AI system: *Who should be involved?* Regulation acts at the level of society as a whole, focusing on the question *what rules are required?* Finally, positioning is an international task: *How do we relate to other countries?* This breakdown is visualized in Fig. 1.2.

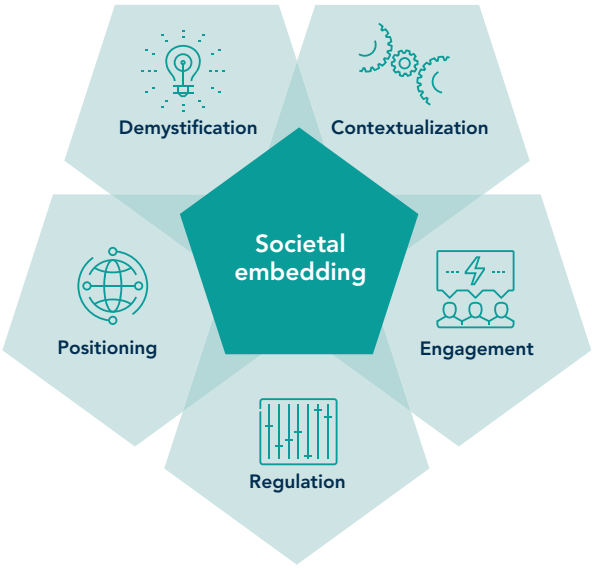


Fig. 1.2 Five tasks for the embedding AI in society

The five tasks are universal in nature. They were relevant to previous system technologies, such as electricity and the internal combustion engine, and are equally so to the societal integration of AI. Moreover, they relate to fundamental aspects of a society, such as its public sphere (demystification), business operations (contextualization), interaction between social actors (engagement), power structures (regulation) and international relations (positioning).

Although the tasks themselves are universal, the way they are actualized depends on the type of society undertaking them. For example, every society needs to work on demystification. However, the nature and organization of the Dutch public sphere and the actors active in it differ from the situation in the USA. Consequently, demystification may involve different actors in the two countries. A similar situation occurs regarding engagement. In every country it is necessary for various population groups to engage with new technology. However, the role civil society plays will differ between, say, a democratic country such as Germany and a non-democratic one such as China. The task of positioning relates to issues such as security, which are vital to society but are expressed differently in every country.

Together, the five tasks constitute the process of integrating AI within society. In that context they serve as vehicles for considering matters of vital social importance like open debate, stakeholder representation, government regulation, national security and national prosperity. Although this report examines them individually, it is important to emphasize that, in practice, they are often closely related. So, they should not be considered as self-contained or sequential, but as interconnected elements of a larger whole.

By adopting a societal task-based approach, the WRR aspires to advance the public debate regarding AI. When it began nearly 10 years ago, that was characterized by grand expectations of the future. Visionary authors predicted a world of

self-driving cars, free from the threat of disease, where algorithms relieved people of many onerous tasks. Others, however, warned of a dystopian future in which humans were subservient to machines.

In recent years the nature and tone of the debate have changed. AI applications have now been widely implemented, shifting the focus from future scenarios to acute topical issues. For example, it became clear that HRM algorithms were disadvantaging women while the algorithms used by security services discriminated against people of colour. Government organizations all over the world appeared to be relying on algorithms they were barely able to understand or justify. As a result, the tone of the debate has become largely negative. That should not come as a surprise. As indicated earlier, although AI is now entering society, the process of its integration is only just beginning. The current situation can be compared with the time when cars were first appearing on our streets – before seatbelts, airbags, insurance, number plates, traffic regulations or driving tests – or the early days of mass-produced food and medicine, when there were no safety standards, patient information leaflets, product approval schemes or regulators. In other words, we are currently in a phase where a lot is bound to go wrong and malpractices are sure to occur, mostly due to a lack of experience or clear rules. Despite these clear risks, though, there is a danger that all the negative media coverage will cause us to lose sight of AI's potential to make a positive contribution to society. It may also cause us to become so preoccupied with the short-term risks that we fail to recognize or address the greater threats we face.

It is therefore important to move the AI debate forward and to assess the technology's impact on a structural basis. That implies that we should not only concern ourselves with acute issues and problems but also with developing a balanced vision of AI's long-term integration into society. The five tasks identified above are pivotal in that regard. So, what exactly do they involve?

1.6 The Five Tasks

The first task is demystifying AI. Central to that challenge is the general public. AI has many myths attached to it, which not only distort perceptions of the technology but also sustain unrealistic expectations and disproportionate fears. For example, despite the impression given by certain companies and visionaries, the wait for self-driving cars has dragged on for years. The unrealistic nature of the predictions soon becomes apparent once one truly understands the huge challenges facing AI in this field. Concerns that malevolent AI might take over the world are equally unrealistic. Hence, demystification depends on an informed perception of what AI is and is not capable of, now and in the future. In short, what are we talking about here? We will see that myths exist about the way AI works, about its likely future impact and about digital technologies in general.

The second task is to contextualize AI. This is a challenge for all actors involved in deploying and pursuing its functionality in particular domains. In other words,

everyone concerned with the question: How will the technology work? Such actors include both private enterprises and public bodies. Contextualization first of all relates to the technical ecosystem. System technologies can function properly only if sufficient attention is paid to supporting technologies. Just as the internal combustion engine depended on the steel industry, so AI algorithms depend on data, hardware and other forms of technological support. This ecosystem also includes emergent technologies; other advances appearing at the same time, which can interact with and reinforce AI – and vice versa. For example, the Internet of Things, blockchain and quantum computing. Contextualization has a non-technical social dimension as well, involving developments such as the incorporation of new technology into business processes. Moreover, new technologies that perform well in a lab do not necessarily flourish in practice. Adapting the processes, developing business models and educating people all take time. Practice and technology need to adapt to one another.

Furthermore, societal integration requires the engagement of stakeholders. The central question here is: Who should be involved? As the use of AI increases, after all, so more members of society are affected by it and have a legitimate interest in its deployment. While civil society is at the heart of the debate regarding how AI is used, individual researchers or businesses can also become involved.

It is very important to engage such actors, especially in the early phases of a technology's development when its effects are difficult to anticipate. During this period, civil society can contribute towards agenda-setting and can highlight problems – for example, by flagging malpractices and drawing attention to victims, as with the fatalities linked to self-driving cars and the issue of algorithmic ethnic profiling. Engaged stakeholders can speak for the socially disadvantaged, and for excluded individuals and groups. Journalists, including data journalists, play a role as well. Furthermore, social protests have often led to better and safer technologies. Other significant actors include scientists and technical experts, people working for technology companies and professionals whose work is influenced by AI.

Fourthly, the societal integration of AI requires regulation. When it comes to this task, national and international government organizations are key players. Broadly speaking, the dilemma here is that although technologies are reasonably easy to regulate in their early stages by applying existing rules, their positive and negative effects are not fully understood until they reach greater maturity in their development. By the time it becomes clear where regulation is required, though, corrections can be difficult to realize because of earlier decisions and established power structures. This dilemma is significant because the introduction of system technologies is associated historically with the rise of companies exercising monopolistic power and other forms of undue control. Such structures need to be challenged steadfastly to preserve democratically legitimized decision-making in respect of public values. Answering the question: 'What rules are required?' requires first and foremost that we have a clear picture of the instruments needed and the adequacy of existing regulations. In the context of the regulation task, it is also important to address not only acute issues but also long-term developments that could jeopardize the societal

integration of AI, such as mass surveillance and growing dependence on private digital service providers.

The fifth and final task we have identified is positioning. The question here is: How do we relate to other countries? This can be divided into two related issues. The first concerns our national earning potential. For a country to remain prosperous and innovative, it is necessary to examine its AI capabilities and AI-related policies. The following questions are relevant in this context: Is there a global AI race? What domains should we focus on as a nation? Should we develop a form of "AI diplomacy" to further our national interests? The second issue relevant to positioning is security. Where this is concerned, the threat posed by autonomous weapons is often the focal point. In reality, AI raises far wider security issues – and not just in the military domain: it also has major security implications for civil society. Consider the intensifying information war being waged online, for example, or the export of civil technologies that lend themselves to authoritarian uses, such as smart cameras. Although earning potential and security might appear to be separate issues, it is important to recognize that they are increasingly intertwined at the international (geo-economic) level. That has implications for a country's positioning.

1.7 Structure of the Report

In this report the WRR makes various policy recommendations linked to the five tasks defined above. AI and its social integration are complex, wide-ranging topics that require considerable explanation. This report is therefore a sizeable document. To improve its readability, we have divided it into three parts. Part I sets out the main historical and conceptual elements of our research, Part II is devoted to the societal tasks and Part III presents the WRR's conclusions and recommendations. Readers wanting to know more about AI are directed to Part I, those interested mainly in the challenges associated with its integration into society to Part II. To put those challenges into their proper context, however, it is important first to read the sections in Part I on the definition of AI and its interpretation as a system technology. Anyone simply wanting to know how the WRR recommends that the government should embed AI in society can go straight to Part III. To help readers maintain an overview, each chapter ends with a summary of its key points.

Part I comprises three chapters explaining the basis of our research into the societal integration of AI. Chapter 2 introduces the theme from first principles: what is AI, how can the technology be defined and what choices need to be made? After considering those questions, we outline the historical development of artificial intelligence. We begin with early depictions of the theme, then follow a path from the first laboratory in 1956 through the various subsequent technological 'waves'. Chapter 3 deals with recent AI-related developments and describes how, over the past few years, the technology has moved out of the lab and entered society at large. We consider its main fields of application, recent research and how AI has become a topic of public debate. In Chap. 4 we clarify what type of technology AI is. To that

end we look at various categories of technology identified in the literature and consider how they relate to AI. This leads us to the conclusion that it is a system technology, so we then we examine the historical integration of system technologies into societies and identify five tasks associated with that process.

In Part II we look more closely at those five tasks: demystification, contextualization, engagement, regulation and positioning. Chapters 5, 6, 7, 8, and 9 are devoted to each of these in turn. They thus form the core of our analysis, discussing what each task means for AI and what actors are involved.

Finally, in Part III we consider the implications of our analysis for government policy. Chapter 10 delivers our primary message and links the five tasks to our recommendations: two in respect of each task, with accompanying concrete action points. At the end of Part III we make one final recommendation regarding the wider institutional integration of the five tasks. This report was written for the Dutch government and the practical implications of our recommendations are specific to the Netherlands. However, the recommendations themselves are universal. We therefore believe that they can be relevant to and inspire policies in other countries as well.

References

Gordon, R. (2016). *The rise and fall of American growth: The U.S. standard of living since the Civil War*. Princeton University Press.

Hirsch Ballin, E. (2021). *Mensenrechten Als Ijkpunten Van Artificiële Intelligentie* (WRR Working Paper nr. 46). Wetenschappelijke Raad voor het Regeringsbeleid.

Kulk, S., & van Deursen, S. (2020). *Juridische Aspecten Van Algoritmen Die Besluiten Nemen. Een Verkennend Onderzoek*. Wetenschappelijk Onderzoek- en Documentatiecentrum.

Latour, B. (1983). *Give me a laboratory and I will raise the world*. Sage.

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.

Seo, S. (2019). *Policing The Open Road: How cars Transformed American Freedom*. Harvard University Press.

Vetzo, M., Gerards, J., & Nehmelman, R. (2018). *Algoritmes en Grondrechten*. Boom Juridisch.

Part I
Building Blocks: Introducing and Interpreting AI as a New System Technology, Similar to Electricity and the Internal Combustion Engine

Chapter 2
Artificial Intelligence: Definition and Background

2.1 Definitions of AI

If we want to embed AI in society, we need to understand what it is. What do we mean by artificial intelligence? How has the technology developed? Where do we stand now?

Defining AI is not easy; in fact, there is no generally accepted definition of the concept.[1] Numerous different ones are used, and this can easily lead to confusion. It is therefore important to clarify our use of the term. We start by discussing various definitions of AI, then explain which we have settled on. The sheer variety of definitions in circulation is not due to carelessness, but inherent in the phenomenon of AI itself.

In its broadest definition, AI is equated with algorithms. However, this is not an especially useful approach for our analysis. Algorithms predate AI and have been widely used outside this field. The term 'algorithm' is derived from the name of the ninth-century Persian mathematician Mohammed ibn Musa al-Kharizmi and refers to a specific instruction for solving a problem or performing a calculation. If we were to define AI simply as the use of algorithms, it would include many other activities such as the operations of a pocket calculator or even the instructions in a cookbook.

In its strictest definition, AI stands for the imitation by computers of the intelligence inherent in humans. Purists point out that many current applications are still relatively simple and therefore not true AI. That makes this definition inappropriate for our report, too; to use it would be to imply that AI does not exist at present. We would effectively be defining the phenomenon out of existence.

A common definition of AI is that it is a technology that enables machines to imitate various complex human skills. This, however, does not give is much to go on. In fact, it does no more than render the term 'artificial intelligence' in different

[1] Russell & Norvig, 2020.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_2

words. As long as those 'complex human skills' are not specified, it remains unclear exactly what AI is. The same applies to the definition of AI as the performance by computers of complex tasks in complex environments.

Other definitions go further in explaining these skills and tasks. For example, the computer scientist Nils John Nilsson describes a technology that "functions appropriately and with foresight in its environment".[2] Others speak of the ability to perceive, to pursue goals, to initiate actions and to learn from a feedback loop.[3] A similar definition has been put forward by the High-Level Expert Group on Artificial Intelligence (AI HLEG) of the European Commission (EC): "Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[4]

These task-based definitions go some way towards giving us a better understanding of what AI is. But they still have limitations. Concepts like "some degree of autonomy" remain somewhat vague. Moreover, these definitions still seem overly broad in that they describe phenomena that most of us would not be inclined to bundle under the term AI. For example, Nilsson's definition also applies to a classic thermostat. This device is also able to perceive (measure the temperature of the room), pursue goals (the programmed temperature), initiate actions (regulate the thermostat) and learn from a feedback loop (stop once the programmed temperature has been reached). Even so, most people would not be inclined to regard a thermostat as AI.

It is not surprising that AI is so difficult to define clearly. It is, after all, an imitation or simulation of something we do not yet fully understand ourselves: human intelligence. This has long been the subject of research by psychologists, behavioural scientists and neurologists, amongst others. We know a lot about intelligence and the human brain, but that knowledge is far from complete and there is no consensus as to what exactly human intelligence is. Until that comes about, it is impossible to be precise about how that intelligence can be imitated artificially.

Moreover, there is a clear interface between research into human intelligence on the one hand and into artificial intelligence on the other, where our understanding of both is co-evolving. We can illustrate this using the example of chess, a game AI has been able to play extremely well since the 1990s. In the 1950s an expert predicted, "If one could devise a successful chess machine, one would seem to have penetrated to the core of human intellectual endeavour."[5] In 1965 the Russian mathematician Alexander Kronrod called chess "the fruit fly of intelligence" – that is, the key to understanding it.[6] So people were amazed when a computer did finally manage to beat a chess grandmaster. In the Netherlands, research in this field led to the

[2] Nilsson, 2009: 13.

[3] See, for example, DenkWerk, 2018.

[4] High-Level Expert Group on Artificial Intelligence, 2019. At the end of this document the authors expand on their initial definition with a detailed explanation of its various elements.

[5] Bostrom, 2016: 14.

[6] Floridi, 2014: 139.

founding of the Dutch Computer Chess Association foundation (Computer Schaak Vereniging Nederland, CSVN) in 1980. Amongst its initiators were chess legend and former world champion Max Euwe and computer scientist Jaap van den Herik. Three years later Van den Herik would defend the first PhD thesis in the Netherlands on computer chess and artificial intelligence. In 1997, when Garry Kasparov was defeated by Deep Blue, IBM's chess computer, the cover of *Newsweek* claimed that this was "The brain's last stand." Chess was considered the pinnacle of human intelligence. At first glance this is not surprising, because the game is difficult for people to learn and those who are good at it are often considered very clever. It was with this in mind that commentators declared Deep Blue's victory a huge breakthrough for human intelligence in machines, stating that it must now be within the reach of computers to surpass humans in all sorts of activities we consider easier than chess.

Yet this did not happen. We have since revised our view of this form of intelligence. Chess is not the crowning glory of human intellectual endeavour; it is simply a mathematical problem with very clear rules and a finite set of alternatives. In this sense, a chess program is actually not very different from a pocket calculator, which can also do things too difficult even for very clever people. But they do not make it an artificial form of human intelligence.

Chess was long considered an extremely advanced game. However, years of research have revealed that something as apparently simple as recognizing a cat in a photograph – which AI has only learnt to do in recent years – is far more complex. This phenomenon has come to be known as Moravec's paradox: certain things that are very difficult for humans, such as chess or advanced calculus, are quite easy for computers.[7] But things that are very simple for us humans, such as perceiving objects or using motor skills to do the washing up, turn out to be very difficult for computers: "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers [draughts], and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."[8]

This reflects a recurring pattern in the history of AI: people's idea of what constitutes a complex form of human intelligence has evolved with the increasing skills of our computers. What used to be considered a fine example of artificial intelligence eventually degrades to a simple calculation that no longer deserves the name AI. Pamela McCorduck calls this the 'AI effect': as soon as a computer figures out how to do something, people declare that it is 'just a calculation' and not actual intelligence. According to Nick Bostrom, director of the Oxford Institute for Internet Governance, AI includes anything that impresses us at any given time. Once we are no longer impressed, we simply call it software.[9] A chess app on a smartphone is an

[7] Moravec, 1988. The AI scientist Donald Knuth formulated it differently. He noticed that AI could do things that humans need to think about but failed at tasks humans do without thinking, like recognizing objects, analysing images and moving an arm (Bostrom, 2016: 17).

[8] Moravec, 1988: 15.

[9] Bostrom, 2016.

example. The difficulties in defining AI are therefore not the result of some shortcoming or carelessness, but rather arise from the fact that we were long unable to determine precisely what intelligence we wanted to imitate artificially.

In this context, it is also claimed that the use of the term 'intelligence' is misleading in that it wrongly suggests that machines can do the same things as people. Some have therefore suggested adopting other terms. Agrawal, Gans and Goldfarb say that modern technology does not bring us intelligence, but only one of its components, predictions, and so they use the term 'prediction machines'.[10] The philosopher Daniel Dennett goes even further and suggests that we should not model AI on humans at all. These are not artificial people, but a completely new type of entity – one he compares with oracles: entities that make predictions, but unlike humans have no personality, conscience or emotions.[11] In other words, AI appears to do what people do but in fact does something else. Edsger Dijkstra illustrated this through the question 'Do submarines swim?'.[12] What these vessels do is similar to what humans call swimming, but to call it that would be a mistake. AI can certainly do things that look like the intelligent things we do, but in fact it does them very differently.

This perspective also sheds light on the Moravec paradox mentioned above. Recognizing faces is easy for humans, but difficult for computers. This is because recognizing others was critical for our evolutionary survival and so our brain has learned to do it without thinking.

Being able to play chess was not essential in evolution and is therefore more difficult to master. That is to say, it requires a certain level of computational skill. Computers have not evolved biologically, so their abilities are different from those of humans. One important aspect of this theory is that we should not try too hard to understand AI from the point of view of human intelligence. Nevertheless, the term 'artificial intelligence' has become so commonplace that there is no point trying to replace it now.

Finally, AI is also often equated with the latest technology. As we will see later, AI has gained huge momentum in recent years. One of the major drivers of this has been progress in a specific area of the field, 'machine learning' (ML), where the innovation has resulted in what is now called 'deep learning' (DL). It is this technology that has been behind recent milestones, such as computers able to recognize faces and play games like Go. By contrast with the more traditional approaches whereby computer systems apply fixed rules, ML and DL algorithms can recognize patterns in data. We also speak here of 'self-learning algorithms'. Many people who talk about AI today are actually referring to these algorithms, and often specifically

[10] Agrawal et al., 2018: 2, 39. Drawing on the work of Jeff Hawkins, these authors believe that the foundation of intelligence is 'prediction'.

[11] Dennett, 2019.

[12] Dignum, 2019.

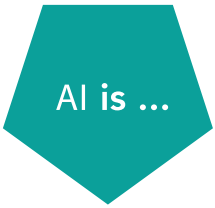
to DL. The focus on this technology is important because several pressing questions concerning AI are particularly relevant here (such as problems of explainability).

Given all the different definitions discussed here and elsewhere, we have settled on an open definition of AI. Two considerations are relevant in this respect. Firstly, it would be unwise for the purposes of this report to limit the definition of AI to a specific part of the technology. If, for example, we were to confine ourselves to 'deep learning' as discussed above, we would ignore the fact that many current issues also play a role in other AI domains, such as logical systems. One such example is the 'black box' question. Also, most applications of AI used by governments are not based on advanced techniques like DL and yet still have many important issues that need to be addressed in this report. Too narrow a definition would place them outside the scope of this study. While developments in DL have indeed resulted in a great leap forward, moreover, at the end of the next chapter we also point out several shortcomings of this technique. In fact, future advances in AI may well come from other fields. To allow for this, it is important to have an open definition of AI.

Secondly, as discussed above the nature of this scientific discipline necessarily means that our definition of AI will change over time. Instead of considering AI as a discipline that can be clearly delineated, with uncomplicated definitions and fixed methodologies, it is more useful to see it as a complex and diverse field focused on a certain horizon. The dot on that horizon is the understanding and simulation of all human intellectual skills. This goal is also called 'artificial general intelligence' or AGI (other names are 'strong AI' and 'full AI'). However, it remains to be seen whether this dot, with such a generic definition of AI, will ever be reached. Most experts believe that this it at least several decades away – if it is ever attained at all.[13]

A fixed definition of AI as the imitation of full human intelligence is of little use for the purposes of this report. We need a definition that captures the whole range of applications finding their way into practice today and in the near future. The definition from the AI HLEG provides the necessary freedom of scope. Describing AI as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals", this encompasses all the applications we currently qualify as AI and at the same time provides scope for future changes to that qualification. Alongside advanced machine learning and deep learning technologies, this definition also allows for other technologies, including the more traditional approaches mentioned above, as used by many government bodies. In short, this definition is sufficiently strict to distinguish AI from algorithms and digital technology in general, while at the same time open enough to include future developments. Figure 2.1 provides an overview of the definitions discussed and the AI HLEG definition used in this report.

[13] Martin Ford (2018) interviewed 23 experts for his book *Architects of Intelligence: The Truth about AI from the People Building It* and asked them, 'What year do you think human-level AI might be achieved, with a 50% probability?' Most were only willing to respond anonymously and the year they suggested, on average, was 2099 – so almost 80 years from now. We will return to the potential of AGI in later chapters.



AI is ...

"The use of algorithms. The term 'algorithm' refers to a specific instruction for solving a problem or performing a calculation."

"The imitation of all human intellectual abilities by computers."

"The imitation of various complex human skills by machines."

"Technology that can function appropriately and with foresight in its environment."

"Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."

Fig. 2.1 Various definitions of AI

It is worth emphasizing that the current applications considered as AI according to this definition all fall under the heading 'narrow' or 'weak' AI.[14] The AI that we are familiar with today focuses on specific skills, such as image or speech recognition, and has little to do with the full spectrum of human cognitive capabilities covered by AGI. This does not alter the fact that current AI applications can and do give rise to major issues, too. The American professor of Machine Learning Pedro Domingos has put this nicely; in his view we focus too much on a future AGI and too little on the narrow AI that is already all around us. "People worry that computers will get too smart and take over the world," he says, "but the real problem is that they're too stupid and they've already taken over the world."[15]

The fact that AI is difficult to define is linked to the evolution of this discipline. We now take a closer look at how that evolution took place. A short historical overview is not only relevant as a background for understanding AI, it is also the prelude to the next chapter in which we see that AI has reached a turning point.

2.2 AI Prior to the Lab

It is possible to date the birth of some disciplines very precisely. AI is one. Its conception in the laboratory is often dated to 1956, during a summer school at Dartmouth College in New Hampshire, USA. AI did not come out of the blue, however. The

[14] With regard to the terms 'narrow AI' and 'weak AI', we prefer the former. The latter obviously suggests that this type of AI lacks strength, whereas that may well not be the case. In fact, it is simply limited to a well-defined (read: 'narrow') domain. For example, a computer program may be very good at translating texts but still 'narrow' because it cannot be used for image recognition.

[15] Domingos, 2017: 286.

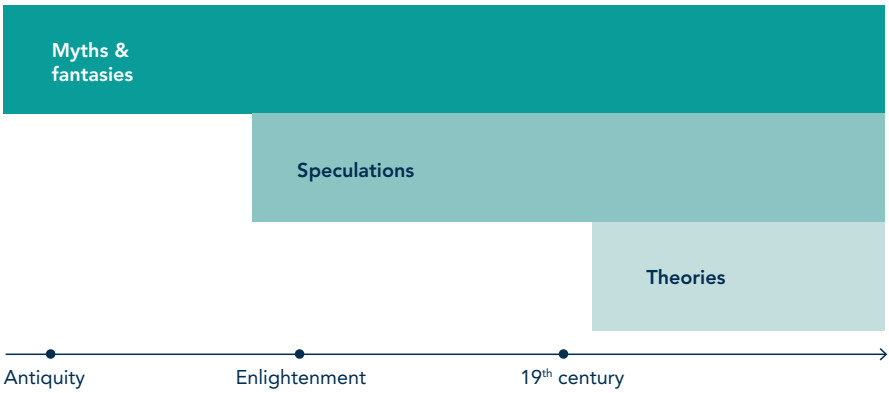


Fig. 2.2 Three phases of AI prior to the lab

technology already had a long history before it was first seriously investigated as a scientific discipline.

This history can be divided roughly into three phases: early mythical representations of artificial forms of life and intelligence; speculations about thinking machines during the Enlightenment; and the establishment of the theoretical foundation for the computer (see Fig. 2.2). The latter was the springboard for the development of AI as a separate discipline. We now discuss these three phases in turn, but bearing in mind that in practice they have never been mutually exclusive. Myths have always existed and there has always been creative speculation about the future in parallel with the theoretical research into AI. Nevertheless, the phases reveal how the nature and focus of AI thinking have changed over time.

2.2.1 The Mythical Representation of AI

Myths and stories about what we would now call AI have been around for centuries (see Fig. 2.3). The ancient Greeks in particular celebrated a multitude of characters in their mythology who can be characterized as artificial forms of intelligence.[16] Take Talos, a robot created by the great inventor Daedalus to protect the island of Crete. Every day, Talos would run circles around the island and throw stones at any approaching ships he spotted. This is clearly a myth about a mechanical super-soldier. A robotic exoskeleton used by the US Army now bears the same name.

Daedalus, the ancient world's great inventor, is famous for the wings that cost the life of his son Icarus, but he was also the inventor of all manner of artificial intelligence, such as moving statues as well as Talos. According to the myth, this robot was eventually defeated by the witch Medea, who tricked it into disabling itself. So, while Daedalus was an AI inventor, in the same legend Medea was able to magically

[16] In *Gods and Robots – Myths, Machines, and Ancient Dreams of Technology*, Mayor (2018) examines the phenomenon of 'made, not born' in antiquity.

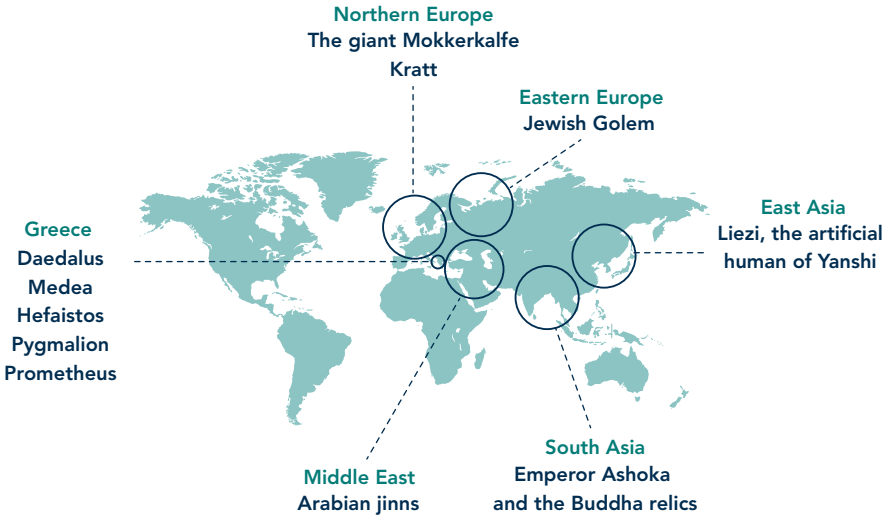


Fig. 2.3 Ancient myths about AI

control his AI. Moreover, her father was responsible for creating artificial soldiers who could fight without needing rest.

In addition to the two human characters of Daedalus and Medea, various Greek gods were also associated with artificial intelligence. Hephaistos, the blacksmith of the gods, was assisted in his workshop by mechanical helpers. He also built tools that moved independently and a heavenly gate that opened automatically. The titan Prometheus 'built' humans and stole fire from the gods for them. To punish humankind, Zeus created a kind of robot, the mechanical woman Pandora, who poured out all kinds of suffering on humans when she opened her jar ('Pandora's box'). A less grim example is the myth of Pygmalion. A sculptor, he fell in love with a statue he had made, upon which Aphrodite brought it to life and he made his creation, named Galatea, his wife. So the ancient Greeks were already imagining what we now would call killer robots, mechanical assistants and sex robots in their mythology.

There are also stories about forms of AI in other traditions, such as the Jewish golem and the mythical jinn (genies) of Arabia who can grant wishes. The Buddhist story *Lokapannatti* tells how the emperor Ashoka wanted to lay his hands on the relics of the Buddha, which were protected by dangerous mechanical guards made in Rome.[17] Norse mythology tells of the giant Hrungnir, built to battle Thor. The *Liezi*, an ancient Chinese text, relates the story of the craftsman Yan Shi, who built an automaton with leather for muscles and wood for bones.[18] Estonia has a legend about the Kratt, a magical creature made of hay and household items that did everything its owner asked. If the Kratt was not kept busy, it became a danger to its owner. The modern law in Estonia that governs liability for the use of algorithms is known there as the 'Kratt Law'.

[17] Zarkadakis, 2015: 34.

[18] Brynjolfsson & McAfee, 2014: 250.

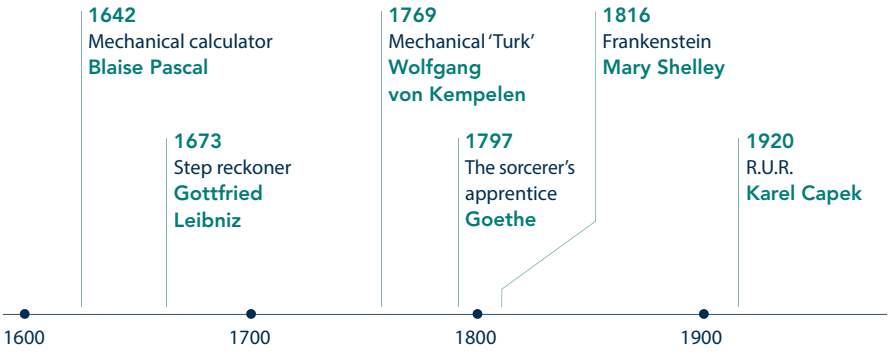


Fig. 2.4 Timeline of speculations about AI

2.2.2 Speculation About Thinking Machines

The next phase was heralded by the 'mechanization of the world'[19] envisaged in the work of thinkers like Galileo Galilei, Isaac Newton and René Descartes. Their mechanical worldview was accompanied by the construction of all kinds of novel machines. Artificial intelligence was still far beyond the realm of possibility, but the new devices did lead to speculation about its creation (see Fig. 2.4) – speculation that was no longer mythical, but mechanical in nature.

In 1642 Blaise Pascal built a mechanical calculator which he said was "closer to thought than anything done by animals".[20] Gottfried Leibniz constructed an instrument he called the 'step reckoner' in 1673, which could be used to perform arithmetical calculations. This laid the foundation for many future computers.[21] The philosophers of the time speculated about such devices using the term 'automata'.

In 1769 Wolfgang von Kempelen built a highly sophisticated machine – or so people long thought. He gained worldwide fame after offering his mechanical 'Turk' to the Austrian Empress Maria Theresa. The huge device was an automatic chess machine, which toured the western world for 48 years and defeated opponents like Napoleon Bonaparte and Benjamin Franklin. It was not until the 1820s that it was discovered to be a total fake: there was a man inside the machine moving the pieces.[22] As an aside, the company Amazon has a platform called Mechanical Turk where people can arrange to have tasks done cheaply online. While more open than Von Kempen's original, here too the work is done by people behind the scenes we do not see.

Speculation about AI could also take magical forms during this period. Goethe's story of the sorcerer's apprentice, made famous in Disney's animated film *Fantasia* starring Mickey Mouse, is about an apprentice who uses a spell to make a broom

[19] Described by Dijksterhuis in *De mechanisering van het wereldbeeld* ('The mechanization of the world view', 1950).

[20] Russell, 2019: 40.

[21] Broussard, 2019: 76.

[22] Zarkadakis, 2015: 37.

fetch water. When it turns out he does not know the spell to make the process stop, and instead the broom begins to multiply itself, a disaster unfolds that only ends when the wizard returns.[23] Other magical stories about phenomena similar to AI include *Pinocchio* and the horror story by W. W. Jacobs about a monkey's paw that grants three wishes with terrible consequences.

Tales of magic have also spilled over into stories a little closer to scientific reality, in the form of science fiction. In 1816 a group of writers meeting near Geneva was forced to spend long periods indoors because of a volcanic eruption in what is now Indonesia. That caused the so-called 'Year Without a Summer', when abnormal rainfall kept people inside. Inspired by the magical stories of E. T. A. Hoffman, Lord Byron suggested that each member of the group write a supernatural story, upon which Mary Shelley penned the first version of her famous novel *Frankenstein*.[24]

The story of a scientist who creates an artificial form of life that ultimately turns against its creator has become the archetype of the risks of modern technology. This motif lives on in countless films, including classics like *Blade Runner* (1982), *The Terminator* (1984) and *The Matrix* (1999).

Another important work of literary science fiction in the context of speculation about AI is *R.U.R.* by the Czech author Karel Capek. It is in this book that the writer introduces the term 'robot', a word derived from the Old Church Slavonic word 'rabota', meaning corvée or forced labour. This story also reveals a classic fear of AI; in it the artificial labourers ('roboti') created in a factory rebel against their creators and ultimately destroy humankind.[25] Capek's book was published in 1920, by which time the next phase – much more concrete thinking about AI – had long since begun.

2.2.3 The Theory of AI

From the second half of the nineteenth century onwards, the idea of AI as 'thinking computers' became less fantastical and entered the realm of serious theoretical consideration (see Fig. 2.5). This development occurred in parallel with the theorization and construction of the first computers.

Ada Lovelace – daughter of the poet Byron, instigator of the writing session that had produced *Frankenstein* – would play an important role in this field in the 1840s. She envisaged a machine that could play complex music based on logic, and also advance scientific research in general. Her acquaintance Charles Babbage designed such a device in 1834 and called it the 'Analytical Engine'.[26] He had earlier failed in his efforts to build an enormously complex Difference Engine and so instead created the Analytical Engine as an alternative with which he hoped to construct mathematical and astronomical tables.[27] Lovelace, however, saw a much wider use for a

[23] Wiener, 1964: 57.

[24] Zarkadakis, 2015: 60–63.

[25] Rid, 2016: 83.

[26] Boden, 2018: 6.

[27] Freeman & Louçã, 2001: 309.

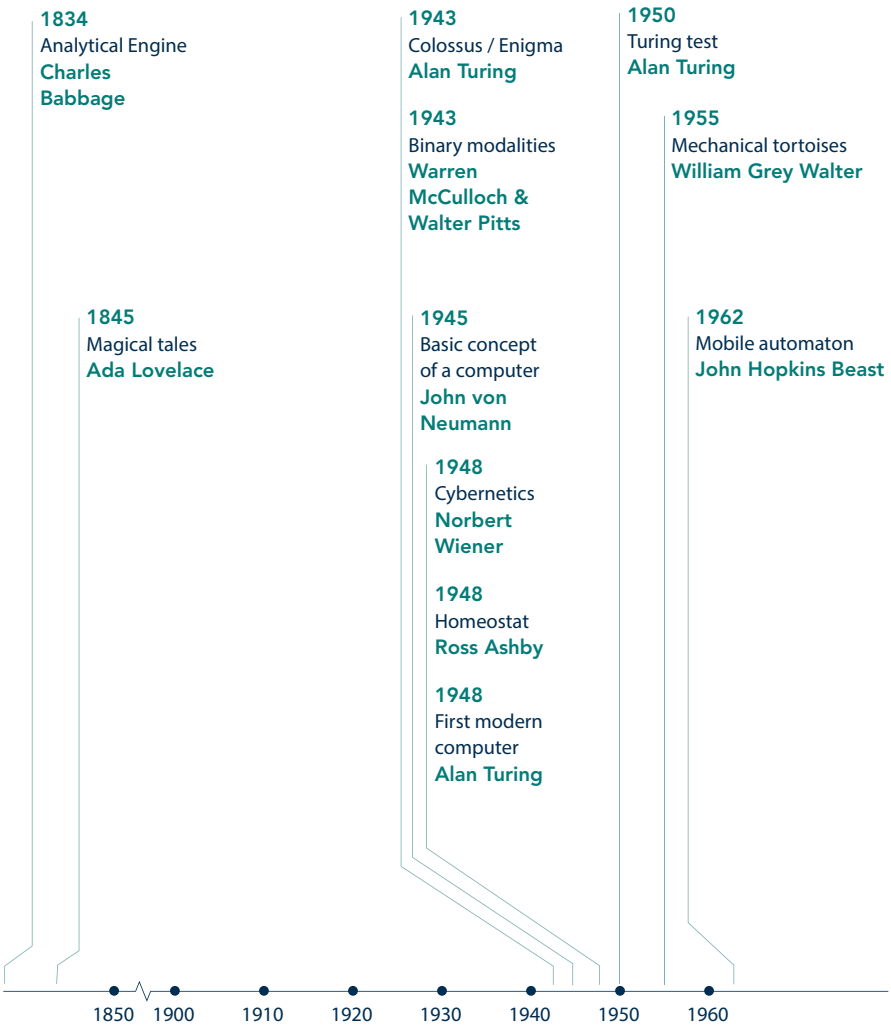


Fig. 2.5 Timeline of theories of AI

'thinking machine' that could reason about "all the subjects in the universe".[28] She even wrote programs for the hypothetical device. However, science at that time was not advanced enough to actually build such computers.

That point would not be reached until the Second World War, when computing power was needed to defend against air raids. The use of fast-moving planes to drop bombs made it impossible for the human operators of anti-aircraft systems to respond quickly enough when relying on their eyesight alone. Instead, their targets' trajectories needed to be calculated mathematically. Research in that field laid the foundations for the modern computer and for another discipline that would emerge

[28] Russell, 2019: 40.

in the 1950s, cybernetics. This work immediately raised questions about automation and human control that are still relevant today.

"The time factor has become so narrow for all operators," a military spokesperson said at the time, "that the human link, which seems to be the only immutable factor in the whole problem and which is particularly fickle, has increasingly become the weakest link in the chain of operations, such that it has become clear that this link must be removed from the sequence."[29]

The development of the computer was given another boost during the war by the British research programme Colossus, which aimed to crack the Nazis' secret communication system known as Enigma. One of the leading lights in this top-secret project at Bletchley Park was Alan Turing, often regarded as the father of both computers and AI. He went on to help develop the first truly modern computer in Manchester in 1948. Two years after that, in 1950, he wrote a paper proposing a thought experiment in the form of an 'imitation game' for a computer pretending to be a human being.[30] This has come to be known as the Turing test. A computer passes if a human is unable to establish that its written answers to their questions were provided by a person or a computer. Variants of this test are still used, for example, to compare AI systems with human abilities such as recognizing images or using language.[31]

Another important theoretical contribution to this field was a paper by psychiatrist and neurologist Warren McCulloch and mathematician Walter Pitts.[32] In this they combined Turing's work on computers with Bertrand Russell's propositional logic and Charles Sherrington's theory of neural synapses. Their most important contribution was that they demonstrated binary modalities (a situation with two options) in various domains and thus developed a common language for neurophysiology, logic and computation. The distinction between 'true and false' in logic was now linked to the 'on or off' state of neurons and the computer values '0 and 1' in Turing machines.[33]

John von Neumann continued to develop the basic concept of a computer with components such as the central processor, memory and input-output devices.[34] Another important founder of AI theory was Norbert Wiener. He coined the term 'cybernetics' in 1948 to describe "the study of control and communication in

[29] Rid, 2016: 37–38.

[30] Turing, 2009 [1950].

[31] There has also been criticism of the use of language in the Turing test. Yann LeCun, a prominent AI scientist, suggested in an interview that there are forms of intelligence that have nothing to do with language (Ford, 2018: 129). Some animals, for example, use less complex language than humans but still form good models of the world and can employ tools.

[32] McCulloch & Pitts, 1943.

[33] In a lecture at Yale in the 1950s, the scientist John von Neumann described the similarity between the computer and the brain as follows: "The nervous pulses can clearly be viewed as (two-valued) markers, in the sense discussed previously: the absence of a pulse represents one value (say, the binary digit 0), and the presence of one represents the other (say, the binary digit 1)." von Neumann, 2012 [1958]: 43.

[34] Freeman & Louçã, 2001: 310.

animals and machines".[35] The key idea was that people, animals and machines could all be understood according to a number of basic principles. The first of these is control: all those entities strive to counter entropy and to control their environment using the principle of 'feedback', which is the "ability to adapt future behaviour to past experience". Through the mechanism of continuous adjustment and feedback, organisms and machines ensure that equilibrium, or homeostasis, is achieved. Wiener used thermostats and servomechanisms as metaphors to explain these processes. Although cybernetics did not last long as a separate scientific field, its core concepts now permeate all manner of disciplines (Box 2.1).[36]

Thanks to such advances, during this period scientists were ready to stop just dreaming and thinking about AI and start actually developing the technology and experimenting with it in the laboratory. The starting gun for this race was fired in 1956.

Key Points: AI Prior to the Lab

- Mythical representations of AI have been around for centuries.
- The most celebrated examples are the ancient Greek stories about Daedalus, Medea, Hephaistos, Prometheus and Pygmalion.
- The mechanization of the world view from the seventeenth century onwards made the construction of all kinds of machines possible. This went hand in hand with speculation about mechanical brains.
- Fictional stories about artificial intelligence appeared from the Industrial Revolution onwards, including *Frankenstein* and *R.U.R.*
- The theoretical foundations for AI were laid when the first computers were built by people like Alan Turing.

Box 2.1: The Homeostat and Electronic Tortoises

In 1948 the Briton Ross Ashby unveiled his 'homeostat', a machine able to hold four electromagnets in a stable position. In that same year *The Herald* wrote of this 'protobrain' that "the clicking brain is cleverer than man's".[37] Another highlight of the cybernetics movement in the 1950s was William Grey Walter's electronic tortoises. These small devices could walk around without bumping into obstacles and locate where in the room their charger was if their battery was weak. Moreover, they also exhibited complex social behaviour as a group. A later example of a cybernetic machine was the John Hopkins Beast, which in the early 1960s was able to trundle through corridors using sonar and a photocell eye to find a charging point.[38]

[35] Wiener, 2019 [1965].

[36] Rid, 2016: 47–52. Famous cyberneticians in various disciplines include the neurophysiologist Warren McCulloch, the physicist Heinz von Foerster, the management theorist Stafford Beer, the philosopher Humberto Maturana, the political scientist Karl Deutsch, the anthropologist Gregory Bateson and the sociologist Talcott Parsons.

[37] Rid, 2016: 53–55.

[38] Moravec, 1988: 7.

2.3 AI in the Lab

2.3.1 The First Wave

As mentioned previously, the beginnings of AI as a discipline can be dated very precisely.[39] After all the myths, speculation and theorizing, artificial intelligence appeared in a lab for the first time in 1956 when a group of scientists made it the subject of a specific event: the Dartmouth Summer Research Project on Artificial Intelligence. This was a six-week brainstorming gathering attended by several of the discipline's founders. The organizers were very optimistic about what they could achieve with this group in a few weeks, as is evident from the proposal they wrote to the Rockefeller Foundation.

> We propose … a 2-month, 10-man study of artificial intelligence … The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.[40]

The proposal was overambitious, and research is still being carried out today in all the areas it mentioned. With this project, however, these scientists formulated a research agenda that launched AI as a discipline.

The summer project was organized by John McCarthy and Marvin Minsky. It was McCarthy who coined the term 'artificial intelligence' in 1956. Minsky was a leading figure in the history of AI and over the years came to be involved in many prominent high-tech projects around the world. The two men also established the Artificial Intelligence Lab at MIT. This was later renamed the MIT Media Lab and is still a centre for the creative use of new technology.[41] Among those present at the summer project were Herbert Simon (Nobel laureate in Economics and winner of the Turing Award, responsible for the idea of 'bounded rationality', amongst other things, and founder of the Carnegie Institute of Technology), John Nash (mathematician, game theorist and another Nobel laureate in Economics) and Arthur Samuel (pioneer of computer games and the man credited with popularizing the term 'machine learning'). These leading scientists were responsible for bringing AI to the lab.

This landmark event heralded a period of great optimism and broad interest in the field of AI, which has come to be known as the first 'AI spring' (or 'wave'). Various programs were developed that could play the board game draughts (checkers), although none was very good yet. The version developed by Samuel did eventually succeed in defeating its human creator, which caused a stir, although he was not

[39] The history of a scientific discipline can be written in several ways. It can focus on the fundamental science, for instance, or on practical inventions and applications. One example is the difference between the development of the natural sciences and the inventions of the Industrial Revolution. In this chapter we combine both perspectives, but the idea of waves in AI is rooted mainly in that of inventions and applications.

[40] Bostrom, 2016: 6.

[41] Broussard, 2019: 69–70.

known as a great player of the game. Wiener wrote in 1964 that, while Samuel was eventually able to beat the program again after some instruction, "the method of its learning was no different in principle from that of the human being who learns to play checkers". He also expected that the same would happen with chess in ten to twenty-five years, and that people would lose interest in both games as a consequence.[42]

Exciting breakthroughs followed when AI systems began focusing on a different category of challenges: logical and conceptual problems. For example, a 'Logic Theory Machine' was built to prove Bertrand Russell's logical theorems. It not only succeeded in proving eighteen of them, it also developed a more elegant proof of one. This was important because, while Samuel was a mediocre draughts player, Bertrand Russell was a leading logician.

The next milestone was the 'General Problem Solver'. This was a program that could, in principle, be applied to solve any problem – hence the name. By translating problems into goals, subgoals, actions and operators, the software could then reason what the right answer was. One example of a problem it solved is the classic logical puzzle of the river crossing.[43]

By the mid-1960s the first students of the AI pioneers were working on programs that could prove geometric theorems and successfully complete intelligence tests, maths problems and calculus exams. So, the discipline was making progress, but its impact outside the lab was very limited. There were some interesting experiments with robots, as in the late 1960s at the Stanford Research Institute; its Shakey the Robot was able to find its way about through reasoning.[44] The American technology company General Electric built impressive robots such as the Beetle and an exoskeleton that enabled humans to lift heavy weights.[45] These robots were not very practical, though.

At the same time, there were grand expectations of AI. In 1965 Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work a man can do".[46] Meanwhile, the British mathematician Irving Jack Good foresaw a machine-induced 'intelligence explosion'. This would also be the last invention of humankind, because machines would now be the most intelligent beings on earth and therefore do all the inventing.[47]

AI caught the imagination of people outside science as well. In 1967 the computer program MacHack VI was made an honorary member of the American Chess Federation, despite having won very few matches.[48] A few years later the film *Colossus: The Forbin Project* was released. In this a computer program is handed control of the US military

[42] Wiener, 1964: 22–24. It would eventually take thirty years for a computer to defeat a chess grandmaster, as we shall see shortly. In any case, people have not lost their interest in these games since sophisticated programs have learned to play them.

[43] Boden, 2018: 10. In this logical problem, three entities all have to cross a river. Only two can cross at the same time. Each entity threatens to harm one of the others, so not every duo can cross together. The problem is: which combinations can be formed to convey everyone to the other side unharmed?

[44] Russell, 2019: 52.

[45] Rid, 2016: 136.

[46] Brynjolfsson & McAfee, 2014: 141.

[47] Rid, 2016: 148. The writer Vernor Vinge would later coin the term 'singularity' for this scenario.

[48] Bakker & Korsten, 2021: 24.

arsenal because it can make better decisions than humans and is unhindered by emotions. After the Soviets reveal a similar project, the two programs start communicating with one another – but in a way that is incomprehensible to their human creators – and subsequently take control of the entire world. Their pre-programmed goal of world peace is achieved, but the price is the freedom of the human race.

This gap between hopeful expectations and harsh reality did not go unnoticed, and from the second half of the 1960s onwards there was increasing criticism of AI research. The philosopher Hubert Dreyfus would remain critical of the potential of AI throughout his life. In 1965 he wrote a study called *AI and Alchemy*, commissioned by the Rand Corporation (the think tank of the American armed forces), in which he concluded that intelligent machines would not be developed any time in the near future. In a 1966 report to the US government, the Automatic Language Processing Advisory Committee concluded that little progress had been made. The National Research Council subsequently phased out its funding of AI. In the United Kingdom, Sir James Lighthill was commissioned in 1973 to conduct a survey of the topic; this brought to light considerable criticism of its failure to achieve the grandiose goals that had been promised. As a result, a lot of research funding was withdrawn in the UK as well.[49]

One problem encountered by many AI systems at this time was the so-called 'combinatorial explosion'. These systems solved problems by exploring all possible options, but they quickly reached the limits of their computing power when dealing with huge numbers of possible combinations. More heuristic approaches, based on rules of thumb, were needed to reduce the number of combinations. However, these did not yet exist. This and other problems – such as the lack of data to feed the systems and the limited capacity of the hardware – meant that progress with AI stalled.

Meanwhile, its practical applications were also proving unreliable. When an AI system was developed during the Cold War, in the 1960s, to translate Russian communications, the results proved less than impressive. One famous example was its translation of "the spirit is willing, but the flesh is weak" as "the vodka is good, but the meat is rotten".[50] During the course of the 1970s, the earlier optimism turned to pessimism. There were too few breakthroughs, so criticism of AI grew, and funding dried up. The first 'AI winter' had set in and put an end to its first wave. Figure 2.6 provides an overview of the emergence of AI as a scientific discipline.

2.3.2 Two Approaches

It is important to note that two distinct approaches to AI gained particular prominence during this first wave. While it is true that there were others as well (we will explain these later), these two still dominate the field to this day. The first is 'rule-based', also known as 'symbolic' or 'logical', AI (along with other names) and emerged in the 1970s in the form of so-called 'expert systems'. Its core principle is

[49] Leung, 2019: 253.

[50] Russell & Norvig, 2021: 21.

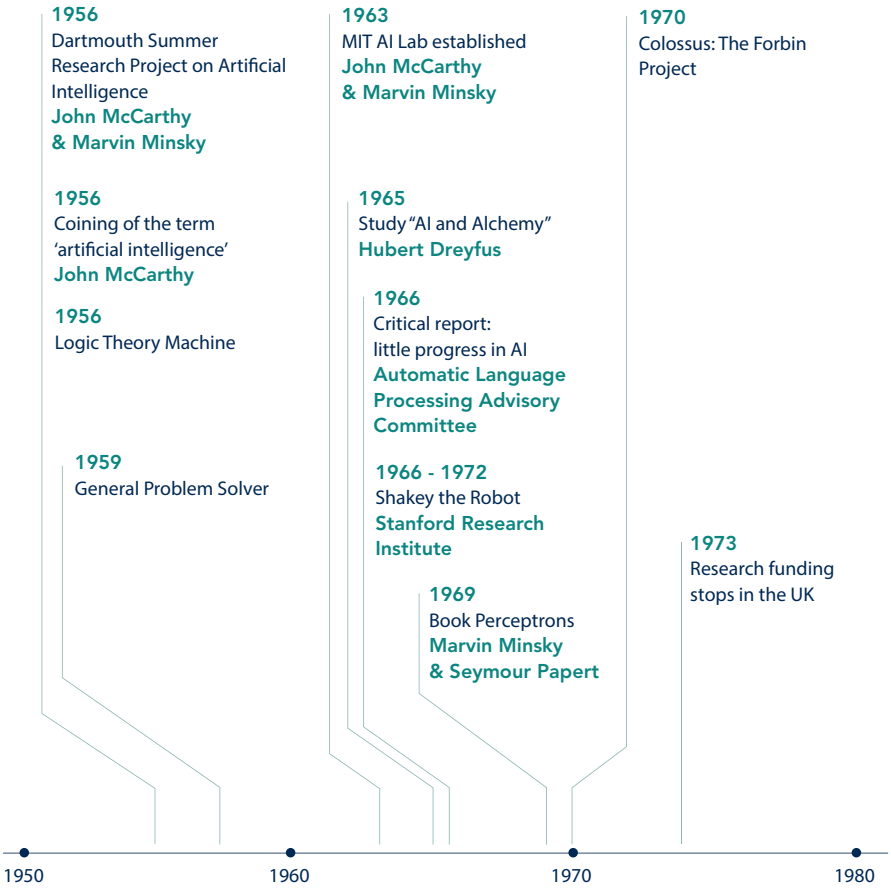


Fig. 2.6 Timeline of the emergence of AI as a discipline (first wave)

that computers learn by encoding logical rules with formulas of the type 'IF X, THEN Y'. The use of logic and rules is also why the term 'symbolic AI' is used, as this approach follows rules that can be expressed in human symbols.

The second approach uses artificial neural networks (ANNs) and is also called 'connectionism'. This includes the deep learning and parallel distributed processing methods that have received a lot of attention in recent years. The central idea here is to simulate the functioning of neurons in the human brain. For this purpose, sets of 'artificial neurons' are built into networks that can receive and send information. These networks are then fed with large amounts of data and try to distil patterns from it. In this case the rules are not drawn up by humans in advance. Most ANNs are based on a principle formulated as early as 1949 by Donald Hebb, a Canadian psychologist, in his book *The Organization of Behaviour*: "Neurons that fire together, wire together".[51] In other words, if two neurons are frequently activated at the same time, they become connected.

[51] Domingos, 2017: 93.

Both approaches to AI were there from the start. While many of the founding fathers at the 1956 summer school followed the rule-based approach, the first artificial neuron was also created around the same time at Cornell University.[52] The difference can be explained as follows. To be able to recognize a cat in a photo, in the first approach a series of 'IF-THEN' rules are established: the presence of certain colours, a given number of limbs, certain facial forms, whiskers, etc., means that it is a cat. With these rules, a program can 'reason' what the data means.

In the second approach, the program might be presented a large number of photos labelled as 'cat' and 'non-cat'. The program distils patterns based on this data, which it then uses to recognize the presence of a cat in subsequent photos. Rather than using labels, another variant of this approach instead presents large numbers of images and then allows the program to come up with its own clustering of cats. In both variants, however, it is not the rules programmed by people, but the patterns identified by the program that determine the outcome.

As already noted, both approaches were explored during the first AI wave. One example of an application of neural networks was Frank Rosenblatt's 'perceptron', an algorithm he invented which learned to recognize letters without these being preprogrammed. This was attracted much media interest in the 1960s. Symbolic AI, however, remained dominant. The Logical Theory Machine and General Problem Solver mentioned earlier were both examples of systems within this strand. For decades it would remain the dominant approach within AI.

The proponents of symbolic AI also expressed much criticism of neural networks. They considered that approach unreliable and of limited use due to its lack of rules. In 1969 Marvin Minsky, an ardent supporter of the symbolic approach, wrote a book called *Perceptrons* with Seymour Papert. This amounted to a painstaking critique of the neural network approach, backed by examples of mathematical proofs of problems it could not solve. To many this appeared to sound the death knell for that approach.[53] Such criticism not only marginalized the position of neural networks, it also contributed towards the onset of the first AI winter.

2.3.3 The Second Wave

In 1982 *Time* magazine named the personal computer its Man of the Year. This coincided with a revival of interest in AI, and the discipline entered a second spring. At the time, the programming language Prolog was used for many logical reasoning systems. In 1982 the Japanese government invested a huge sum in a Prolog-based AI system in the form of the Fifth-Generation Computer Systems Project.[54] This was a far-reaching, ten-year partnership between the government and industry and

[52] Greenfield, 2017: 214.

[53] From an interview with Geoffrey Hinton (Ford, 2018: 83).

[54] Russell, 2019: 271.

was intended to boost the discipline in Japan by establishing a 'parallel computing architecture'. At a time when there was widespread fear of Japanese economic growth, several Western countries quickly followed suit with their own projects.

To keep up with the competition, the US established the Microelectronics and Computer Technology Corporation (MCC), a research consortium. In 1984 MCC's principal scientist, Douglas Lenat, launched a huge project called Cyc. Initiated with the full support of Marvin Minsky, this is still running today and involves collecting vast amounts of human knowledge about how the world works.[55] In 1983 DARPA, the scientific arm of the US Department of Defense, announced a Strategic Computing Initiative (SCI) that would invest one billion dollars in the field over ten years.[56] Both the Japanese and the American research projects took a broad approach to AI, with hardware and human interfaces also playing an important role, for example.[57] In 1983 the United Kingdom announced its response to the Japanese plans in the form of the Alvey Programme.

One important development during this second wave was the emergence in the 1970s of expert systems within symbolic AI. These are a form of rule-based AI where human experts in a particular domain are asked to formulate the rules for a program. One example was MYCIN, a program trained by medical experts to help doctors identify infectious diseases and prescribe appropriate medication. The Dendral project involved the analysis of molecules in organic chemistry. Expert systems were also developed to plan manufacturing processes and solve complex mathematical problems; for example, the Macsyma project. Such systems thus found practical applications outside the lab.

Some were developed in the Netherlands, too, in the 1980s and tested in pilot projects. These addressed themes including the implementation of social security and criminal sentencing policies.[58] In part thanks to specific research programmes and funding provided by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO) and various universities, but also by a number of government departments, the Netherlands was even able to establish an international profile with a relatively large research community in the field of legal knowledge-based systems. An important early facilitator in this respect was JURIX, the Foundation for Legal Knowledge-Based Systems, an organization of 'legal tech' researchers from the Netherlands and Flanders. It has held annual international conferences since 1988; their proceedings – all available online – testify to the rich Dutch and Flemish academic history of research on and development of AI applications in the legal domain.[59] Another prominent platform is the Benelux Association for Artificial Intelligence (Benelux Vereniging voor Kunstmatige Intelligentie, BNVKI), originally formed in the Netherlands in 1981 (as the NVKI) but later

[55] Domingos, 2017: 35.

[56] Leung, 2019: 254.

[57] Russell & Norvig, 2020: 24.

[58] Hage & Verheij, 1999.

[59] www.jurix.nl/proceedings/

connecting scientists from Belgium and Luxembourg as well. The US Office for Technology Assessment has called expert systems "the first real commercial products of about 25 years of AI research"[60] and in 1984 the front page of *The New York Times* reported that they held out "the prospect of computer-aided decisions based on more wisdom than any one person can contain".[61]

Nevertheless, the results of this second wave were ultimately disappointing. The big ambitions of the major national projects were never achieved, either in Japan, the US or Europe. Their poor results were why the US SCI drastically scaled down its funding. Among the problems to limit the potential of these projects were hardware issues. This period culminated with the bankruptcy of several specialized companies in the field in the late 1980s.[62] But the expert systems also had their own problems. They tended to be highly complex, so minor errors in the rules had disastrous consequences for the results and systems could fail when two rules contradicted each other.[63] The Cyc project is still ongoing but has failed to live up to expectations throughout almost four decades of existence.[64] By the late 1980s, therefore, another AI winter had set in: the second wave had run out of momentum.

2.3.4 The Third Wave

In the 1990s, however, AI again began to attract attention and eventually flourish anew. Initially, the logical systems approach had several successes. One of the most iconic of these was the victory of IBM's Deep Blue program over chess grandmaster Garry Kasparov, in 1997. At the time this was considered a fundamental breakthrough. The successor to that program, named Watson, later participated in the US television quiz show *Jeopardy!*, in which contestants have to formulate questions to match given answers. In 2011 Watson defeated the game's reigning human champions. This was seen as proof that AI was approaching mastery of human language, another major breakthrough. Both cases are examples of the use of symbolic AI, in which the lessons of chess masters and answers from previous players of *Jeopardy!* were fed to the programs as rules. At the same time, however, experts were becoming increasingly dissatisfied with this approach.

---

[60] Leung, 2019: 259.

[61] Dreyfus & Dreyfus, 1986: ix.

[62] Leung, 2019: 255.

[63] The idea of expressing the limits of human behaviour and language in rules had been explored earlier by philosophers such as Ludwig Wittgenstein (Wittgenstein, 1984).

[64] According to Ray Kurzweil, a proponent of neural networks, Cyc has actually achieved almost nothing (Ford, 2018: 233). That, however, is an oversimplification. Such projects form the foundations of techniques such as knowledge graphs, which are now important for the functioning of search engines like Google. This also demonstrates why the two approaches are not mutually exclusive and in practice often go hand in hand.

Although both events were huge landmarks in the eyes of the public, in reality the truth was more prosaic. Stuart Russell describes how the foundations of chess algorithms were laid by Claude Shannon in 1950, with further innovations following in the 1960s. Thereafter, these programs improved according to a predictable pattern, in parallel with the growth of computing power. This was easily measurable against the scores recorded by human chess players. The linear pattern predicted that the score of a grandmaster would be achieved in the 1990s – exactly when Deep Blue defeated Kasparov. So that was not so much a breakthrough as a milestone that had been anticipated as part of a predictable pattern.[65] Deep Blue won by brute force, thanks to its superior computing power. Moreover, various chess champions had fed heuristic principles into its software. Instead of the smart computer beating the human, this victory could also be seen as the triumph of a collective comprising a computer program and numerous human players over a single grandmaster.[66] It was man and machine together that were superior to a human opponent.

The computer's victory in *Jeopardy!* is also questionable. It would be incorrect to claim that the program could understand the complex natural language of humans. The game has a very formalized question-and-answer design, and many of the questions can be found on a typical Wikipedia page. This makes them relatively easy to answer for a program that can rapidly search mountains of information for keywords; that does not require an in-depth understanding of language.

While these logical systems only began to attract attention in the 1990s, other forms of AI had been making progress for far longer and the momentum eventually shifted towards the neural network approach. This trend had already begun in the mid-1980s when fundamental research into the so-called 'backpropagation algorithm' (in which multiple layers of neural networks are trained) improved the process of pattern recognition. At about the same time the US Department of Defense recognized that its funding programme had been unfairly neglecting the neural networks approach. Under the banner of 'parallel distributed processing', neural networks returned to centre stage in 1986. In a book published the previous year, John Haugeland had introduced the term GOFAI ('good old-fashioned AI') – a phrase which has since become a pejorative term for symbolic AI. In the same period Judea Pearl began applying probability theory rather than logical reasoning to AI.

Breakthroughs below the radar were thus undermining the dominant rule-based approach. A paper on backpropagation was rejected for a leading AI conference in the early 1980s and, according to Yann LeCun, researchers at the time even used code words to mask the fact that they were working with neural networks.[67] It took time for the importance of this new approach to become recognized. For example, Jeff Hawkins said in 2004 that AI had fewer skills than a mouse when it came to image recognition.[68]

[65] Russell, 2019: 62–63.

[66] Ihde, 2010.

[67] From an interview with Yann LeCun (Ford, 2018: 122).

[68] Tegmark, 2017: 79.

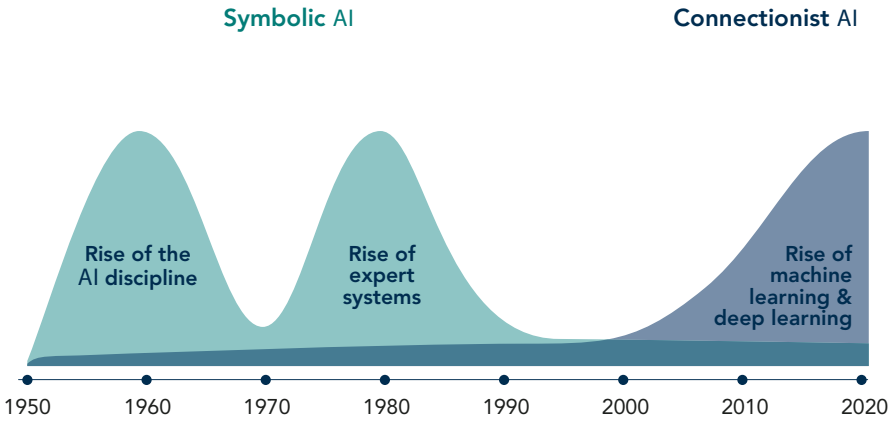


Fig. 2.7 The transition from a symbolic to a connectionist AI

At that time, it was thought it would take another century before a computer could beat a human in the Asian game go, which has many more combinations of moves than chess.[69] In fact, Google's AlphaGo program defeated world champion Lee Sedol in 2016. This was made possible thanks to recent breakthroughs in the approach to neural networks, in which researchers such as Yann LeCun and Andrew Ng played an important role. But it is Geoffrey Hinton who is often seen as the father of those advances. Together with David Rumelhart and Ronald Williams, he had already popularized the use of the backpropagation algorithm in a paper published in *Nature* in 1986. That algorithm traces the contribution made by the output layer back to hidden layers behind it, where individual units are identified that need to be modified to make the algorithm work more effectively. For a long time, the 'backprop' had only a single hidden layer, but more have recently been distinguished. Backpropagation thus addresses a central problem of ANNs: the representation of hierarchy. Relationships can now be distinguished at different levels and the success factors of the algorithm are also determined at all levels (called 'credit assignment').[70] Such neural networks have since been used, for instance, to simulate the price of shares on the stock exchange. Figure 2.7 shows the historical development of the two approaches to AI.

In 1989 Yann LeCun applied backprop to train neural networks to recognize handwritten postcodes. He used convolutional neural networks (CNNs), where complex images are broken down into smaller parts to make image recognition

[69] Tonin, 2019: 1.

[70] In the book *Perceptrons*, which was highly critical of the neural networks approach, Minsky and Papert demonstrated that it was unable to solve the problem of the 'exclusive OR' (XOR). But Rumelhart, Hinton and Williams showed that backpropagation could learn XOR.

Box 2.2: Three Forms of Machine Learning
ML can be subdivided into three different forms: supervised, unsupervised and reinforcement learning. In supervised learning, a program is fed data with labels as in our earlier example of 'cat' versus 'non-cat'. The algorithm is trained on that input and then tested to see if it can correctly apply the labels to new data.

Unsupervised learning has no training step and so the algorithm needs to search for patterns within the data by itself. It is fed large amounts of unlabelled data, in which it starts to recognize patterns of its own accord. The starting point here is that clusters of characteristics in the data will also form clusters in the future. Supervised learning is ideal when it is clear what is being searched for. If the researchers themselves are not yet sure what patterns are hidden within data and are curious to know what they are, then unsupervised learning is the more appropriate method.

more efficient. This was another important contribution to contemporary AI programs.[71]

In another paper, written in 2012, Hinton introduced the idea of 'dropout', which addresses the specific problem of 'overfitting' in neural network training. That occurs when a model focuses so strongly on training with existing data that it cannot effectively process new information. Hinton's work gave an enormous boost to the applicability of neural networks in the field of machine learning. The use of multiple layers in the training process is why it is called 'deep' learning; each layer provides a more complex representation of the input based on the previous one. For example, while the first layer may be able to identify corners and dots, the second one can distinguish parts of a face such as the tip of a nose or the iris of an eye. The third layer can recognize whole noses and eyes, and so it goes on until you reach a layer that recognizes the face of an individual person (Box 2.2).[72]

The third form is applicable in other contexts, such as playing a game. Here it is not about giving a right or wrong answer or clustering data, but about strategies that can ultimately lead to winning or losing. In these cases, the reinforcement learning approach is more suitable. The algorithm is trained by rewarding it for following certain strategies. In recent years reinforcement learning has been applied to various classic computer games such as Pacman and the Atari portfolio, as well as to 'normal' card games and poker. The algorithm is given the goal of optimizing the value of the score and then correlates all kinds of actions with that score to develop an optimum strategy.

In 2012 Hinton's team won an international competition in the field of 'computer vision' – image processing using AI. They achieved a margin of error of 16%,

[71] Marcus & Davi, 2019: 52.
[72] Domingos, 2017: 117.

whereas no team before them had ever managed less than 25%. A few years earlier the same team had been successful in using neural networks for speech recognition after a demonstration by two students in Toronto. But the gains in computer vision in 2012 were the real revelation for many researchers.[73] Deep learning proliferated, and in 2017 almost all the teams in the competition could boast margins of error lower than 5% – comparable with human scores. That improvement continues to this day. The application of DL has since gained momentum, with the scientific breakthroughs using neural networks prompting an explosion of activity in this approach to AI. We are currently at the height of this latest AI summer. In the next chapter we look in more detail at the developments that has set in motion outside the lab: in the market and in wider society.

It is clear that the rapid expansion of AI in recent years has its origins in fundamental scientific research. Big companies like Google have subsequently rushed to hire talented researchers in this field, but it is scientists at universities who have been responsible for the most important breakthroughs.

In addition to these academic milestones, two other factors underlie the recent rise and application of AI. The first is the growth in processing power, as encapsulated in Moore's Law. This pattern, that the number of transistors on a chip roughly doubles every two years, has been observed consistently in the computer industry for decades. It means that more and more computing power is becoming available while prices continue to fall. Hence the fact that the smartphones of today surpass the computing power of the very best computers of only a few decades ago. We noted earlier how the first 'AI winter' was caused in part by the combinatorial explosion. The increase in computing power provided the solution to this problem. A further leap in that power came from the chip industry, using graphic processing units (GPUs) rather than the classic central processing units (CPUs). GPUs were originally developed for complex graphics in the gaming industry but were subsequently found to enable many more parallel calculations in AI systems as well.[74] Since 2015, tensor processing units (TPUs) specifically designed for ML applications have also come into use.

The other factor that has contributed to the current AI wave is the increase in the amount of data. This is closely linked to the rise of the internet. In the past algorithms could only be applied to a limited range of data sources. In recent decades, however, as people have started to use the internet more and more, and directly and indirectly to generate a lot more digital information, the amount of data available for AI systems to analyse has increased significantly.

The 'digital breadcrumbs' we leave behind on the internet are now food for training AI algorithms. But we are helping with this training in other ways, too. By tagging personal names in photos on Facebook, for example, people provide algorithms with labels that can be used to train facial recognition software. One specific dataset that is very important for this kind of training is ImageNet, an open database of

[73] From an interview with Geoffrey Hinton (Ford, 2018: 77).

[74] Kelly, 2017: 38.

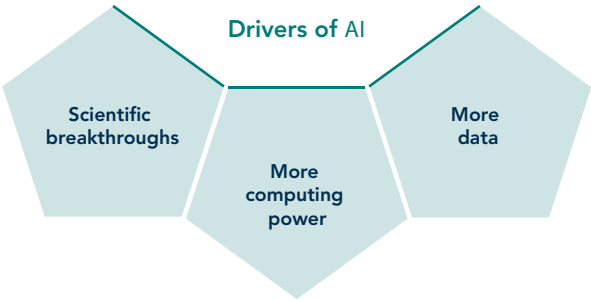


Fig. 2.8 Three drivers of progress in AI

more than 14 million hand-labelled images. The 'internet of things' (the growing number of sensors and connections in the physical environment) is also contributing to the growth in data.

The triad of scientific breakthroughs, greater computing power and more data has allowed AI to take off in a big way recently (see Fig. 2.8). As mentioned, this expansion has been driven mostly by the application of machine learning as part of the neural network approach, and within ML by the development of deep learning.

Key Points: AI in the Lab

- In the lab AI has ridden three waves of development. Between these were two 'winters' when scientific progress ground to a halt, hardware capacity was inadequate, and expectations were not met.
- The first wave began with the Dartmouth Summer Research Project in 1956. At that time AI was used mainly for games such as draughts, in early robots and to solve mathematical problems. Two further waves, dominated by progress in symbolic AI and then neural networks, would follow.
- The second wave began in the 1980s, driven in part by the international competition between Japan, the US and Europe. This produced expert systems, the first major commercial applications of AI.
- The third wave began in the 1990s with major achievements in symbolic AI, but only properly gained momentum some years later due to advances in the field of machine learning and its subfield of deep learning. The scientific breakthroughs in this area, together with increases in computing power and data volumes, are the driving force behind this wave, which continues to this day.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Bakker, S., & Korsten, P. (2021). *Artificiële Intelligentie Als Een general purpose technology: Strategische Belangen Van Verantwoorde Inzet In Historisch Perspectief*

(WRR Working Paper nr. 41). Wetenschappelijke Raad voor het Regeringsbeleid. Available at: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Boden, M. (2018). *Artificial Intelligence: A very short introduction*. Oxford University Press.

Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

DenkWerk. (2018). *Artificial Intelligence in Nederland: Zelf Aan Het Stuur*. Available at: https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf

Dennett, D. (2019). What can we do? In J. Brockman (red.), *Possible minds: Twenty-five ways of looking at AI* (pp. 41–53). Penguin.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to develop and use AI in a responsible way*. Springer.

Domingos, P. (2017). *The master algorithm: How the Quest for the ultimate learning machine will remake our world*. Penguin Random House.

Dreyfus, H., & Dreyfus, S. (1986). *Mind over Machine*. The Free Press.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

Freeman, C., & Louçã, F. (2001). *As time Goes By: From the industrial revolutions to the information revolution*. Oxford University Press.

Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.

Hage, J., & Verheij, B. (1999). Rechtsinformatica: De Stand Van Zaken In De Wetenschap. In A. Oskamp and A. Lodder (reds.), *Informatietechnologie voor juristen. Handboek voor de jurist in de 21e eeuw* (pp. 65–92). Kluwer.

High-Level Expert Group on Artificial Intelligence. (2019). *A definition of AI: Main capabilities and scientific disciplines*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

Ihde, D. (2010). *Embodied technics*. Automatic Press/vip.

Kelly, K. (2017). *The Inevitable: Understanding the 12 technological forces that will shape our future*. Penguin.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Marcus, G., & Davi, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. Vintage.

Mayor, A. (2018). *Gods and Robots: Myths, machines, and ancient dreams of technology*. Princeton University Press.

McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115–133.

Moravec, H. (1988). *Mind Children: The future of robot and human intelligence*. Harvard University Press.

Nilsson, N. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.

Rid, T. (2016). *Rise of the machines: A cybernetic history*. WW Norton & Company.

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A modern approach* (4th ed.). Pearson.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A modern approach*. Pearson.

Tegmark, M. (2017). *Life 3.0: Being Human in the age of Artificial Intelligence*. Penguin.

Tonin, M. (2019). Artificial Intelligence: Implications for NATO's Armed Forces. *149 stctts 19 E rev. 1 fin.*

Turing, A. (2009 [1950]). Computing machinery and Intelligence. In R. Epstein, G. Roberts, and G. Beber (reds.), *Parsing the turing test*. Springer.

von Neumann, J. (2012 [1958]). *The Computer and the Brain.* Yale University Press.
Wiener, N. (1964). *God and Golem, Inc.: A comment on certain points where cybernetics impinges on religion*. MIT Press.
Wiener (2019 [1965]). Cybernetics: Or control and communication in the animal and the machine. MIT Press.
Wittgenstein, L. (1984). *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Suhrkamp.
Zarkadakis, G. (2015). *In our own image: Will artificial intelligence save or destroy us?* Ebury Publishing.

Chapter 3
AI Is Leaving the Lab and Entering Society

Since the birth of AI in 1956, various applications of the technology have left the lab and spread through society. Expert systems have been in widespread use for decades and the first neural networks entered the financial sector some time ago. Thus far, however, the impact has been modest due to the limited scope for utilizing such forms of AI.

That picture is now changing. As AI has gathered momentum, many applications have started to appear throughout society and the economy. As explained in the previous chapter, AI's acceleration is driven by scientific advances coupled with increasing computational power and data availability. This chapter considers how AI is making its presence felt in society. We begin by identifying a set of indicators that demonstrate the momentum it now has – ranging from publications and patents to investment and employment. We then discuss the various types of AI currently in use, including image recognition, speech recognition and robotics. That analysis reveals just how widely AI applications are now distributed in many countries. We go on to describe how, largely as a result of AI's entry into society, the technology has become the subject of public debate. Finally, we look at the future of the laboratory. AI may have moved from lab to society, but it remains a technology heavily reliant on fundamental research.

3.1 Momentum from Lab to Society

3.1.1 Scientific Activity

AI's definitive and wide-ranging transition from the research laboratory into everyday settings started gathering momentum in about 2010. That movement was preceded by an upsurge of scientific activity. The World Intellectual Property Organization has released a study showing a considerable increase in the number of

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_3

AI-related publications over the past 20 years: up an average of 8% annually between 1996 and 2001, rising to 18% between 2002 and 2007.[1] After 2015 annual growth surged again to 23%, and in 2018 AI-related papers accounted for 2–3% of all published articles worldwide[2] –almost three times the proportion in the late 1990s.

3.1.2 Practical Potential

In that same period, the deep learning-based advances in speech and image recognition referred to in the previous chapter opened the door to a wide variety of potential practical applications. We also see a marked rise in the number of AI-related patents granted: the average annual increase was 8% between 2006 and 2011, but 28% in the years 2012–2017.[3] AI's share of all new patents jumped in the last two of those years from less than 1.5% to nearly 2.5%.[4] Half of all AI inventions ever patented date from 2013 to 2018.[5] In short, the surge in academic activity since the early 2010s has been accompanied by a wave of AI patents.

Looking more closely at the patent grants, we see that growth has been greatest in the domain of machine learning. Some 40% of all AI patents refer to that technology. Within this domain, deep learning has been the fastest-growing discipline with patent grants increasing by 175% between 2013 and 2016.[6] Zooming in on the fields of application discussed in the next section, image processing or computer vision is the most prominent, accounting for about half of all AI patents in the period.[7] In other words, a great deal of innovation is taking place in AI. The increasing importance of practical applications is also apparent from the software development data: since 2014 the amount of AI-related open-source software (OSS) has increased at three times the pace of other forms of OSS.[8]

3.1.3 Rising Investment: AI Is Becoming a Business

The growth in patent grants reflects the business community's increasing interest in AI. From about 2010 onwards, companies such as Google, IBM and Microsoft began working with neural networks for speech recognition. Google has been using

[1] World Intellectual Property Organization, 2019.

[2] Baruffaldi et al., 2020; Perrault et al., 2019.

[3] World Intellectual Property Organization, 2019.

[4] Baruffaldi et al., 2020.

[5] World Intellectual Property Organization, 2019: 39.

[6] World Intellectual Property Organization, 2019; Baruffaldi et al., 2020.

[7] World Intellectual Property Organization, 2019; Baruffaldi et al., 2020.

[8] Baruffaldi et al., 2020: 32.

these networks on Android smartphones since 2012. The use of computer vision by big technology companies has been on a similar upward trajectory. In 2014 Google acquired the British company DeepMind, a global leader in AI research with many 'firsts' to its name, including the first AI go victory over a human champion, Lee Sedol.

Enhanced AI language capability has been deployed in Google Translate since 2016,[9] and in 2017 Intel spent €14 billion to acquire the Israeli company Mobileye, a specialist in driver assistance and autonomous driving systems. Facebook, Amazon, Apple, Microsoft and other hardware and software companies have also been acquiring AI start-ups in recent years to boost their capabilities in this field. Whereas barely ten such acquisitions were registered in 2010, there were more than 240 in 2019.[10]

Major tech corporations have also been recruiting prominent AI scientists. Geoffrey Hinton joined Google, Yann LeCun went to Facebook and Andrew Ng has worked for both Google and the Chinese company Baidu. In their public statements, the executives heading up such companies have explicitly stated their interest in AI. In a 2016 letter to shareholders, Amazon's Jeff Bezos wrote that machine learning was crucial to improving core operations. The following year Google CEO Sundar Pichai delivered a speech announcing that the firm was moving from a 'mobile-first world' to an 'AI-first world'.[11] Similarly, Microsoft's Satya Nadella wrote to company personnel in 2018 setting out organizational changes linked to the reallocation of resources to the cloud (online storage) and AI.[12] Chinese tech giants such as Baidu, Tencent and Alibaba have also been saying for years – in some cases before their American counterparts[13] – that AI is central to their business strategies. For example, the first research centre Alibaba ever opened outside China was an AI-focused facility in Singapore.

Commercial interest is not confined to the 'big tech' sector. The business landscape includes a wide range of young companies with AI at the heart of their operations. They include China's ByteDance and Face++, US firms Airbnb, Shazam and Tesla, Israel's Waze and the Europe-based Spotify and Booking.com. It is the European Commission's stated ambition that three out of every four companies should be using AI by 2030.[14]

Global investment in AI start-ups has been increasing steadily for some years. Researchers at Stanford University estimated total private investment in this segment at US$40 billion dollars in 2018, up from $1.3 billion in 2010. During that period, investment increased by an average of nearly 50% a year.[15] Although

[9] From an interview with Yoshua Bengio (Ford, 2018: 27–28).

[10] CB Insights, 24 June 2021.

[11] Agrawal et al., 2018: 179.

[12] Leung, 2019: 248.

[13] CB Insights, 26 April 2018.

[14] European Commission, 9 March 2021a.

[15] Baruffaldi et al., 2020: 82.

quantitative investment estimates vary, depending on the definitions and methodologies used, the upward trend is unmistakable. Like the total amount invested, the number of investments also increased: from 200 in 2011 to 1400 in 2017. Based on those trends, the OECD has concluded that investors are recognizing the potential of AI.[16]

Taking a broader view, Stanford University estimates that total investment in AI businesses was nearly US$70 billion dollars in 2020[17] – five times as much as in 2015. Between 2015 and 2020, therefore, AI firms around the world received a huge injection of funds. In recent years 60% of all AI investment has gone into machine learning.[18] For a long time the bulk of that was directed towards the development of autonomous vehicles, in line with the focus on computer vision referred to above.[19] In 2018 they accounted for 30% of the capital invested in AI start-ups, with the number of businesses testing such vehicles in California increasing sevenfold. In 2020, however, the COVID-19 pandemic brought about a realignment, with the healthcare and pharmaceutical sectors now attracting the lion's share of investment.[20]

3.1.4 Economic and Employment Impact

Various consultancy firms have made predictions about the implications of AI's definitive entry into society. They envisage that, because of its generic nature, the technology will influence almost all business sectors and have considerable economic impact. In 2017, for instance, PwC forecast that AI could be contributing as much as US$15.7 trillion to the world economy by 2030.[21] The same report identified healthcare, automotive manufacturing, financial services, transport and logistics, ICT, media and retail as the sectors where the impact would be greatest. Deloitte also foresees AI's commercial importance increasing rapidly and suggests that the window of opportunity for a business to gain a competitive advantage from it is very narrow. Firms need to involve themselves quickly if they do not want to miss the boat.[22] In a 2018 report McKinsey predicted that 70% of the world's businesses would make use of AI and that the technology had the potential to boost global gross domestic product (GDP) by 1.2% a year.[23] More recently, McKinsey analysed AI's economic potential for a number of countries identified as Europe's 'digital

[16] Baruffaldi et al., 2020: 1.
[17] Zhang et al., 2021: 93.
[18] Tonin, 2019.
[19] Baruffaldi et al., 2020: 90; OECD, 2018: 3.
[20] Zhang et al., 2021: 97.
[21] Rao & Verweij, 2017.
[22] Loucks et al., 2019.
[23] Bughin et al., 2018.

leaders'. If they succeed in adopting AI and pursue sound investment strategies, the analysts say, GDP growth could increase by 1.4% a year.[24]

Meanwhile, US researchers have demonstrated that AI's entry into society can also boost employment. The number of available AI-related positions went up from 0.3% of all US vacancies in 2012 to 0.8% in 2019. Having stood at 0.26% in 2010, the proportion of jobs accounted for by AI-related roles reached 1.32% in 2019.[25] Moreover, AI has become one of the most popular fields of study for postgraduate researchers in computer science in North America. In 2010 the proportion of PhD graduates in AI taking jobs in industry was about the same as the percentage going into academia. Since then, though, the balance has shifted: in 2019 more than half went on to take industry jobs, while fewer than a quarter followed academic careers.[26] According to technology expert Tim O'Reilly, 'data scientist' is now the most coveted job title in Silicon Valley. The McKinsey Global Institute estimates that, in 2018, the US already had between 140,000 and 190,000 fewer machine learning experts than it needed.[27]

3.1.5 Governments Are Also Focusing on AI

It is not only through commercial activities and private-sector applications that AI is entering society; a wide variety of public organizations are also contributing towards the transition. Police services use the technology to investigate and fight crime, social security agencies use it for fraud detection and various AI-based control initiatives were launched during the COVID-19 pandemic. Although no global historical overview is available, the European Commission estimates that roughly 230 public-sector AI applications were in use in 2019.[28] It seems very likely that the actual number was higher; in the Netherlands alone, 74 public-service projects were making use of AI that year.[29]

Further evidence of AI's societal traction is provided by the growing number of national AI strategies being produced. Once it became clear that AI had reached the point where various practical applications were in the offing and the business community was investing heavily, many governments began developing strategies to reap the associated benefits. First came the Pan-Canadian Artificial Intelligence Strategy in March 2017, in which the Ottawa government announced plans to invest C$125 million in AI. Singapore, Japan and the United Arab Emirates followed suit later that year. China then published the New Generation Artificial Intelligence

[24] McKinsey & Company, 2020.

[25] Perrault et al., 2019.

[26] Zhang et al., 2021: 118.

[27] Domingos, 2017: 9.

[28] Misuraca & Van Noordt, 2020.

[29] Van Veenstra et al., 2019.

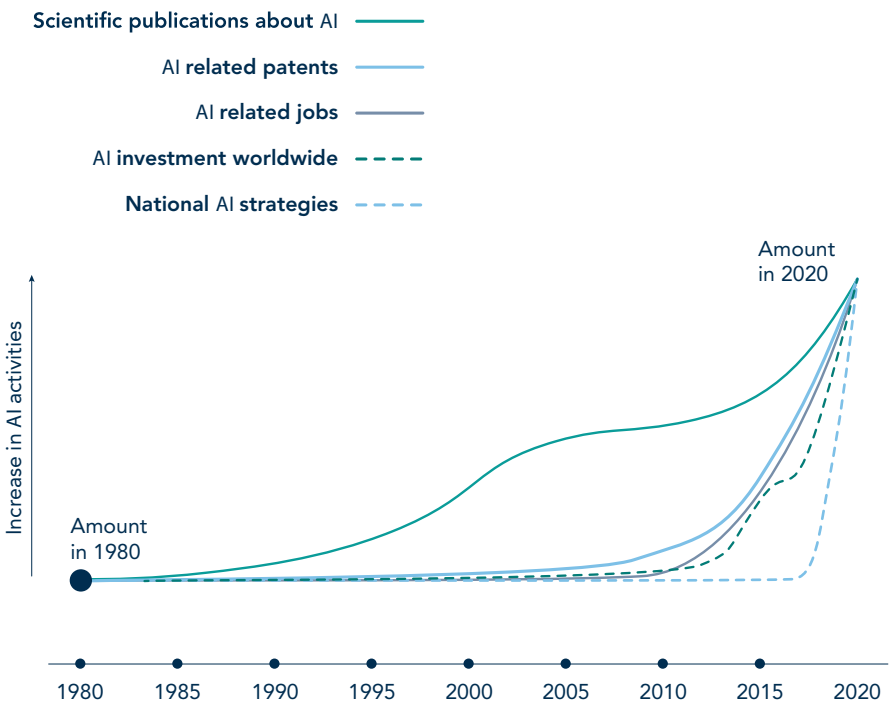


Fig. 3.1 AI gathers momentum outside the lab

Development Plan, setting out its ambition to be the absolute global leader in AI by 2030. Soon afterwards strategies were presented by Finland, the US, France, the UK, Germany and other countries. As part of its commitment to 'a Europe that is ready for the digital age', the EC also began a number of AI-related programmes accompanied by a European Action Plan for AI[30] and a data strategy.[31] Since then, dozens of nations have produced action plans for utilizing AI, including less obvious countries such as Kenya, India and Mexico.[32] The flow of publications hit a peak in 2019 when twenty national AI strategies appeared; a total of about sixty are now in circulation. There is also one international AI strategy: the EU's Co-ordinated Plan on Artificial Intelligence (2018).[33]

Following the acceleration of AI development from around 2000 onwards, it is apparent from the increasing number of patent grants, the growing level of private investment, the appearance of new business models, the growth of AI-related employment and the publication of national strategies that we have reached a new chapter in the history of AI: the technology is entering society. Figure 3.1 illustrates

[30] European Commission, 2021b [2018].

[31] European Commission, 2020.

[32] Holoniq, 9 April 2020; Future of Life Institute, undated (a); Van Roy et al., 2021.

[33] European Commission, 2021b [2018].

this progress using the indicators referred to above. It is therefore pertinent to ask what mechanisms are at work here and what forms is AI taking in society. We address those questions in the next section.

Key Points – Momentum from Lab to Society
- Since the 2010s, AI's migration from lab to society has gained momentum. Advances made in the laboratory provide a springboard for practical application of the technology.
- One reflection of AI's new practical potential is an increasing number of patent grants. Half of all patented AI inventions were registered between 2013 and 2018.
- Big tech companies are openly committed to AI, new businesses are springing up with AI at the heart of their operations and private investment in AI is increasing substantially throughout the world.
- Because of its generic nature, AI is expected to have a major economic impact. Demand for AI experts is growing in the jobs market, while more and more PhD graduates in the subject are finding employment in the commercial sector.
- Governments are also turning their attention to this theme: more than sixty countries have now developed national AI strategies.

3.2 The Practical Application of AI

AI has thus made the transition from the lab to society. As a result, we nowadays encounter all kinds of applications of the technology in our everyday lives: chatbots, smart cameras, translation apps, recommendation systems, risk analyses, driving systems and so on. In practice, AI takes many different forms which may be divided into several broad groups based on the type of task performed. Within the discipline, various classification systems are used. For the purpose of this overview, we distinguish five types of AI: applications for predictive analysis (machine learning), for image processing (computer vision), for language (natural language processing) and speech (speech recognition) and for the performance of physical tasks (robotics). All of these are already visible around us. Figure 3.2 provides an overview of the five types, which are considered individually below.

3.2.1 Machine Learning

The most common type of AI is machine learning. That can be slightly confusing, because the same term is also used for the form of technology currently dominant within AI. In this case, however, 'machine learning' refers to a particular type of

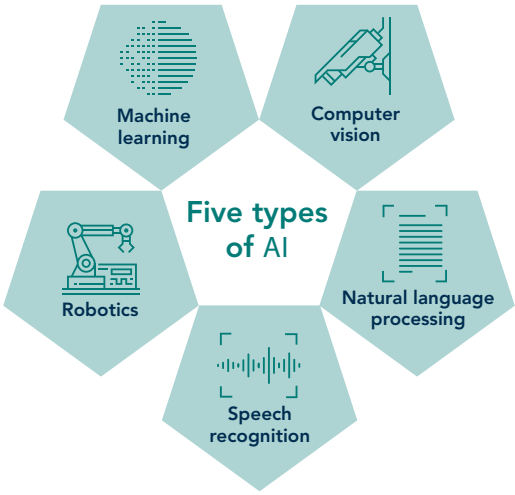


Fig. 3.2 Five types of AI in practical use

application for predictive or advanced analytics, which is used to identify patterns in datasets as a basis for making predictions. Although machine learning technology can be used in other types of AI as well, this form is characterized by prediction being the primary task. It could thus also be referred to as 'predictive systems'.

The ability to use data to make better-informed estimates about the future has huge potential value in many different contexts. The organization of energy supplies is a good example. Google's DeepMind has developed an AI system that uses weather forecasts and turbine data to predict the inflow of energy from wind farms 36 h in advance.[34] Optimum use can then be made of wind power, despite the variability of the elements.

Because risk forecasting has always played an important role in financial services, machine learning is now widely used in that sector. Examples include AI-based credit rating, where a person's creditworthiness is predicted based on their credit history and personal data.

Machine learning is also used for fraud prevention. For instance, Mastercard uses a system called Decision Intelligence to detect abnormal, potentially fraudulent activity by analysing transaction patterns.[35] There are also AI applications for customers, including systems that predict financial trends to help inform investment decisions.[36] Like banks and insurers, local authorities and police forces are looking into machine learning in the fight against fraud and other forms of crime. The UK's Department of Work and Pensions uses AI to assess benefit claims and estimate the probability of fraud, for example.[37] In the Netherlands, the Ministry of Social Affairs and Employment and the country's local authorities introduced System Risk

[34] DeepMind, 26 February 2019.

[35] Mastercard, undated.

[36] ING's Smart Working Capital Assistant and Katana Lens.

[37] Gov.uk, undated; Marr, 29 October 2018.

Indication (SyRI) to tackle benefit fraud. This approach proved controversial, however, and was ultimately deemed unlawful by the courts. Meanwhile, some Dutch local authorities use AI to predict which of their residents are liable to fall into debt and may therefore require assistance.

AI is also deployed in police work, in the form of prediction systems. Many examples of 'predictive policing' can be found in the US, where AI is used, for example, to assess the risk of reoffending. Forces in other countries are investigating the scope for using machine learning to provide intelligence. For instance, the Dutch police have a Crime Anticipation System (CAS) that detects patterns of criminality and predicts where and when robberies are most likely to occur. Based on this output, surveillance and preventive activities can be tailored to the anticipated risk. Almost all of the 168 police districts in the Netherlands are currently using a version of CAS.[38] Furthermore, the Dutch police experiment with machine learning to predict which cold cases have the highest chance of a breakthrough and are therefore worthy of further investigation.[39]

AI's accurate predictive capabilities can be valuable in other sectors as well. Some supermarket chains have announced plans to experiment with dynamic pricing as a tool for minimizing waste and maximizing income. They could also use machine learning for product-range optimization or automated discounting. This would involve an algorithm analysing data on product shelf-life, outlet location, weather conditions and historical sales patterns to make predictions.[40]

In the media industry, machine learning helps tailor products and services to consumers' wishes. The most familiar examples are platform services like Netflix, YouTube and Spotify, which use AI to make relevant recommendations based on users' previous choices. Predictive technologies of this kind are known as 'recommender systems'. Machine learning-aided personalization has become an important pillar of e-commerce as well. Online retailers like Amazon, Alibaba and Zalando use AI to compile user profiles and adapt their marketing accordingly.

Similarly, advertising can be aligned with the interests and sensibilities of individual users. Known as microtargeting, this technique lends itself not only to commercial applications but also to political ends. Political microtargeting made waves around the world when it became known that the company Cambridge Analytica had used Facebook data to disseminate personalised advertising during the US presidential election and the UK's Brexit referendum in 2016 (see Box 3.1). However, microtargeting is a widespread phenomenon that has been occurring in many countries for quite some time.[41] Investigative journalists have found that almost all political parties in the Netherlands engage in bespoke online messaging.[42]

---

[38] Waardenburg et al., 2020: 70.

[39] Politie, 23 May 2018.

[40] Albert Heijn, 20 May 2019.

[41] Prins, 2017; Zuiderveen Borgesius et al., 2018.

[42] Davidson and Delhaas, 22 April 2020. The planned Political Parties Act will regulate such activities (Dutch national government, 26 June 2019).

Box 3.1: Cambridge Analytica and Microtargeting
Microtargeting is directing particular messages at particular people. AI is used for 'psychographic profiling', so that the content shown to an individual is tailored to their personal profile, thus (supposedly) maximizing its effectiveness. The best-known examples of the dark side of this technique involve the data-mining firm Cambridge Analytica, which was closely associated with the 2016 Trump and pro-Brexit campaigns. Machine learning was applied to huge volumes of data on people's online behaviour to build an understanding of public thinking, thus enabling targeted messaging on Facebook and other platforms to influence the way individuals voted. Cambridge Analytica has since ceased trading, but predictive systems of a new type are now being developed: 'multi-agent artificial intelligence' (MAAI). It is claimed that these can predict behaviour even more accurately, opening the way for more precise influence by putting targeting strategies to the test in simulated communities.[43]

3.2.2 Computer Vision

Our second main type of AI relates to image recognition, also known as computer vision. This is about automating the observation, analysis and interpretation of visual information. That may be in the form of photographs, videos or live input from the physical world. Its development has been accelerated by the increasing availability of digital imagery. Social media and smartphones have facilitated a veritable explosion of images, some publicly available, which can be used to train computer vision algorithms. Indeed, we now communicate increasingly through images – 'If there isn't a pic, it didn't happen!' Since Instagram was launched, users have uploaded about 50 billion photos, while 350 million photos a day are posted on Facebook and 500 h of video material are added to YouTube every minute.[44]

One of the best-known applications of computer vision is facial recognition. Moving beyond the mere detection of a face in an image by a computer, this entails the computer actually identifying whose it is. Camera input is analysed and features such as chin proportions, eye separation and cheek roundness are measured with millimetre-level accuracy. The computer the translates this data into a code representing the unique characteristics of a face, enabling it to be recognized when next encountered.

Facial recognition software is built into some smartphones, enabling users to unlock their phones simply by looking into the camera – in other words, to use their face like a password. Various apps use the technique in a similar way so that, for example, a PIN is not needed to authorize a payment.

[43] Hern, 30 July 2019; Lewis & Hilder, 23 March 2018; Lawton, 2 October 2019.
[44] Apple, 24 June 2020; Smith, 18 September 2013; Wojciki, 14 February 2020.

China leads the world in the state use of facial recognition. The technology is widely deployed by the police there and for the surveillance of urban public spaces.[45] Many US government organizations, including the police, investigative agencies and border forces, use facial recognition as well.[46] Although currently controversial in Europe, most countries here are experimenting with the technology for use in airports, stadiums, schools and casinos as well as law enforcement.[47] According to AlgorithmWatch, a research and lobby organization concerned with algorithms and AI, facial recognition is used by police forces in at least eleven European nations.[48] However, the EU plans to introduce strict controls on its deployment in public places; the recently proposed Artificial Intelligence Act would prohibit such use except where strong grounds exist in its favour.[49]

Facial recognition is by no means the only application of computer vision. It is also crucial for self-driving vehicles. Autonomous and semi-autonomous cars currently under development by Tesla, BMW, Volvo, Audi and Uber are equipped with multiple cameras that scan the surrounding space and recognize objects, road markings, traffic signs and traffic lights. Other applications of computer vision are intended primarily for monitoring of the physical environment. Examples include the detailed inspection of roads, bridges and machines with a view to facilitating prompt maintenance and the automated detection of vehicles and objects. In Amsterdam, for instance, cameras read the number plates of vehicles entering the city's low emissions zone and the details of any not entitled to be there are sent to the agency responsible for issuing and collecting traffic fines. During the COVID-19 pandemic, computer vision has been utilized in various countries to scan public spaces for people who might not be respecting the rules on social distancing.[50]

As well as lending itself to applications in public space, computer vision has great potential for the agricultural and livestock sectors and the food industry. It can be used to monitor and harvest crops, for example, and also play an important role in so-called 'precision agriculture'.[51] Computer vision is suitable for animal-welfare applications, too, with cameras used to monitor behaviour.[52] Dutch start-up OneThird has developed a fruit-and-vegetable scanner that can accurately estimate their remaining shelf life by means of image recognition.[53] Such information

[45] Chen, 12 October 2017; Simonite, 3 September 2019, 3. It was previously revealed that the Chinese government also uses facial recognition to trace and monitor Uyghurs, an Islamic minority group (Mozur, 14 April 2019).

[46] The digital rights lobby organization Fight for the Future maintains a map of all the places where the US government uses facial recognition technology (Fight for the Future, undated).

[47] Chiusi et al., 2020.

[48] Kayser-Bril, 11 December 2019.

[49] European Commission, 2021c.

[50] A 'one-and-a-half-metre monitor' was deployed in Amsterdam, for example (Amsterdam Algorithm Register, undated).

[51] Tian et al., 2020.

[52] Serket, undated.

[53] OneThird, undated.

facilitates better decision-making and thus helps minimize waste. When a consignment of tomatoes, say, arrives at a distribution centre, the decision might be taken to send them for immediate processing because it is possible to see that they will be unsaleable by the time they reach the shops.

Although progress is being made in this field, the applications of computer vision are still often limited to specific tasks in specific domains. In clinical medicine it has proven relatively successful in the form of 'image-based diagnostics'[54]: images are scanned for particular irregularities that could indicate a disorder, helping radiologists, dermatologists and pathologists to detect and diagnose illness.[55] Such successes have been aided by healthcare being a data-rich sector, and much of that data being visual, so there is ample material to train the algorithms.

Computer vision also has the potential to improve the quality of medical imagery and help surgeons perform operations. The US Food and Drug Administration (FDA, the agency responsible for regulating medical devices) has recently approved ten diagnostic tools based on the technology for use in hospitals.[56] Computer vision-enabled apps have also been developed so that people can check themselves for health issues, in most cases skin conditions; one is the Dutch SkinVision utility, another Google's recently unveiled Derm Assist. Users scan their own skin and are then given advice on any follow-up that may be appropriate.[57] Although the medical world has been quite critical of such apps,[58] the examples we give do illustrate how computer vision can be utilized in practice.

3.2.3 Natural Language Processing

Our third general type of AI application automates the reading, analysis and generation of human language. The 'holy grail' of natural language processing is algorithms that can understand human language well enough to perform tasks requiring the interpretation of text. Language processing algorithms dissect sentences in various ways; for example, by distinguishing letters and words, labelling text elements and reading both left-to-right and right-to-left. This enables inferences to be made regarding the meaning of the text. Like computer vision, natural language processing has undergone a period of accelerated development in recent years, driven by

[54] A meta-analysis has revealed that the diagnostic performance of deep learning systems is similar to that of human medical professionals (Liu et al., 2019). However, the authors qualify that conclusion by saying that most of the studies included in the meta-analysis were not externally validated, meaning that the results did not support the general conclusion that AI was as good at image-based diagnosis as human doctors.

[55] Yu et al., 2018.

[56] Topol, 2019: 46.

[57] SkinVision, undated; Bui & Liu, 18 May 2021.

[58] Freeman et al., 2020.

advances in deep learning. Supported by sophisticated learning technology, the models can now be trained to understand human language more quickly and easily.

Because language is central to the way we communicate and how we gather, record and transfer knowledge, the potential applications of sophisticated natural language processing are enormous. In another parallel with computer vision, though, current systems are limited to specific tasks that require relatively little actual understanding of text input. Examples include tools that auto-correct, auto-complete or check text as it is typed, as well as automated translation systems like Google Translate.[59] Spam filters and search engines also make use of natural language processing. Google's search algorithm, for instance, applies two techniques when processing each query. First it links the words entered to relevant words in documents. The algorithm then ranks the various documents containing the words in question on the basis of assumed quality and relevance, as determined from the number of previous clicks on the page – a process known as 'page ranking'. This application of natural language processing has revolutionized the way we find information online. But it does not involve any true understanding of human language.

Another example is 'messenger bots', the automated chat systems that many organizations use for website-based customer support. Here AI helps provide customers with prompt, efficient assistance. In such applications, language processing is actually combined with expert systems: the algorithm analyses a question and, using a decision tree, selects the most appropriate reply or follow-up question. The Dutch police use such a chatbot to help people report internet fraud online; it checks that the report is complete, makes a preliminary appraisal of the case and advises the victim as to their best course of follow-up action.

3.2.4 Speech Recognition

Speech recognition is the AI domain concerned with the detection, analysis and interpretation of spoken human language. It involves the use of algorithms to distinguish words and sentences in spoken language and convert them to text – speech-to-text translation. One field in which this kind of application is being tested is healthcare, where AI systems transcribe discussions between doctors and patients.[60] A natural language processing tool then analyses the result, identifies important clinical information and produces a summary of the consultation – the aim being to reduce doctors' administrative workload and thus ultimately yield better consultation reports.[61] The same technology can also work in reverse, converting text into speech. As, for example, when a device reads an e-book out loud, or a speech

[59] Lewis-Kraus, 14 December 2016.

[60] Van Buchem et al., 2021. The study shows that such digital scribe systems are very promising. However, they are not currently used in clinical practice anywhere, making it impossible to draw conclusions regarding their value in clinical settings. Ajami, 2016.

[61] Wouda and Hutink 2019.

computer acts as a voice for someone who cannot speak or has difficulty doing so (such as a patient with motor neurone disease).[62]

Voice-controlled smart assistants like Apple's Siri, Google Assistant, Microsoft's Cortana and Amazon's Alexa combine the two technologies described above to enable spoken communication between human and computer. After responding to 'wake words' such as 'Siri' or 'Alexa', the tools are able to perform all sorts of tasks: searching the internet, compiling to-do lists, playing music, making restaurant reservations and so on. All the user has to do is give a clear spoken command. Speech recognition technology converts their speech into text, then natural language processing interprets the written information and determines what action is required.

Unlike people, who could speak and listen before they invented writing, computers find written language easier to process than the spoken word. Speech recognition is considerably more difficult because of the variability of spoken language and the noise in audio streams; picking out the words, identifying them and converting them into a type of text the computer can process is extremely challenging. Nor is interpretation of the speech signals themselves straightforward. When we speak, the sounds we make are not separated into distinct words. What a computer hears is very like what a person hears when listening to a language they are totally unfamiliar with: a continuous stream of sound, with individual words very hard to distinguish. Yet telling them apart is essential if we are ever to translate those words into a language we understand.

The problem posed by speech recognition thus differs fundamentally from the interpretation of written language or images. Unlike computer vision and natural language processing, speech recognition involves the processing of a single input variable – sound waves – that changes dynamically over time. The great challenge is distinguishing words and sentences within this input, so that they can be translated into a language the algorithm is able to process.

A further challenge is that some of the meaning of speech is conveyed by changes in volume, cadence and tone – the characteristics of spoken language. Effective interpretation therefore depends on more than simply distinguishing words from one another. The phonetic aspects need to be detected and interpreted as well, in order to determine the meaning of what is being said. Another stumbling block is homophones: words that sound the same but mean different things, such as 'hour' and 'our' or 'air' and 'heir'. Their interpretation depends on the context: both the narrow context of the sentence and the wider context of the situation, the speaker and so on.

As in other domains, advances in machine learning have led to progress in the field of speech recognition since it has become possible to process much greater volumes of speech data to train the algorithms. Relatively successful practical applications of speech-to-text and text-to-speech conversion are now viable, providing

[62] In collaboration with Google, DeepMind is currently working on a project to develop a text-to-speech program that would enable someone with a speech impediment to speak with their own voice via a computer (Chen et al., 18 December 2019).

that the speech is clear in both auditory and content terms. However, much spoken communication is unclear in one or both of these respects. Consequently, speech recognition technology has not yet reached the stage where it can be used reliably on a wide scale and for a range of purposes. To a large extent this is attributable to the limitations of natural language processing and AI's ability to truly understand language. Although it has made the transition from lab to society, AI is still far from being a mature technology – a point we return to in the final section of this chapter.

3.2.5 Robotics

In this report the term 'robotics' is applied to the type of AI used in combination with robots. Robotics brings together all types of AI: the ability to reason and learn, to see and hear, to communicate and to understand. However, it differs from other AI disciplines in that it additionally involves physical processing: the ability to manipulate objects.

A robot needs to be able to move and undertake physical actions to perform tasks. Those may be so-called 'dull, dirty, dangerous and dear jobs' or activities in which robots can outperform people. Examples include space exploration, the clean-up operations following the nuclear accident at Fukushima and defusing bombs.[63] However, robotics are important as well in the context of innovations in healthcare, retailing, manufacturing, livestock husbandry, agriculture and horticulture. Autonomous vehicles can also be regarded as a form of robotics. Robots thus come in countless shapes and sizes, making a precise definition of the word very difficult. Joseph Engelberger, a pioneer in the field of industrial robotics, addressed that challenge with a variation on the classic one-liner often used in respect of familiar but undefinable things: "I can't define a robot, but I know one when I see one."

In classic robotics, expert systems play an important role. They are particularly suitable for standardized tasks in situations where a choice needs to be made from a number of predefined courses of action. For that reason, robots are currently used mainly in controlled manufacturing and port environments. Deploying them in highly dynamic and often chaotic everyday human settings, such as on the roads, involves far more complex challenges. Coping with the variety and spontaneity of such situations requires a degree of understanding of how the world works; the robot needs to be able to observe its surroundings, assess situations, predict plausible future scenarios and decide, in a dynamic setting, which of all the possible courses of action is most appropriate for the circumstances.[64] A system flexible enough to operate in the world outside the lab must therefore be underpinned by such understanding.

[63] One example being the PackBot developed by Endeavor Robotics (previously iRobot), part of flir ugs (unmanned ground systems), which supplies robot systems to the US military (flir ugs, undated).

[64] Marcus & Davis, 2019: 113.

Because a robot of this kind would have to cope with the open-ended nature of our world, in which the possibilities are endless, it is crucial that it incorporate a wide range of capabilities. At present, though, the limitations of the other forms of AI effectively restrict the practical potential of this kind of robotics. For example, robots currently find it difficult to pick up a dark handkerchief from a dark table of their own accord, because computer vision is not yet sufficiently sensitive to light. Progress in the other branches of AI and advances in machine learning are therefore vital to the further development of robotics. Although the hardware and human control aspects are already quite impressive, everyday tasks require extremely refined motor control, planning and perceptual capability. While a commercial glasshouse may seem an orderly environment, for instance, it is still extremely difficult for a robot to pick a tomato without squashing it.

Three big players in the field of robotics are Boston Dynamics, which specializes in the simulation of human movement by robots ('humanoids'), DJI, a specialist manufacturer of drones for consumer use, and Amazon Robotics, with a focus on automated logistics. Amazon develops and deploys robots capable of efficiently navigating large warehouses and thus optimizing sorting processes. To do that the machines have to make allowances for and co-operate with one another, which they manage very successfully in facilities they are deployed in. Amazon's sorting robots perform specific tasks and operate in environments that are predictable and surveyable – for robots, at least.

By contrast, Boston Dynamics is aiming to develop robots that are far more flexible both physically and 'mentally', enabling them to be used for a variety of purposes. Previously owned by Alphabet (Google) but sold to Japanese technology giant SoftBank in 2017, Boston Dynamics is well known from its impressive video footage of two and four-legged robots such as Atlas and BigDog. They can stand and move around in ways that closely resemble the locomotion of people and animals. However, the company has yet to develop any commercial products. But Chinese technology firm DJI does already operate commercially; in fact, it is the global market leader in unmanned aircraft (drones) for aerial photography and video applications.

As the real-world examples above illustrate, AI is making its presence felt in society through robotics, speech recognition, natural language processing, computer vision and machine learning. Indeed, its practical applications are now so numerous and varied that it is impossible to compile a comprehensive overview. Nevertheless, the appendix to this report lists examples of AI applications in various sectors of the economy, primarily to provide an impression of their huge breadth and diversity. The simple observation that AI is today utilized in many different ways and contexts emphasizes the extent to which it is now becoming established within society. That process is not inconsequential. In the next section we consider the societal dynamics set in motion by AI's transition from the lab to society.

Key Points – The Practical Application of AI

- Having made the transition from lab to society, AI now has a variety of practical applications. We distinguish five general types of AI in everyday use.
- Machine learning: AI for predictive analysis. One familiar example is the recommender systems that personalize internet content suggestions.
- Computer vision: AI for the observation and analysis of visual information, such as recognizing faces or road signs.
- Natural language processing: AI for the interpretation of everyday human language. Chatbots use this technology, for example.
- Speech recognition: AI for spoken language processing. Voice-controlled assistants, such as Apple's Siri and Amazon's Alexa use this type of AI.
- Robotics: the combination of various AI capabilities with physical functionality. Examples include robots that transport goods inside warehouses.

3.3 AI as a Phenomenon in Society

AI's transformation from something researchers investigate into something used in everyday life has clear repercussions for society as a whole. Moving out of the lab inevitably implies moving into the public arena. The world AI has now entered is one of divergent interests and forces, and its arrival there has triggered investment, experimentation, discussion and alarm. Visions and strategies to utilize AI to maximum effect have been published, but also open letters and reports calling for its regulation. In short, the appearance of this new technology is making waves within society. As well as evolving technologically, AI is developing as a societal phenomenon (see Fig. 3.3). Although that process is still very much in progress, we can already discern a number of trends. To aid understanding of the current situation, in this section we consider society's various responses to the arrival of AI and the shifting emphasis in them.

3.3.1 Interest in AI as a Revolutionary Technology

The scientific blossoming of artificial intelligence after the most recent 'AI winter' focused attention on the technology's countless potential applications. That led to the appearance of several iconic books on its future, which often present the latest advances as the beginning of a new era. Visionary Ray Kurweil speculates on an imminent 'singularity' in which human and computer intelligence merge to form a

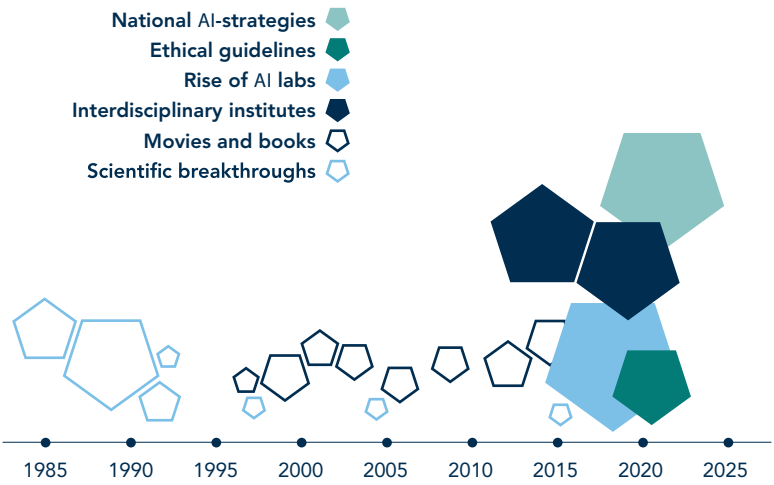


Fig. 3.3 AI's development as a societal phenomenon, per selected indicators

single superintelligent entity.[65] Scientists Erik Brynjolfsson and Andrew McAfee place AI at the heart of 'the second machine age', in which machines relieve humans not only of physical tasks but also cognitive ones.[66] Philosopher Nick Bostrom views such developments as a serious threat, however: as AI becomes cleverer and faster than us, it becomes hard for humanity to maintain control over it.[67] Another philosopher, Luciano Floridi, refers to a 'fourth revolution' in which digital technologies like AI fundamentally change our world view and our understanding of ourselves.[68] Klaus Schwab, financial backer and chair of the World Economic Forum, talks of a 'fourth industrial revolution' when the application of smart technology transforms the way we work and live, just as the steam engine, electricity and digitalization did previously.[69]

A future in which human life is closely intertwined with AI is also a popular theme for the film industry. In parallel with the recent 'AI spring', movies such as *Her* (2013), *Ex Machina* (2014) and *Transcendence* (2014) depict a future where AI reaches a critical threshold of intellectual ability. As such, these films serve as a form of 'scenario thinking': they portray imagined situations in which humans have emotional relationships with AI and can even fall in love with it (*Her*), in which AI can pass the ultimate Turing test and become so like us that it is no longer possible to distinguish between human and machine (*Ex Machina*) or in which AI becomes a dangerous, barely controllable source of power (*Transcendence*). Although the

[65] Kurzweil, 2005.

[66] Brynjolfsson & McAfee, 2014.

[67] Bostrom, 2016.

[68] Floridi, 2014.

[69] Schwab, 2016.

idea of the superintelligent computer has long been a source of inspiration for screen writers, these recent productions have made the specific term 'AI' familiar to the general public.

3.3.2 Applied Research and the Run on Talent

Besides compelling screen depictions, what has mainly stimulated public interest in new AI technologies is their practical potential. That has also made AI economically attractive for the business community and governments. We have already described how private investment in AI has increased considerably all around the world. In addition, businesses and governments have teamed up with research institutes to set up special 'AI labs' where links are forged between fundamental science and practical requirements. In many parts of the world laboratories of this kind have been established to cater for particular sectors of the economy, ranging from agriculture and mobility to retail and manufacturing, healthcare and education to public administration. Others are addressing the societal aspects of AI; these are often known as ELSI (ethical, legal and social implications) or ELSA (ethical, legal and societal aspects) labs.

The first applied research facility in the Netherlands devoted to AI began life in 2015. That was the QUVA Deep Vision Lab, a joint initiative by the University of Amsterdam and Qualcomm dedicated to translating computer vision research into industrial applications. Similar projects proliferated in the years that followed, with the Innovation Center for Artificial Intelligence (ICAI) founded in 2018 by the University of Amsterdam and VU Amsterdam playing an important co-ordinating and supporting role. The Netherlands now has twenty ICAI labs, where companies including Bosch, TomTom, KPN, ING, Ahold-Delhaize and DSM, as well as hospitals, the national police and government bodies, collaborate with universities and research centres to develop innovative AI solutions.

With businesses also exploring AI's potential in many different fields, and developing applications for them, an enormous demand has arisen for talent in this domain. That in turn has sparked debate in various countries as to how best to nurture and retain people with the necessary skills.[70] At the beginning of this chapter we pointed out that most AI-related PhD graduates in the US are now choosing careers in industry. Other countries are experiencing a similar 'brain drain' from academia to the business community, but in many cases with their trained specialists moving abroad to boot.[71] In Europe, prominent scientists from more than twenty countries have written an open letter on the subject to sound the alarm and call on policymakers to invest in the European research climate.[72]

[70] Hoeks, 12 April 2019; Delcker, 27 June 2018; Boland, 2 September 2018.

[71] Elsevier, 2018.

[72] ELLIS, 2018.

3.3.3 AI Action Plans

The focus on the potential of AI is also reflected in a proliferation of national and international AI strategies. Most such documents deal primarily with the economic opportunities, often within those sectors already important for the countries in question.[73] The OECD observes that the goal of national strategies is usually to boost national productivity and competitiveness by harnessing AI.[74] They are therefore concerned primarily with the development and utilization of AI through mechanisms like research funding, enhanced support infrastructures and encouraging business interest. For the same reason the development and retention of talent is an important feature of many strategies.[75]

Although the main thrust of an AI strategy is typically the definition of an innovation agenda, many additionally address societal and ethical aspects. However, the passages devoted to these points are often subordinate to the economic plans and usually less substantive and action-oriented. In Europe the rationale for discrepancies of that kind tends to be that, in order to align AI with our values, we need to be in the technological vanguard.

The Dutch think tank DenkWerk produced a report entitled *AI in Nederland* (AI in the Netherlands) in 2018. This stressed the urgent need for the country to commit seriously to artificial intelligence, arguing that it was being left behind in terms of investment in the private sector and other forms of government support. DenkWerk pointed to the 'enormous societal potential' of AI and called on the government to formulate a national agenda for its development and application. The report urged immediate action, saying, "This is not a matter that should first be considered for two years".[76] That same year DenkWerk helped to initiate work on a national AI agenda. AiNed, a coalition of corporate, academic and government partners, then published a report substantiating the earlier call for urgent action. That argued that AI should be made a national priority to protect and enhance the nation's prosperity and international status. With a view to accelerating the development of AI and differentiating the Netherlands on the global stage, AiNed formulated several objectives as the basis for a national strategy.

The strategy was eventually published in autumn 2019. Following the release of the AiNed report, a task force was formed. Led by employers' confederation VNO-NCW, this also involved the Ministry of Economic Affairs and Climate Policy, which assumed responsibility for realizing the objectives formulated. The first major practical step was to create a Dutch AI Coalition, a platform for collaboration between businesses, government bodies, non-governmental organizations and research institutes to catalyse AI development. The coalition quickly announced its intention to promote the formation of AI labs and to work with the government to

[73] Mols, 2019.

[74] OECD, 2019: 121.

[75] Mols, 2019.

[76] DenkWerk, 2018.

develop an AI strategy. Another outcome is the Strategic Action Plan for AI (SAPAI), presented by the Ministry of Economic Affairs and Climate Policy with the support of the ministries of Justice and Security, the Interior and Kingdom Relations, Social Affairs and Employment and Education, Culture and Science.

In this plan the government sets out a pathway for the period ahead and describes the first practical initiatives to accelerate AI development and raise the Netherlands' profile in this field. The SAPAI defines a three-track policy. Track 1 relates to utilization of the societal and economic opportunities offered by AI, track 2 to creation of a conducive ecosystem and track 3 to safeguards. Inclusion of the third track reflects how the debate has shifted, with attention focusing not only on the economic opportunities afforded by AI but also increasingly on the impact of its applications. The SAPAI was presented together with documents devoted to 'AI, civic values and human rights'[77] and 'Safeguards against risks associated with government data analyses'.[78]

The private investment, the formation of AI labs and the launch of national AI strategies are indicative of the increased interest in AI outside the scientific community. Much of that focuses on the potential of AI. There is growing awareness that it has now reached a certain level of maturity and so the time has come for the appropriate actors to realize its potential. AI is on the agenda, particularly the economic agenda. Stories about new applications and the doors they can open appear regularly in the media. As a result, the general public has become aware of the technology. Many may not understand quite what AI is, but they know of its existence.

3.3.4 Interest in the Practical Effects of AI

As AI has become the focus of increased attention, questions about its impact have arisen. The technology's introduction to the real world has led people to consider its implications for everyday life. In some recent books the emphasis has shifted from the revolutionary nature of AI to the possible consequences of its use in real life. Various authors have highlighted potential problems associated with its transition from the lab to the mainstream. Moral, societal, political, legal and economic issues have all been raised. AI's effect on society and its core values has become a matter of public debate.

In her book *Weapons of Math Destruction* (2017), Cathy O'Neil warns of the harmful effects that the careless and short-sighted use of algorithms can have on people's lives.[79] Meredith Broussard has a similar message, coining the term 'technochauvinism' to describe how mankind can be insidiously degraded by the idea

[77] *Kamerstukken II* 2019/20, 26 643, no. 642.

[78] *Kamerstukken II* 2019/20, 26 643, no. 641.

[79] O'Neil, 2016.

that technology is capable of meeting every human need.[80] Shoshana Zuboff has also expressed concerns, but about the actors controlling the technology rather than the technology itself. She warns of 'surveillance capitalism', an economic philosophy based on excessive data gathering and use of predictive algorithms, allowing big tech companies to exercise unprecedented influence over our behaviour. In *The Algorithmic Society* (2020), various authors highlight the association between data, algorithms and power and describe how that association can distort the relationship between citizen and state. In his most recent work even Stuart Russell, author of the definitive AI handbook, expresses concern about AI's effect on the real world. A system that works well in a technical sense may nonetheless have undesirable effects, Russell writes. He argues that this makes it important to keep AI permanently under control: "What's worse than a society-destroying AI? A society-destroying AI that won't switch off."[81]

Living with AI has become an important theme. In 2016 the World Economic Forum was devoted to designing a world of smart technologies such as AI. In the same year G7 IT ministers agreed with the OECD that international talks should be held regarding the development of AI and its economic and societal implications. The OECD has since been increasingly active in this field, organizing conferences and encouraging international policy discussions. In 2019 the organization presented its principles for AI, identifying the technology's effect on people and society as an important theme and setting out a framework for responsible further development. All OECD member states, the G20 and ten other countries have endorsed the principles, which have thus become the first intergovernmental guidelines on AI.

3.3.5 Social Organizations Become Involved

Attention is shifting from AI's economic impact to its societal impact. At the international level, UNESCO has also started taking an interest. In 2018 an entire edition of the organization's magazine, *The UNESCO Courier*, was devoted to the opportunities and threats society associates with AI. In her contribution, Director-General Audrey Azoulay stressed the importance of an ethical debate on AI. Unsurprisingly, she saw a role here for UNESCO: "It is our responsibility … to enter this new era with our eyes wide open."[82] Her organization is seeking to discharge that responsibility by working on a global ethical standard for AI, to serve as a basis for the development of national policy.[83] The European Commission has meanwhile established the High-Level Expert Group on AI (AI HLEG), a body charged with

[80] Broussard, 2019.

[81] Sample, 24 October 2019.

[82] Šopova, 2018.

[83] The draft of this document was presented in 2020 (Ad Hoc Expert Group, 7 September 2020).

providing European governments with an ethical framework for the technology. Following the publication of the European strategy for the development of AI, responsible development in which the effects of AI are taken into account is now also on the agenda. The AI HLEG's guidelines and recommendations for 'trustworthy AI' are intended to promote awareness among policymakers of the ethical and societal aspects of AI and to provide a framework for managing them.[84]

Concepts like 'human-centric' AI and ethical, humane and responsible AI are being mentioned with increasing frequency, indicating growing interest in the relationship between AI applications and human society and values. That trend is also reflected in the proliferation of publications about AI and ethics.[85] Research institutes devoted specifically to the societal implications of AI have been established too. Even in Silicon Valley, the crucible of AI's technological development, the Stanford Institute for Human-Centered AI has been set up specifically to investigate the technology's human and societal impact.

The AI Now Institute, also in the US, is perhaps the most prominent example of a research centre concerned with the societal effects of AI. Its annual reports serve as important catalysts of worldwide debate. Since the first appeared in 2016, AI Now's messaging has become increasingly clear: having initially called for research into the effects of AI, the institute has moved on to arguing that certain applications should be prohibited, sometimes at least provisionally, and to setting out specific requirements for the responsible use of the technology.

The changing tone of these recommendations illustrates how the public debate on AI has developed in recent years: from promoting awareness of its effects to a substantive discourse about how certain values can be impacted and protected. This trend is apparent in the Netherlands too. The Rathenau Institute – the Netherlands Organization for Technology Assessment – began by advancing the cause of public debate regarding the impact of digital technologies such as AI, but more recently has become an active contributor to the discussion on how that impact should be managed. As experience and research have made the practical effects of new technologies clearer, so firmer ideas have emerged as to how undesirable effects should be countered. The debate many commentators were advocating a few years ago is actually happening today and become increasingly substantive.

3.3.6 Sectoral Interest in AI

Growing recognition of AI's practical potential has attracted attention from research centres and consultancies concerned primarily with non-technological fields. Having previously thought of AI as part of the general issue of digitalization, organizations active in such domains as education, healthcare, security, infrastructure

[84] High-Level Expert Group on Artificial Intelligence, 2019a, c.

[85] Zhang et al., 2021: 130.

and the law have in recent years turned their attention to the specific question of AI's implications for their disciplines.

Since 2018, various sectoral bodies in the Netherlands have published studies and advisory reports addressing the significance of AI for their particular domains. The Advisory Council on International Affairs (Adviesraad Internationale Vraagstukken, AIV) and the Advisory Committee Public International Law (Commissie van Advies inzake Volkenrechtelijke Vraagstukken, CAVV) have reported on the military applications of AI,[86] while the Netherlands Environmental Assessment Agency (Planbureau voor de Leefomgeving, PBL) has considered what smart algorithms could mean for mobility,[87] the Netherlands Centre for Ethics and Health (Centrum voor Ethiek en Gezondheid, CEG)[88] and the Council for Public Health and Society (Raad voor Volksgezondheid en Samenleving, RVS)[89] have explored the implications of AI in the healthcare sector and Dialogic was commissioned by the Ministry of Education, Culture and Science to investigate AI's impact on education.[90]

These explorations and recommendations add texture to the AI-debate: the different contexts in which AI can be applied, demonstrate the breadth and diversity of its prospective impact. Meanwhile, more experience is being gathered with the deployment of AI and this reveals the difficulties and risks in the step towards its practical application. Examples of discrimination by algorithms, accidents with self-driving cars and the disappearance of the human dimension through excessive 'algorithmization' give food for thought for the ways in which we want to integrate AI into society.

3.3.7 The Dark Side of AI

Paralleling the growing sector-specific interest in AI, more attention has been paid to the dark side of the technology. Examples of algorithmic discrimination, accidents involving self-driving cars and dehumanization associated with excessive reliance on technology have prompted people to reflect on how they want AI integrated into society. In Europe, the EU's Agency for Fundamental Rights (FRA) is investigating potential implications in its field. The role of AI in the development of autonomous weapons, the use of facial recognition by local authorities and police forces and the status of 'big tech' have all become topics of public concern. The potentially harmful side of AI is starting to dominate debate.

[86] AIV and CAVV, 2015.

[87] PBL, 2017.

[88] CEG, 2018.

[89] RVS, 2019.

[90] Van der Vorst et al., 2019.

Moreover, the risks posed by AI and its dual-use potential, combined with the speed at which the technology is currently evolving, are adding a degree of urgency to the debate: if its increasing use is not to have undesirable consequences, not only do we need a clear picture of the risks but we also have to respond accordingly. Various US states and cities have now prohibited the use of facial recognition by the police and in public places.[91] The European Commission's draft Artificial Intelligence Act seeks to do the same, except in special circumstances such as where compelling security considerations exist. Both the US and Europe have for some time been looking at possible ways to curb the burgeoning power of big tech corporations through competition law.

Campaign groups and lobby organizations are taking up AI-related causes as well. They include, firstly, groups dedicated to addressing problems of a particular kind – privacy or digitalization issues, for example – that have developed an interest in potential abuses associated with the use of AI. In Europe, EDRi (European Digital Rights) and other groups dedicated to protecting rights and freedoms in the digital environment are now concerning themselves with AI. Secondly, we are now seeing groups dedicated specifically to AI-related issues. They include AlgorithmWatch in Germany, which systematically surveys and critically evaluates the international use of algorithmic systems. Another development is that some major human-rights organizations, such as Amnesty International, Hivos, Human Rights Watch and UNICEF, have also started to take an interest.[92] In short, there is a growing movement within civil society concerned with the negative effects of AI.

3.3.8 On the Policy Agenda

Within government too, AI-related issues are commanding greater attention. In part as one aspect of the wider debate about digitalization and privacy, as recognized by the research institutes and advisory councils. In the early 2010s discussion of digitalization was dominated by questions relating to big data and privacy. At that time the EU was working on the General Data Protection Regulation (GDPR) with a view to providing a legal framework to enhance the protection of personal data, particularly in the digital domain.

The focus on privacy gave rise to interest in transparency as well, another principle prominent in the GDPR, and both figured in the political debate regarding AI from the outset. Alongside the more general debate regarding digitalization, the discourse around big data is gradually transforming into a discussion about how that is processed using ever more intelligent algorithms. In its advisory report *Big Data*

[91] In the states of California, Oregon and New Hampshire, the use of biometric surveillance technology by the police is against the law, while all use of facial recognition in public places is prohibited in the cities of Portland, San Francisco and Oakland.

[92] Since 2016 Amnesty has had a separate department, Amnesty Tech, dedicated to digital technology matters including AI.

in a Free and Secure Society, the WRR has already highlighted the crucial role that algorithms play in big-data processes.[93]

Since 2018, however, the public debate regarding AI has broadened discernibly. At the European level that was the year in which the AI HLEG was set up. It went on to publish a set of *Ethical Guidelines for Trustworthy AI* (2019), which introduced such concepts as unfair bias, accountability and welfare to the discussion. Its effect has been to raise the profile of issues of discrimination and human control. The latter became an important principle in the European White Paper on AI (2020) and the subsequent draft Artificial Intelligence Act (2021), which is considered in more detail in Chap. 7.

To begin with, however, it is the immediate challenges associated with AI applications that command most attention. In response, efforts are being made to find practical means to address those challenges. One idea that is regularly floated is the establishment of an 'AI authority' or 'algorithm watchdog' to supervise the use of artificial intelligence. Other efforts to manage its direct effects include the development of standards for AI applications by organizations such as CEN-CENELEC at the European level and the ISO at the global level, as well as the ongoing legislative initiatives.

Meanwhile, a second broadening of the public debate is now discernible. There is interest not only in AI's implications for public values in particular contexts, but also increasingly in its impact on society as a whole. The work of the Council of Europe's Ad Hoc Committee on Artificial Intelligence (CAHAI) illustrates this trend; it takes a broad interest in AI's relationship with human rights, democracy and the rule of law. The number of governments commissioning research into its societal impact and convening advisory committees to provide wide-ranging policy-support information is further evidence of a growing recognition that AI has potential implications for all aspects of society and therefore requires structural attention.

3.3.9 Ethics

As awareness of the effects of AI has grown, ethics have become an important feature of the debate in recent years. Governments and private-sector actors have developed ethical codes and guidelines on the responsible use of AI, and university technology programmes have been adding ethics modules to their curriculums.[94] In both technical and social studies, increasing interest in the wider relationship

[93] WRR, 2016.

[94] Ethics is now one of the research domains within Delft University of Technology's AI programme, while Eindhoven University of Technology has made Ethics a compulsory module. The University of Amsterdam includes a module entitled Fairness, Accountability, Confidentiality & Transparency in its programme.

between AI and society has become evident.[95] Alongside their technology-based AI professorships, several Dutch universities have recently created chairs covering its societal and community aspects.

More systemic study of the implications of AI is coinciding with the emergence of a degree of 'ethics fatigue'. Although in practice many things covered by that term have little to do with ethics, there is growing dissatisfaction with the plethora of codes and guidelines that AI is expected to comply with. These often fail to reflect the complexity of the field in practice and provide an inadequate framework to prevent abuses and undesirable developments. It seems that more structural safeguards are required to ensure that AI is aligned with our common values, and that is shifting attention beyond its actual applications to the broader dynamics of its integration into society.

3.3.10 Interest in the Societal Integration of AI

We are currently at a stage where there is widespread interest in AI as a multifunctional technology with great economic potential. It has also become clear that its use will have a transformative effect on established practices and could lead to undesirable situations. Until recently most attention focused on the short term and much of it on specific values. However, the scope of the debate has now broadened to encompass AI's effects in a variety of domains and its impact on a wider range of values. Whereas the debate initially related mainly to matters of privacy, transparency and human control, there is now also interest in how AI affects other values, such as sustainability (see Box 3.2).

The Dutch government's request to the WRR for advice on AI, which led to this report, was signed by nine different ministers, indicating how its impact is relevant to all areas of policy and has the potential to affect all their core values. Interest in the effect of AI in particular contexts has effectively coalesced into an interest in its impact on society as a whole.

Since AI first appeared on the public agenda as a revolutionary technology, its effects and especially its risks have gradually become more important topics of debate. Now that it is being put to practical use in more and more spheres, and is set to find even wider applications in the future, interest in its impact is becoming more structural: as we develop AI, how can we safeguard the things that we value as a society, our civic values? To answer this question, we must look beyond AI's immediate effects and consider the longer term. Safeguarding civic values depends not only on the robustness of our technical systems, but also on the structure of society itself.

---

[95] The AI programmes at Utrecht University and VU Amsterdam include modules entitled, respectively, Philosophy of AI and AI & Society. The Faculty of Social Sciences at Maastricht University now offers a Bachelor's degree programme entitled Digital Society, which explores the societal impact of digital technologies like AI.

Box 3.2: AI and Sustainability

As AI enters everyday life, there is increasing interest in its impact on society. Issues concerning privacy, equal treatment, autonomy and security have become the focus of growing debate. Another pertinent subject – with a societal and political profile that has so far been quite low, but now attracting more and more attention from researchers – is the effect on sustainability.

There is an optimistic school of thought that AI can make a substantial contribution to enhancing sustainability. The UN's annual AI for Good conference is devoted to topics like ecological objectives in relation to AI. Furthermore, we are now seeing numerous initiatives through which AI is indeed adding substantively to sustainability. The best known are projects to make more efficient use of energy and improve wind and solar energy forecasting. But AI is also being used for smart farming. Amsterdam-based start-up Connecterra uses Google algorithms in livestock husbandry, for example.[96] There have also been interesting initiatives in nature conservation. One is eBird's use of machine learning algorithms in ornithology and utilization of the output data for bird protection. Another example is the use of AI by Global Fishing Watch for population monitoring. Finally, the EU's Destination Earth initiative (DestinE), for which the use of AI is also envisaged, should not be overlooked.[97]

On the other hand, there is a growing body of evidence that AI can have negative impacts on sustainability. The CO_2 footprint of the global computing infrastructure is already greater than that of the aviation industry at its zenith. Running a single natural language processing algorithm is associated with emissions equivalent to 125 return flights between New York and Beijing.[98] Furthermore, AI is being used to maximize fossil energy production and to promote non-sustainable consumption.

Peter Dauvergne has claimed that for every example of AI having a positive impact on sustainability, there are multiple cases of negative impacts. He attributes that to the wider political economy and the power structures associated with the technology. As long as the landscape is characterized by a commercial logic focused on exploitation, AI will not have a positive net effect on sustainability.[99] Achieving its potential in this area, he argues, will require changes to power structures, to the actors using AI and to the purposes for which the technology is used.

[96] Dauverge, 2020.

[97] European Commission, undated (n.d.).

[98] Crawford, 2021.

[99] Dauverge, 2020.

At this point in the development of AI we face the challenge of determining what is needed to achieve the technology's structural integration within society. Before considering the various aspects of that issue, it is important first to consider the role the lab will continue to play in the process.

Key Points – AI as a Phenomenon in Society
- AI's transition from lab to society has generated a societal dynamic.
- At first that dynamic was characterized by interest in AI as a revolutionary technology. Initially, the primary focus was the economic opportunities.
- As more practical experience has been gained, the potential negative consequences of using AI have become clearer. As a result, interest in the opportunities is increasingly accompanied by consideration of the risks, and a public and political debate has arisen.
- The AI debate initially focused on specific values such as privacy, non-discrimination and transparency and on application of the technology in particular contexts. However, AI's wide range of potential uses has broadened the debate to cover its impact on society as a whole and all the associated civic values.

3.4 The Future of the Lab

We have seen how AI has moved out of the lab and become a feature of society in many different respects, in the form of numerous applications and wide-ranging public debate. What implications do such developments have for the future of the lab? We should not expect that now AI has established itself within society, the lab's significance or dynamism will decline – that would be a mistake. The transition from lab to society does not imply that AI is a perfected technology requiring no further development, or that from now on attention should focus solely on its applications.[100] Rather, the breadth of the lab-to-society transition implies that a wider range of issues related to AI's societal integration will warrant attention, as we outline in the next chapter.

Despite all the activity in the application sphere, lab development remains vital for at least two reasons. The first is that, despite all the advances made in recent years and the innovations they have enabled, AI still has significant limitations. The

[100] That was suggested by computer scientist Kai-Fu Lee (Lee, 2018: 143). Of course, he recognizes that AI can develop further, but he believes that the discoveries made in recent years are so significant that it is unlikely that equally important breakthroughs will emerge in the near future. For example, he suggests that, since Geoffrey Hinton's important paper, a decade has passed without any equally revolutionary advances in machine learning. He argues that applications therefore warrant more attention than fundamental research. That is not the message of this report, however.

current methods offer no answer to a variety of questions. People are already talking about the limits to the capabilities of deep learning, for example. While it is impossible to say whether this will lead to a third 'AI winter', it is clear that the technology still has a long way to go, and that significant progress will require further fundamental research. The second reason for the lab's continuing importance relates to the particular nature of AI. It is a form of technology in whose application the lab must remain involved. Strictly speaking, then, AI has not left the lab but extended beyond it.

3.4.1 The Need for Fundamental Research

Various experts have made the point that access to more and better data is the key to overcoming many of the current limitations to machine learning. Furthermore, interesting developments are taking place in this particular field, like the use of generative adversarial networks (GANs) in which multiple algorithms are used to improve one another. One algorithm generates something new, such as an image of a bird, and in response the other algorithm indicates whether it recognizes it as a bird or not. If not, the first algorithm continues refining the image until the second one is 'convinced'.

Ray Kurzweil believes that such simulation methods can resolve many of the problems associated with data shortages. For example, rather than self-driving cars having to learn in real traffic, with all the attendant dangers, they could travel millions of kilometres in simulated worlds without putting anyone at risk.[101] Similarly, defence robots could be trained within a simulation so that they are more advanced prior to deployment in the real world. Another promising approach is federated learning, where data to train a machine learning algorithm is not loaded onto a central server but algorithms are refined by adjusting their parameters with those from other datasets, without combining the actual data. This approach is particularly suited for use with privacy-sensitive data such as hospital records.

Despite such developments, scientists believe that innovation remains necessary, partly because machine learning appears to have inherent limitations. For example, progress is needed in the field of computer vision if it is ultimately to be used for autonomous vehicles or security applications. Now its algorithms are relatively easy to fool, as demonstrated by experiments showing that tiny traffic signs too small for the human eye to see were treated like the real thing by self-driving cars. Current algorithms look for patterns, so if a minute sign closely matches the pattern sought then it will be interpreted with confidence as a real one. Its abnormal dimensions are not noticed by the existing algorithms.

[101] From an interview with Ray Kurzweil (Ford, 2018: 230).

This makes such algorithms vulnerable to adversarial attacks – with potentially disastrous consequences in the case of a self-driving car. A recent study demonstrated that changing a single pixel in an image can confuse an AI algorithm. The military use of AI is another field where vulnerabilities like these can have serious repercussions; it has been shown, for instance, that an image classifier can be fooled into identifying a machine gun as a helicopter.[102] The same attack strategy could be deployed for other purposes too. Google uses an algorithm to classify videos for the protection of intellectual property rights and so on. Researchers at the University of Washington showed that this could be tricked by inserting random images into a video for fractions of a second.[103] In an incident in the US, a police officer who was being filmed started playing music, presumably in the belief that YouTube's algorithms would prevent the video being shared on intellectual property rights grounds.[104]

3.4.2 Superficial and Inefficient

Numerous other shortcomings with machine learning illustrate that a great deal of lab work is still required, as AI pioneers have themselves acknowledged. Yoshua Bengio has argued that deep neural networks can learn superficial statistical regularities from datasets, but not higher abstract concepts. They therefore lack the type of understanding needed for certain tasks and forms of communication. Geoffrey Hinton and Demis Hassabis, founder of DeepMind, have both stated that general artificial intelligence is currently nowhere near being a workable reality.[105]

Pioneer Hinton is critical of current methods and has highlighted various shortcomings.[106] One is inefficiency. Machine learning is more like human learning than earlier technologies were. For example, images are recognized by identifying patterns rather than by following fixed rules. In that respect the machine and human learning processes are alike. Nevertheless, major differences also exist, and humans are still able to learn far more efficiently. A small child only needs to see a few apples to acquire the ability to recognize apples in the future. By contrast, machine learning algorithms need to be shown thousands of images of apples before they are trained to identify the fruit. Furthermore, while the volume of data available globally to train algorithms is increasing, the situation varies from domain to domain; many still have a shortage. Another problem is that, for certain applications, high error rates during the training phase entail serious dangers.

[102] Tonin, 2019: 6.

[103] Agrawal et al., 2018: 200.

[104] Thomas, 9 February 2021.

[105] Marcus & Davis, 2019: 62.

[106] Dickson 2 March 2020.

3.4.3 Common Sense

A related problem is AI's lack of common sense. The earlier example of the tiny road signs perfectly illustrates this. Current algorithms are designed to be capable of processing all possible images and therefore cannot distinguish between plausible and implausible contexts. Although they recognize patterns, they are not good at ascribing significance to them. CAPTCHAs – completely automated public Turing tests to tell computers and humans apart – are a good example. They are the tests you come across on the internet that ask you to prove you are not a robot by, for example, selecting all the photos in a group that include trees. Passing requires common sense. Even if a picture includes only a small part of a tree, a human can usually see what it is by drawing conclusions from the surrounding objects, such as bushes.

For an algorithm, however, the limited number of data points available as a basis for recognition is usually problematic. CAPTCHAs therefore reveal the current limitations of machine learning in situations that call for common sense. Machine learning algorithms have no access to the collective knowledge acquired elsewhere by other programs. They therefore have trouble answering questions that humans can answer without hesitation, like "Who is taller, Prince William or his young son Prince George?" or "If you stick a pin into a carrot, will you make a hole in the pin or the carrot?"[107]

Humans answer such questions by drawing on a large pool of implicit knowledge. When we speak, we do not provide all the relevant information because we assume that the listener will make deductions based on the context. If someone instructs a taxi driver to take them to the airport as quickly as possible, the driver knows that they are not expected to drive without regard for the rules of the road or the safety of other road users. An algorithm, however, lacks that kind of implicit background knowledge. In other words, the language is underspecified, and no fact exists in isolation.[108] Stuart Russell cites the example of progress in the field of physics. By analysing data from telescopes, an algorithm can develop new knowledge. However, progress depends on more than merely studying additional data. The formulation of hypotheses and the selection of factors for inclusion from the universal data pool rely on prior knowledge of physics, which does not exist in a form an algorithm can process.[109]

[107] Marcus, 2018: 12.

[108] Marcus & Davis, 2019: 136–139.

[109] Russell, 2019: 83.

3.4.4 *Lack of Transparency*

Another shortcoming is that current machine learning algorithms lack transparency, which often makes it extremely difficult to ascertain how they come to a given conclusion. In many cases the decision-making process can be uncovered, but that does not necessarily imply that it is explainable: knowledge is not the same as understanding. A decision to classify something based on a pixel-level detail is unfathomable for humans, for example. In many cases this is not a great problem. If the algorithm's decision relates to something like a security risk, a mortgage application or a medical diagnosis, however, opacity has serious implications. In such contexts, explainability is therefore a requirement.

Some experts believe that the complexity of the algorithms needed for applications of this kind does not present an insurmountable problem. Hassabis, for example, takes the view that we are currently building the systems and that the construction phase will be followed by a process of reverse engineering aimed at understanding how they actually work. He therefore believes that, within a decade, most systems will no longer be black boxes.[110] Yann LeCun suggests that it is as if we are still in the process of inventing the internal combustion engine but are already worrying about brakes and seatbelts. Such problems can be addressed at a later stage, he argues.[111]

Other experts see poor explainability as inherent to the technology, making a different approach necessary. According to Judea Pearl, human knowledge is expanded not by a blind process but by building and testing models of reality. Current machine learning approaches are limited because they focus on correlation, not causality.[112] He draws an analogy with the difference between Babylonian and Greek astronomy. While the Babylonians were able to make very accurate predictions, better than the later Greeks, the process they used was unreproducible – a black box – and the mechanisms underpinning the predictions were not understood. The Greek approach was based on understanding those mechanisms, and the emphasis on causality proved central to the subsequent development of science. Pearl thus regards the current non-model-based approach to machine learning as inadequate.[113]

3.4.5 *Old and New Approaches*

With a view to addressing shortcomings of this kind, alternative approaches are now being developed. Several build on 'good old-fashioned AI', the symbolic technology with which the discipline began. Such rule-based systems are used, for

[110] From an interview with Demis Hassabis (Ford, 2018: 178).

[111] From an interview with Yann LeCun (Ford, 2018: 136).

[112] From an interview with Judea Pearl (Ford, 2018: 363). 211 Pearl, 2019: 18.

[113] Pearl, 2019: 18.

example, where the amount of available data is limited. Siemens has a system built on that principle to control gas turbine processes in its factories. Without predefined rules the turbines would have to run for a century to train an algorithm to do the job effectively by means of machine learning.[114] It is also difficult to apply machine learning in situations where that would imply using large volumes of privacy-sensitive data and generating results with an opaque basis. In such cases, top-down logical systems may offer a solution. Another, related suggestion is to use rule-based systems in combination with machine learning to predict outcomes and thus deduce what rules are being followed, so making the results more transparent. People like Yann LeCun and Nick Bostrom believe that the future lies in adding structure and modelling to existing machine learning techniques.[115]

In a variation on the hybrid approach, efforts are being made to code common sense into algorithms. For example, DARPA has a Machine Common Sense programme. This is creating models that distinguish between various categories, such as objects, locations and actors, as happens in human cognition. There are also related approaches that involve building certain principles into algorithms so that they do not have to learn everything from scratch; these work like the inductive biases that influence learning in children. At an early age children learn the basic physics of objects, how they move through space and, for example, that they cannot pass through each other. Principles of that kind guide and accelerate the learning process so that a child does not have to see thousands of examples of something before it can recognize the item in question. One approach that works that way is the graph network, in which objects are represented by circles and relationships by lines.[116] Geoffrey Hinton is working on 'thought vectors' to better capture the meaning of language,[117] while the Allen Institute for AI's Project Mosaic is endeavouring to program common sense into computers.

Another set of approaches has close links to neuroscience. Just as neural networks were inspired by the workings of the brain, so there are now initiatives to create neuromorphic chips. If successful, future computers could be fitted with chips modelled on the workings of neurons. The European Union's Human Brain Project is aimed at building a brain made up of computers,[118] as is the BRAIN Initiative in the US.[119]

Besides the two familiar approaches of symbolic AI and artificial neural networks, Margaret Boden distinguishes another three. They are evolutionary programming, cellular automata and dynamical systems.[120] In his quest to find the ultimate algorithm, Pedro Domingos has identified three new techniques that can

[114] Wilson et al., 14 January 2019.

[115] Ford, 2018: 78, 108, 126.

[116] Waldrop, 2019.

[117] Marcus & Davis, 2019: 128.

[118] Marsh, 10 January 2019.

[119] Domingos, 2017: 118.

[120] Boden, 2018: 5–6.

contribute to the search alongside the symbolic and neural approaches. Genetic programming is an approach used in the design of electronics and the optimization of factories.[121] There are also Bayesian methods, such as naive Bayes classifiers and hidden Markov models, which are used for spam filters, speech recognition systems, cleaning up data series and so on.[122] Finally, there are analogy-based systems like the nearest-neighbour algorithm used by support vector machines. The analogy-based approach has been used for modelling the solar system and atoms, and to produce music in the style of particular composers.[123]

Although machine learning has taken off in recent years and has rapidly been adopted in a wide variety of domains, it has its limitations and alternatives are being investigated. Each of these has its own strengths and weaknesses, meaning that the most suitable approach differs from one application to another. It may be that in the future the emphasis will be placed on selecting the right approach for each application, with no particular one regarded as universally preferable. Many experts see hybrid approaches as the future, arguing that human intelligence works in a similar way. For example, the unconscious recognition of familiar patterns is attributable to neural networks whereas unfamiliar situations are addressed using conscious reasoning, which is more akin to symbolic AI. From all this we can safely conclude that AI is far from perfected and so fundamental research is going to remain very important.

3.4.6 The Lab Belongs with AI

As indicated earlier, the second reason why the lab will continue to play an important role relates to the nature of AI itself and of digital technology more generally. Whereas the traditional pattern is for something to be invented in a lab, then developed at a factory into a finished product for sale to the customer, digital products are characterized by a different dynamic. It is normal for their developers to remain involved in their application. Consider, for example, the difference between traditional television and a streaming service like Netflix. In the former a broadcaster airs a programme and then receives feedback from viewers that can be used to create new, improved output. With a streaming service the user remains connected to the provider's platform, and it learns from their behaviour in real time, enabling immediate adaptation. A digital product can therefore be regarded not as a finished product but as a semi-finished one. Streaming platforms, smart thermostats, healthcare apps and all other digital products are continuously adapted and improved in use.

This is reflected in the structure of the technology industry. A 'lean start-up' will quickly develop a 'minimal viable product' (MVP). In many cases this does not

[121] Domingos, 2017: 133.

[122] Domingos, 2017: 151–155.

[123] Domingos, 2017: 199.

work very well and is marred by numerous 'bugs', but is at least useable. Once in use it can learn and improve. Because of this 'the lab' in a sense remain present within the product and continues to play a major role in its further refinement. This is why collaborative initiatives like the AI labs referred to earlier, where scientists ('the lab') are in contact with businesses and/or government agencies, are common in the industry. We could even go so far as to say that with AI's transition from lab to society, the lab itself has entered mainstream society. Another way of looking at it is that society has been absorbed by the lab to become a 'living lab'. Facebook, for example, develops new services by continually running experiments involving the platform's users. It should be acknowledged, however, that this practice has sometimes proven highly controversial, as with Facebook's experiments aimed at influencing users' emotions.[124]

The particular dynamics of digital product development gives rise to a range of issues, which we consider in later chapters. One is the 'technical debt' problem, whereby it can be difficult to rectify a shortcoming in an MVP.[125] Furthermore, the development process is liable to entail a variety of risks. Because rollouts are not initially developed to end-product status, users may be exposed to something with undesirable or even harmful effects. By the time those are detected, the damage has already been done.

The semi-finished nature of AI products also presents challenges for regulators. For example, vehicles are normally tested by the responsible authorities before they can be used on public roads. This approach is workable in a world where the vehicles are end products, but not when they are semi-finished and liable to change once in use – as with a Tesla that receives a software update. The possibility of a continuous testing regime is therefore being investigated. In the healthcare sector too, the functionality and safety of devices has traditionally been tested prior to licensing on the assumption that their basic functions will not subsequently change. An AI product is constantly evolving, however, and changes can be implemented remotely. Such 'lab dynamics' therefore require a more dynamic approach to testing. We return to this topic in Part 2. For now, it is important to recognize how the lab remains involved with and indistinguishable from an AI product after its practical rollout.

The lab is thus slated to continue to play a major role in the future of AI. The limitations of the current approaches are such that further fundamental research is required, while the very nature of AI implies that the lab will always be associated with the technology's practical application. In the interests of further technical advances and AI's successful integration within society, it is therefore important to keep sight of the lab's role, to involve it in practical implementation and to ensure that it has adequate resources and talent.

[124] See, for example, NOS, 29 June 2016, and Rutkin, 25 June 2014.

[125] Marcus & Davis, 2019: 188.

Key Points – The Future of the Lab

- Although AI is now making the transition to society, the lab remains as relevant as ever. There are two main reasons for this.
- First, current AI methodologies have a variety of shortcomings. They are superficial, inefficient, lacking common sense and opaque. Fundamental research therefore remains very important to address drawbacks of this kind.
- Second, as with digital technology in general, lab research continues even after an AI product enters practical use. So in fact the lab itself enters society together with the product.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

AIV en CAVV. (2015). *Autonome Wapensystemen: De Noodzaak Van Betekenisvolle Menselijke Controle* (Nr. 97 AIV/Nr. 26 CAVV). Adviesraad Internationale Vraagstukken en Commissie van Advies Inzake Volkenrechtelijke Vraagstukken. Available at: https://www.adviesraadinternationalevraagstukken.nl/documenten/publicaties/2015/10/02/autonome-wapensystemen

Ajami, S. (2016). Use of Speech-To-Text Technology for documentation by healthcare providers. *The National Medical Journal of India, 29*(3), 148–152.

Albert Heijn. (2019, May 20). *Albert Heijn Zet Kunstmatige Intelligentie In Tegen Voedselverspilling*, Albert Heijn Nieuws. Available at: https://nieuws.ah.nl/albert-heijn-zet-kunstmatige-intelligentie-in-tegen-voedselverspilling/

Apple, C. (2020, June 24). Instant Gratification: The history of Instagram. *Spokesman*. Available at: https://www.spokesman.com/stories/2020/jun/24/how-instagram-hit-one-billion-users/

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfield, D., & Squicciarini, M. (2020). *Identifying and measuring developments in Artificial Intelligence: Making the impossible possible* (OECD Science, Technology and Industry Working Papers 20(5)). OECD Publishing.

Boden, M. (2018). *Artificial Intelligence: A very short introduction*. Oxford University Press.

Boland, H. (2018, September, 2). Britain faces an AI brain drain as tech giants raid top Universities. *The Telegraph*. Retrieved from https://www.telegraph.co.uk/technology/2018/10/24/britains-ai-industry-must-avoid-brain-drain-us-mps-warn/

Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI Frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute. Available at: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.pdf?shouldIndex=false

Bui, P., & Liu, Y. (2021, May 18). Using AI to help find answers to common skin conditions. [Blog] Google.nl. Available at: https://blog.google/technology/health/ai-dermatology-preview-io-2021/

CB Insights. (2018, April 26). *Rise of China's Big Tech in AI; what Baidu, Alibaba, and Tencent are working on*. CB Insights Research Briefs. Available at: https://www.cbinsights.com/research/china-baidu-alibaba-tencent-artificial-intelligence-dominance/

CB Insights. (2021, June 24). *Despite a Pandemic Slump, the AI sector remains hot for acquirers.* CB Insights Research Briefs. Available at: https://www.cbinsights.com/research/artificial-acquisitions-trends-annual-deals/

CEG. (2018). *Digitale Dokters: Een Ethische Verkenning Van Medische Expertsystemen*. Centrum voor Ethiek en Gezondheid. Available at: https://www.ceg.nl/binaries/ceg/documenten/signalementen/2018/07/04/digitale-dokters%2D%2D-een-ethische-verkenning-van-medische-expertsystemen/webversie_CEG_Digitale_dokters_Een_ethische_verkenning_van_medische_expertsystemen.pdf

Chen, S. (2017, October 12). China to build giant facial recognition database to identify any citizen within seconds. *South China Morning Post*. Available at: https://www.scmp.com/news/china/society/article/2115094/china-build-giant-facial-recognition-database-identify-any

Chen, Y., Casagrande, N., Zhang, Y., & Brenner, M. (2019). *Using Wavenet Technology to Reunite speech-impaired users with their original voices*, DeepMind.nl, 18 December. Available at: https://deepmind.com/blog/article/Using-WaveNet-technology-to-reunite-speech-impaired-users-with-their-original-voices

Chiusi, F., Fischer, S., Kayser-Bril, N., & Spielkamp, M. (2020). *Automating Society Report 2020*. AlgorithmWatch. Available at: https://www.ivir.nl/publicaties/download/Automating-Society-Report

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Dauverge, P. (2020). *AI in the Wild – Sustainability in the age of Artificial Intelligence*. MIT Press.

Davidson, D., & Delhaas, R. (2020, April 22). *Als De Politiek In Ieder Oor Een Andere Belofte Fluistert*. Argos. Available at: https://www.vpro.nl/argos/lees/nieuws/2020/microtargeting-in-Nederland.html

DeepMind. (2019). *Machine learning can boost the value of wind energy*, Deepmind.nl, 26 February. Available at: https://deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy

Delcker, J. (2018, June 27). *Merkel warns of AI brain drain to foreign tech companies*. Politico. Retrieved from: https://www.politico.eu/article/merkel-artificial-intelligence-warns-brain-drain-to-foreign-tech-companies/

DenkWerk. (2018). *Artificial Intelligence in Nederland: Zelf Aan Het Stuur*. Available at: https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf

Dickson, B. (2020, March 2). Understanding the limits of CNNs, one of AI's Greatest Achievements. *TechTalks*. Available at: https://bdtechtalks.com/2020/03/02/geoffrey-hinton-convnets-cnn-limits/

Domingos, P. (2017). *The master algorithm: How the Quest for the ultimate learning machine will remake our world*. Penguin Random House.

ELLIS. (2018). *Open letter*. ELLIS Society. Retrieved from: https://ellis.eu/letter

Elsevier. (2018). *ArtificiaI Intelligence: How knowledge is created, transferred, and used*. Elsevier. Retrieved from: https://www.elsevier.com/connect/resource-center/artificial-intelligence

European Commission. (2020). *A European Strategy for Data*, COM(2020) 66 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX:52020DC0066

European Commission. (2021a). *2030 Digital Compass: the European way for the Digital Decade*. Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118

European Commission. (2021b [2018]). *EU coordinated action plan on AI 2021 review*, COM(2021) 205 final. Available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-reviewhttps://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

European Commission. (2021c). *Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Commission. (n.d.). *Destination Earth*. Available at: https://digital-strategy.ec.europa.eu/en/policies/destination-earth

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

Freeman, K., Dinnes, J., Chuchu, N., Takwoingi, Y., Bayliss, S. E., Matin, R., Jain, A., Walter, F., Williams, H., & Deeks, J. (2020). Algorithm based smartphone Apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies. *British Medical Journal, 368*(127), m127.

Hern, A. (2019, July 30). Cambridge Analytica did work for leave. EU, Emails confirm. *The Guardian*. Available at: https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm

High-Level Expert Group on Artificial Intelligence. (2019a). *Ethics Guidelines For Trustworthy AI*. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on Artificial Intelligence. (2019b). *Policy and investment recommendations for trustworthy AI*. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

Hoeks, G. (2019, April 12). Europese Wetenschappers In Verweer Tegen Braindrain Kunstmatige Intelligentie. *Het Financieele Dagblad*. Available at: https://fd.nl/economie-politiek/1297131/europese-wetenschappers-in-verweer-tegen-braindrain-kunstmatige-intelligentie-lnl1ca8zhDaO

Holoniq. (2020, April 9). *The 2020 AI strategy landscape: 50 National Artificial Intelligence strategies shaping the future of humanity*. Available at: https://www.holoniq.com/notes/50-national-ai-strategies-the-2020-ai-strategy-landscape/

Kamerstukken II 2019/2020 26 643, nr. 641. (2019, October 8). *Waarborgen Tegen Risico's Van Data-Analyses Door De Overheid*, Kamerbrief. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/10/08/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid.pdf

Kamerstukken II 2019/2020 26 643, nr. 642. (2019, October 8). *AI, Publieke Waarden En Mensenrechten*, Brief regering. Available at: https://www.tweedekamer.nl/downloads/document?id=1a5131a6-0f2e-4b5e-917f-8f3e2cf0a144&title=AI%2C%20publieke%20waarden%20en%20mensenrechten.pdf

Kayser-Bril, N. (2019, December 11). *At least 11 police forces use face recognition in the EU, AlgorithmWatch reveals*. AlgorithmWatch. Available at: https://algorithmwatch.org/en/face-recognition-police-europe/

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.

Lawton, G. (2019, October 2). AI can predict your future behaviour with powerful new simulations. *New Scientist*. Available at: https://www.newscientist.com/article/mg24332500-800-ai-can-predict-your-future-behaviour-with-powerful-new-simulations/

Lee, K. F. (2018). *AI Superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Lewis, P., & Hilder, P. (2018, March 23). Leaked: Cambridge Analytica's Blueprint for Trump victory. *The Guardian*. Available at: https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory

Lewis-Kraus, G. (2016). The Great A.I. Awakening. *The New York Times Magazine*. Available at: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Liu, X., Faes, L., Kale, A., Wagner, S., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J., Schmid, M., Balaskas, K., Topol, E., Bachmann, L., Keane, P., en Denniston, A. (2019). 'A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis', The Lancet Digital Health, 1: e271-e297.

Loucks, J., Hupfer, S., Jarvis, D., & Murphy, T. (2019). *Future in the balance? How countries are pursuing an AI advantage*. Deloitte Center for Technology, Media & Telecommunications. Available at: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/public-sector/lu-global-ai-survey.pdf

Marcus, G. (2018). *Deep learning: A critical appraisal*. Available at: https://arxiv.org/pdf/1801.00631.pdf?ut

Marcus, G., & Davi, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. Vintage.

Marsh, H. (2019, January 10). Can man ever build a mind? *Financial Times*. Available at: https://www.ft.com/content/2e75c04a-0f43-11e9-acdc-4d9976f1533b

McKinsey & Company. (2020). *How nine digital front-runners can lead on AI in Europe*. McKinsey & Company. Available at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/how%20nine%20digital%20frontrunners%20can%20lead%20on%20ai%20in%20europe/how-nine-digital-frontrunners-can-lead-on-ai-in-europe.pdf

Misuraca, G., en van Noordt, C. (2020). *AI Watch – Artificial Intelligence in public services: Overview of the use and impact of AI in Public Services in the EU*. Publications Office of the European Union.

Mols, B. (2019). *Internationaal AI-beleid. Domme data, slimme computers en wijze mensen*. WRR Working Paper.

Mozur, P. (2019, April 14). One Month, 500,000 Face Scans: How China is using A.I. to profile a minority. *The New York Times*. Available at: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

NOS. (2016, June 29). *Stiekem experiment Op Facebook*. NOS. Available at: https://nos.nl/artikel/668173-stiekem-experiment-op-facebook

OECD. (2018). *Private Equity Investment in Artificial Intelligence*. OECD Publishing.

OECD. (2019). *Artificial Intelligence in Society*. OECD Publishing.

O'Neil, C. (2016). *Weapons of Math destruction: How Big Data increases inequality and threatens democracy*. Penguin.

PBL. (2017). *Mobiliteit En Elektriciteit In Het Digitale Tijdperk. Publieke Waarden Onder Spanning*. Planbureau voor de Leefomgeving. Available at: https://www.pbl.nl/sites/default/files/downloads/pbl-2017-mobiliteit-en-elektriciteit-in-het-digitale-tijdperk-1874.pdf

Pearl, J. (2019). The limitations of opaque learning machines. In J. Brockman (red.) *Possible minds: Twenty-five ways of looking at AI* (pp. 13–19). Penguin.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. (2019). *The AI Index 2019 Annual Report*. Stanford University, Human-Centered AI Institute. Available at: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

Politie. (2018, May 23). *Nieuwe technologie in oude politiezaken*. Available at: https://www.politie.nl/nieuws/2018/mei/23/00-nieuwe-technologie-in-oude-politiezaken.html

Prins, C. (2017). Politiek Profileren. *Nederlands Juristenblad, 92*(38), 2799.

Rao, A., & Verweij, G. (2017). *Sizing the Prize: What's the real value of AI for your business and how can you capitalise?* PricewaterhouseCoopers. Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.

Rutkin, A. (2014, June 25). Even online, emotions can be contagious. *New Scientist*. Available at: https://www.newscientist.com/article/mg22229754-900-even-online-emotions-can-be-contagious/?ignored=irrelevant#.U7EHbI21au8

RVS. (2019). *Waarde(N)Volle Zorgtechnologie. Een Verkennend Advies Over De Kansen En Risico's Van Kunstmatige Intelligentie In De Zorg*. Raad voor Volksgezondheid en Samenleving.

Sample, I. (2019, October 24). Human Compatible By Stuart Russell Review – AI and our future. *The Guardian*. Available at: https://www.theguardian.com/books/2019/oct/24/human-compatible-ai-problem-control-stuart-russell-review

Schwab, K. (2016). *The Fourth Industrial Revolution*. Random House.

Simonite, T. (2019, September 3). Behind the rise of China's Facial-Recognition Giants. *Wired*. Available at: https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/

Smith, C. (2013, September 18). Facebook users are uploading 350 million new photos each day. *Business Insider*. Available at: https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?IR=T

Šopova, J. (2018). Audrey Azoulay: Making the most of Artificial Intelligence. *The UNESCO Courier, 3*, 36–41. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000265211

Thomas, D. (2021, February 9). Is this Beverly Hills Cop playing Sublime's 'Santeria' to avoid being live-streamed?. *VICE*. Available at: https://www.vice.com/en/article/bvxb94/is-this-beverly-hills-cop-playing-sublimes-santeria-to-avoid-being-livestreamed

Tian, H., Wang, T., Yadong, L., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation – A review. *Information Processing in Agriculture, 7*(1), 1–19.

Tonin, M. (2019). Artificial Intelligence: Implications for NATO's Armed Forces. *149 stctts 19 E rev. 1 fin.*

Topol, E. (2019). High-performance medicine: The convergence of human and Artificial Intelligence. *Nature medicine, 25*(1), 44–56.

van Buchem, M., Boosman, H., Bauer, M., Kant, I., Cammel, S., & Steyerberg, E. (2021). The digital scribe in clinical practice: A scoping review and research Agenda. *NPJ Digital Medicine, 4*(57), 1–8.

Vorst, T. van der, Jelicic, N., de Vries, M. en Albers, J. (2019). *De (On)Mogelijkheden Van Kunstmatige Intelligentie In Het Onderwijs, Nr. 2018.068.1828*. Dialogic. Available at: https://www.dialogic.nl/wp-content/uploads/2019/04/Dialogic-De-onmogelijkheden-van-kunstmatige-intelligentie-in-het-onderwijs-v1.0.116.pdf

van Roy, V., Rossetti, F., Perset, K., en Galindo-Romero L. (2021). *AI watch – National strategies on Artificial Intelligence: A European Perspective* (EUR 30745): Publications Office of the European Union.

van Veenstra, A. F., Djafari, S., Grommé, F., Kotterink, B., & Baartmans, R. (2019). *Quick scan AI in de Publieke Dienstverlening*. TNO. Available at: www.rijksoverheid.nl/documenten/rapporten/2019/04/08/quick-scan-in-de-publiekedienstverlening

Waardenburg, L., Sergeeva, A., & Huysman, M. (2020). Predictive policing Ontcijferd: Een Etnografie Van Het 'Criminaliteits Anticipatie Systeem' in De Praktijk. In J. Janssens, W. Broer, M. Crispel, and R. Salet (reds) *Informatiegestuurde politie* (Cahiers Politiestudies 54) (pp. 69–88). Gompel & Svacina.

Waldrop, M. (2019). News feature: What are the limits of deep learning? *Proceedings of the National Academy of Sciences, 116*(4), 1074–1077.

Wilson, H., Daugherty, P., & Davenport, C. (2019, January 14). The future of AI will be about less data, not more. *Harvard Business Review*. Available at: https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more

WIPO. (2019). *WIPO Technology Trends 2019: Artificial Intelligence*. World Intellectual Property Organization.

Wojciki, S. (2020, February 14). YouTube at 15: My personal journey and the Road Ahead, *blog*. Available at: https://blog.youtube/news-and-events/youtube-at-15-my-personal-journey

Wouda, F., & Hutink, H. (2019). *Artificial Intelligence In De Zorg: Begrippen, Praktijkvoorbeelden En Vraagstukken*. Nictiz. Available at: https://www.nictiz.nl/wp-content/uploads/Rapport_artificial_intelligence_in_de_zorg.pdf

WRR. (2016). *Big data In Een Vrije En Veilige Samenleving*. Amsterdam University Press.

Yu, K., Beam, A., & Kohane, I. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering, 2*(10), 719–731.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *The AI index 2021 annual report*. Stanford University, Human-Centered AI Institute. Available at: https://arxiv.org/ftp/arxiv/papers/2103/2103.06312.pdf

Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Fataigh, R., Irion, K., Dobber, T., Bodo, B., & De Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review, 14*(1), 82–96.

Chapter 4
AI as a System Technology

Now we know what AI is and have seen how the technology has made the transition from the lab to society in recent years, we turn our attention to the process of embedding AI into society. What is required to incorporate AI into our society? To answer that question, this chapter presents a framework within which AI can be viewed as a particular type of technology, namely a system technology, with a number of historical precedents. By viewing AI in this way, we can draw various conclusions from the history of other system technologies. That in turn provides a basis for reflecting on what we need to do with AI and how we can address the many issues associated with it. It is not our intention to imply that history always repeats itself or that technological development has deterministic characteristics. We do not set out a rigid framework but identify general patterns that shed light on the present, while recognizing that the past and the present differ. By adopting this approach, we seek to look beyond the current situation and thus beyond the whims of the day.

Various prominent commentators have drawn parallels between AI and other technologies. According to researcher Andrew Ng, the impact of AI is "comparable to that of electricity a century ago."[1] Google's CEO Sundar Pichai and his predecessor Eric Schmidt have also compared AI with electricity. Indeed, Pichai went so far as to liken AI to fire.[2]

In the policy world too, similar comparisons are often made. In a paper on the strategic implications of AI, Michael Horowitz wrote that it is not an isolated technology but similar to general-purpose technologies such as electricity and the internal combustion engine.[3] The breadth of AI's potential applications has also been highlighted by the European Commission's European Political Strategy Centre: "It

[1] Lynch, 4 May 2021.

[2] Goode, 19 January 2018; Morozov, 2013: 1.

[3] Horowitz et al., 2018.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_4

is hard to think of any sector of society that will not be transformed by AI in the years ahead."[4] The EU accordingly regards AI as a 'key enabling technology' (KET).

In short, many authors and organizations have hinted at similarities with earlier technologies. But few have gone on to make a detailed comparative analysis. We therefore seek to do that in this chapter. To that end we consider the implications of placing AI in the same bracket as technologies such as electricity. We discuss the literature on various types of technology, focusing particularly on the concept of general-purpose technologies, and we introduce the term 'system technology'. By tracing the historical development of system technologies, we identify a number of general patterns in the way they are embedded in society. From there we define five overarching tasks associated with the process of societal integration. In Part 2 of this report, we consider how each of the five overarching tasks applies to the embedding of AI within society.

4.1 Classification of Technologies

Academics have long been interested in how different types of technology exert a general influence over the economy and society. An early example is the Kondratiev wave theory, in particular as elaborated by Joseph Schumpeter. He observed that periods of high economic growth alternate with periods of lower growth, a pattern he attributed to the effect of new technologies; sets of new technologies periodically boosted growth, after which the effect gradually waned over time. According to Schumpeter, such dynamism was inherent to a capitalist market economy: "it is essential to understand that capitalism is an evolutionary process … 'industrial mutation' … is constantly bringing about revolutionary change to the structure of the economy from the inside, constantly destroying the old structure and creating a new one."[5] Such reasoning forms the basis of the familiar concept of creative destruction.

In his acceptance speech when presented with the Nobel Prize for Economics in 1971, Simon Kuznets introduced the idea of 'epochal innovations' driving periods of great economic development. Innovation scientists Carlota Perez and Chris Freeman have written about a similar phenomenon, which they refer to as 'new technology systems' and 'technological revolutions'.[6] A new technology system is a powerful and conspicuous cluster of new and dynamic technologies, products and industries that lead to major change throughout the economy and ultimately to economic growth. Perez has identified five such clusters since the Industrial Revolution, including the eras of steam power and railways, of steel and electricity, of oil, cars and mass production and of information and telecommunications. She argues that

[4] European Political Strategy Centre, 2018.

[5] Quotation from Joseph Schumpeter in Juma, 2016: 17.

[6] Freeman & Louçã, 2001: 144.

each of these brought its own 'techno-economic paradigm': a way of thinking and acting, leading to the relevant technologies becoming integral to the fabric of society.[7] Alessandro Nuvolari has made a significant addition to Perez's theories by emphasizing that the observed effects are attributable not so much to individual technologies as to blocks of radical innovations that together bring about revolution.[8] Some researchers accordingly take the view that innovation consists not of the development of major new things but of the combination of things that already exist.[9]

4.1.1 General-Purpose Technologies

AI can be classified based on its general, transformative impact on society. It is useful to view the technology in relation to the concept of the general-purpose technology (GPT). GPTs are technologies whose potential applications are not specific, like those of a lawnmower, toaster or microscope, but generic insofar as they lend themselves to countless, highly diverse purposes. GPTs can therefore have a major influence on the economy and society. Timothy F. Bresnahan and Manuel Trajtenberg introduced the concept in an article published in 1992,[10] which cited three criteria for the classification of a technology as a GPT. First, a GPT is highly pervasive, being utilized in numerous sectors, production processes and products. Second, there is great scope for its technical improvement, meaning that the cost of the technology keeps falling and its efficiency increasing. Third, a GPT spawns numerous 'innovational complementarities', leading to generalized economic productivity improvements.

A large body of literature on the concept of the GPT is now available. However, that has not led to the adoption of a uniform definition[11] or consistent use of the term. Some authors recognize only a small number of historical GPTs, while others argue that there have been many throughout human history, going back as far as the domestication of livestock and the forging of bronze. One author suggests that the literature identifies twenty-eight technologies as GPTs.[12]

Another topic of debate amongst academics is the existence of technologies that have great societal impact but are not particularly generic. Examples include the printing press and the steamship: technologies whose applications are limited but

[7] Perez, 2003: 8–11. Belgian economist Luc Soete is also active in this field.

[8] Nuvolari, 2019.

[9] Brynjolfsson & McAfee, 2014: 78.

[10] Bresnahan & Trajtenberg, 1995.

[11] For example, Gavin Wright (2000) defines GPTs as "deep new ideas or technologies that have the potential to significantly influence numerous sectors of the economy".

[12] See the Working Paper Artificial intelligence as a general-purpose technology – Strategic interests in responsible use in a historical perspective (Bakker & Korsten, 2021) produced by Freedomlab for the WRR.

have radically changed society. The technologies most widely recognized and cited as GPTs are the steam engine, electricity, the internal combustion engine and IT.[13]

Notwithstanding the qualifications made above, a number of interesting studies in recent years have related AI explicitly to the concept of the GPT. Following a conference organized by the US National Bureau of Economic Research (NBER) in 2017, for instance, a collection of papers entitled The Economics of Artificial Intelligence was published in 2019. The first part, entitled 'AI as a GPT', includes contributions by renowned technology researchers and economists and contains various interesting analyses that we draw on in this report, although – in keeping with the nature of the original conference, but in contrast to our own focus – they are concerned primarily with the macroeconomic effects of AI.

Also of interest in this context is the thesis by Jade Leung of Oxford University, entitled Who will govern artificial intelligence? Learning from the history of strategic politics in emerging technologies. In this she places AI alongside aerospace technology, biotechnology and cryptography as an example of what she calls 'strategic GPTs', and in that context emphasizes the relationship between governments and new technologies, particularly in the defence sector. Leung identifies three key actors here, the government, business and the research community, and demonstrates that each has different aims, instruments and limitations, which may converge in certain phases but are at odds in others.

Various researchers have recently sought to place AI in a broad historical perspective without making explicit use of the term 'GPT'. In a polemic on Andrew MacAfee and Erik Brynjolfsson's famous book The Second Machine Age, Carlota Perez wrote a nine-part series of articles entitled Second Machine Age or Fifth Technological Revolution? In these she explores how today's digital technologies – including AI – compare with previous technologies.[14]

All of these studies, and particularly the perspective developed in them, are relevant to the theoretical framework we use to view AI in this report. We additionally draw on a number of more empirical studies of the effects of specific technologies. Sarah A. Seo has written about the best-known application of the internal combustion engine, namely the motor car. She also demonstrates how this symbol of freedom has simultaneously led to an enormous increase in the power that the state – particularly the police – have over citizens' private lives.[15] In a general survey of a series of technologies ranging from tractors and margarine to electricity and GMOs, Calestous Juma investigates the dynamics of social resistance to new technologies.[16] This report thus draws not only on research into GPTs but also analogies

[13] For a critical analysis of the various uses of the term 'GPT', see Field, 2008.

[14] Perez, 2017–2020.

[15] Seo, 2019.

[16] Juma, 2016.

with recent technologies such as genetically modified organisms (GMOs) and nanotechnology, which have interesting parallels with AI.[17]

4.1.2 AI as a GPT

The question we need to ask at this point is whether AI can actually be regarded as a GPT. But the answer seems quite clear: although its global impact is currently in its early stages, it already appears that AI is indeed a GPT. If we consider Bresnahan and Trajtenberg's three criteria for classification as a GPT, a strong case can be made for saying that all apply to AI.

The first of these criteria is pervasiveness. Although AI's perfusion of the economy and wider society has gathered pace only in recent years, the technology is already used in a variety of sectors and products. Earlier in this part of the report (2.2), we presented a range of examples illustrating how AI is being used in manufacturing, agriculture, the public sector, entertainment, financial services and medical practice. Given that versatility, it is already apparent that AI is well on the way to pervading society and the economy.

The second criterion is inherent potential for technical improvement, leading to lower cost and increased efficiency. Again, it is evident that AI passes this test. In Chap. 1 we highlighted how Moore's Law states that computing power doubles every 2 years, opening the way for the further improvement of AI technologies. We also saw how scientific research has fuelled the development of new and improved technologies. As a result, the application of AI has passed numerous milestones in recent years. Furthermore, as highlighted in our discussion of the future of the lab, promising new technologies are being developed, which are expected to further boost the performance and efficiency of AI.

Finally, classification as a GPT depends on the presence of complementary innovations that lift general productivity. Numerous signs of a positive influence on general productivity can already be discerned, but AI is simply too young for us to demonstrate conclusively the existence of complementary innovations. Nevertheless, various authoritative research bodies and consultancies, including Accenture, PwC, McKinsey and Deloitte, have forecast major productivity increases over the decade ahead. We set out the three defining characteristics of system technologies in Fig. 4.1.[18]

[17] In its report Health significance of nanotechnologies, the Health Council of the Netherlands describes nanotechnology from the perspective of an enabling technology, an approach that has interesting intersections with our discussion of GPTs.

[18] For a detailed analysis of AI as a GPT, see Bakker & Korsten, 2021.

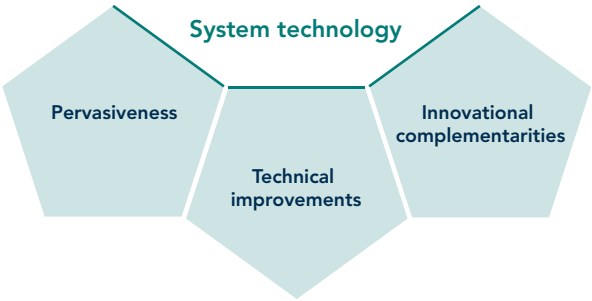


Fig. 4.1 The three defining characteristics of system technologies

4.1.3 AI as a System Technology

We can conclude, then, that AI satisfies the three criteria for classification as a general-purpose technology. The GPT concept and the wealth of literature considering AI as such a technology provide useful starting points to understand what kind of technology we are dealing with. Nevertheless, we have chosen not to apply the term 'GPT' here. Rather, we have elected to define AI as a 'system technology'. That choice reflects significant focal differences between our analysis and the literature on GPTs.

Firstly, the GPT literature from the earliest Kondratiev wave sources to the recent NBER study has a strong focus on the macroeconomic effects of the technologies in question. Many researchers seek to quantify the effects of the technologies they study. That gives rise to debate as to whether and how a GPT can be shown to support a prolonged increase in economic growth. Given the huge number of variables to be accounted for, a model capable of demonstrating such an effect has to be extremely complex. By contrast, we have chosen to concentrate not on the quantitative effects of system technologies but primarily on the qualitative changes they bring about.

Secondly, the literature on GPTs pays particular attention to historical classifications. As indicated earlier, there is considerable debate as to how many historical technologies may be considered GPTs. One researcher recognizes dozens, Perez distinguishes five clusters, authors such as Chandler refer to three Industrial Revolutions,[19] Schwab identifies four and Brynjolfsson and McAfee speak of two 'machine ages'. Furthermore, many authors make use of highly schematic timelines with precise start and end dates for individual technologies. This report differs from those approaches in that we refrain from introducing such demarcations. Because we are concerned mainly with qualitative impact rather than quantitative effects, we do not need to commit ourselves to a strict classification system or definite start and end dates for technologies. What we are seeking to do is highlight general patterns. To that end we concern ourselves primarily with a small number of previous system technologies – the steam engine, electricity, the internal combustion engine and the

[19] Freeman & Louçã, 2001: 145.

computer – and draw pragmatically on historical sources to identify relevant parallels.

Another reason for not adopting the term 'GPT' is that it emphasizes a technology's versatility. We prefer 'system technology' because we wish to emphasize the systemic nature of certain technologies and to broaden the focus to their systemic effects on society. In the context of 'system technology', therefore, the word 'system' has two implications. First, it implies that the technology consists of a system with multiple components. Electricity, for example, works in conjunction with generators, cables, batteries and so on. Similarly, AI is part of a wider technical system of data and hardware. The second implication of 'system technology' is that the technology influences a variety of systems and processes within society. Exercising such influence involves a complex process of adaptation, trial and negotiation. In other words, our chosen term reflects the process of societal integration and the associated qualitative effects.

4.1.4 Similarities and Differences Between AI and Earlier System Technologies

AI is a system technology and therefore comparable with earlier technologies of that type such as the steam engine, electricity and the internal combustion engine (Fig. 4.2). Furthermore, we can define AI even more precisely given that particular characteristics make it more similar to one technology than another in certain respects. For example, the internal combustion engine and the steam engine are tangible, whereas AI is like electricity in being intangible to some extent. It does not exist in isolation but only as part of a product or service. In that sense, devices such as toasters, lamps and radios that work by means of electricity are comparable with thermostats, watches and machines that work by means of AI.

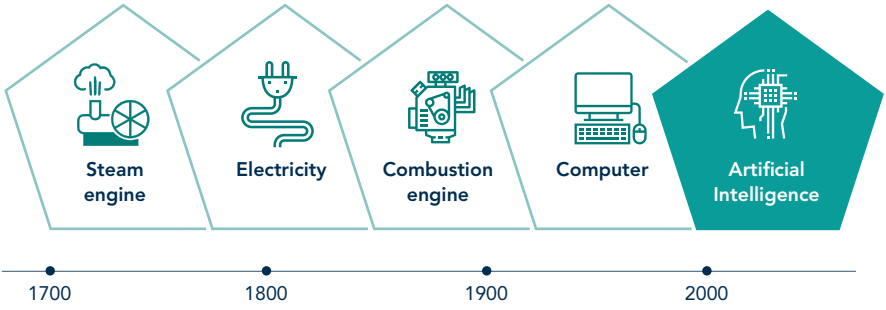


Fig. 4.2 AI as a new system technology

Another respect in which AI is more like electricity than the internal combustion engine is that it is 'technology-radical', rather than 'use-radical'. The descriptor 'technology-radical' is applied to technologies driven primarily by technical and scientific progress; their development is propelled by the curiosity of researchers, without any clear notion of how or for what purpose the technology will ultimately be used. By contrast, 'use-radical' implies a clear understanding of the applications from the outset, with commercial factors playing a role early on. The development of use-radical technologies is goal-oriented. That was the case with the internal combustion engine. Like with electricity, researchers were working on AI long before people recognized the lucrative applications we are now aware of.

A distinction can also be drawn between system technologies in which governments play an obvious role from the start and those whose development has no such involvement. The first group includes technologies developed specifically for defence purposes and dependent on the defence sector for their further application and development. This differentiates them from 'civilian-first' technologies, whose development is attributable mainly to their economic potential. Governments have more control over the development of the first group of technologies than the second. Space technology is an example of a technology developed with direct government involvement, while biotechnology is an example of one whose development fits the second model. However, both are examples of what Jade Leung calls strategic GPTs.[20]

With AI, the US military institute DARPA was a key financier in the early stages. Nevertheless, early government-funded AI research was of a fundamental nature and military applications represent only a small portion of its full range of uses. In that respect AI is more akin to biotechnology than aerospace technology. However, it differs from biotechnology insofar as the latter's developers are largely attached to major (academic) laboratories, whereas innovation in the field of AI is more decentralized. That has implications for researchers' ability to define universal standards.

It is important to consider not only such technical similarities and differences between AI and other system technologies, but also how AI compares in terms of its societal and temporal context. Take the role of the government, for example. The steam engine was developed in a laissez-faire climate in the UK, with the government playing only a very limited role. On the other hand, the combustion engine and the motor car were developed in an era when government economic policy was led by Keynesian thinking. Although governments now exercise considerable influence over the economy by means of standardization and legislation, AI emerged at a time when there was significant resistance to strong guidance of the economy. It is important to bear those circumstances in mind when seeking to identify historical patterns that are instructive in relation to AI.

The societal context of AI also differs from that of earlier technologies in terms of the mobilization of social actors. Increasing prosperity and the progress of

[20] Leung, 2019.

democratization have empowered more people to express themselves in the public arena. Whereas enterprises and governments could once shape society with relative ease, nowadays civil society, the academic community, individuals and the media have much more influence than in the past. The mobilization of these actors therefore plays a more significant role in relation to AI and its integration into society than it did in relation to earlier system technologies.

This phenomenon ties in with what Trajtenberg calls the 'democratization of expectations': factory workers during the Industrial Revolution had little power because most struggled to make ends meet. We return to this point in Sect. 4.5, in relation to the Luddites. Today far more people participate in public life and workers have much better representation. Moreover, people are less inclined to bear the cost of technological change while also having greater expectations in terms of sharing in the benefits of such change.[21]

The world today is not only more democratized than in the past but also more globalized. Consequently, the issues associated with AI have always been more global. The extensive nature of modern markets and the consequently wide geographical impact of AI's applications are relevant in this context, as is the existence of all manner of international constraints such as trade agreements, human rights and technical standards. Interestingly enough, the rise of earlier system technologies has often been an impulse for the formation of new international organizations for standardization,[22] and these are now playing a role in relation to AI. Examples include entities active in the fields of telecommunications and the internet, standardization bodies such as the ISO and international engineering associations such as the IEEE. Although the development and embedding of earlier technologies had an international dimension, the significance of that dimension has increased over time under the influence of globalization.

One final difference between the development context of AI and that of earlier system technologies is the increased level of organization and communication amongst scientists. The scientific community was not well integrated at the time of the steam engine's development, whereas academic organizations, codes of conduct and standards now exert significant influence.

4.1.5 The Techno-Economic Paradigm of AI

Finally, Carlota Perez's notion of the techno-economic paradigm warrants attention.[23] She argues that major technological change leads not only to new products and services but also to new ways of thinking and working and new principles of organization. For example, the Industrial Revolution led to the rise of factories

[21] Trajtenberg, 2018: 178.

[22] Kaiser & Schot, 2014.

[23] Perez, 2003.

while electricity enabled 'networked' production. Similarly, the invention of the internal combustion engine gave us not only cars but also the conveyor belt. Fordism, Taylorism and just-in-time production are all derived from organizational principles. Although it is too early to characterize the techno-economic paradigm of AI in definitive terms, we can already discern certain outlines that follow earlier forms of digitalization but also exhibit new features.

We believe that three aspects of the techno-economic paradigm of AI are already distinguishable. The first relates to changes in the nature of objects and products. As discussed at the end of the previous chapter, in the digital domain we are dealing not so much with end products as with semi-finished ones. A digital product is never finished. Unlike traditional products and services, which ultimately leave the factory and are sold, digital products are constantly being revised and adapted. By means of updates, digitally-enabled objects such as computers, cars, cameras and medical devices are always changing. In the words of Kevin Kelly, everything is in a continuous 'state of becoming'.[24] Or, as Luciano Floridi puts it, 'things' are being replaced by '-ings', such as interact-ing, process-ing, network-ing, do-ing and be-ing.[25]

Related to this is the phenomenon that physical objects that acquire a digital aspect cease to be discrete entities. In this regard Adam Greenfield highlights porosity as a common characteristic of modern-day technologies. The boundaries between objects and between user and platform, and even the walls of our homes, have become porous due to bilateral interconnection and intermingling. Numerous actors are therefore involved with and present in all those products. The changes to the nature of physical objects raise a variety of security, privacy and responsibility questions.

A second feature of the technical paradigm of AI is, paradoxically, that while objects associated with individuals are becoming more transparent, much of the technology is becoming invisible. At a meeting in Davos in 2015, Eric Schmidt predicted that the internet would disappear. He did not mean that it would fall into decline but was referring to an idea derived from an influential 1991 article by Mark Weiser entitled The Computer for the twenty-first Century.[26] That introduced the concept of 'ubiquitous computing', an omnipresent architecture of digital technology. According to Weiser, "The most prominent technologies are those that disappear. They become integral to the fabric of daily life, with the result that we cease to be aware of them." Hence, "computers can disappear into the background".[27]

Luciano Floridi has made the same point using a metaphor. He suggests that we are now living on the 'piano nobile', the central upper storey of a Renaissance home visible from the outside. However, below us are numerous servants – in our case

[24] Kelly, 2017: 9–27.

[25] Floridi, 2014: 183.

[26] Zuboff, 2019: 200.

[27] Weiser, 1991.

digital servants – at work in the service rooms.[28] An interesting feature of this spatial metaphor is that it emphasizes the existence of a vertical structure. The building has multiple superimposed levels, not all of which are visible. Benjamin Bratton sees in such verticality the core of digital technology.[29] He argues that we used to live in a horizontal world, with people, objects and countries adjacent to each other on the map. Digitalization has introduced a vertical structure, however, the layers of which are formed by internet addresses, cloud services and data centres running through everything largely unnoticed. In the world of technology, the 'stack' is a familiar concept: an entity made up of superimposed layers of hardware, software, network and applications. The existence of that largely invisible layering raises questions regarding power relationships and dependencies.[30] Jose van Dijck uses another metaphor to describe the vertical structure of digital technology. She refers to the tree-like structure of platformization, focusing attention on the power concentration associated with, for example, vertical integration.[31]

One final aspect of the technical paradigm of AI that warrants attention relates to Floridi's concept of technology. He argues that the idea of technology as an instrument is problematic because it suggests that a person uses an instrument, and by doing so exercises influence over the outside world. That obscures the fact that much of our technology today acts not act on an external physical world but on other technologies. Our attention should therefore be directed towards that 'inter-technological' dynamism. Floridi calls technologies that act on other technologies 'second-order technologies'. One example is a brake, which acts on the wheel of a car. In that case the process is activated by a person pressing the brake pedal.

However, the world of AI is complicated by the existence of 'third-order technologies': technologies that cause other technologies to act on yet other technologies, without human intervention. In an autonomous vehicle, for example, the decision to activate the brake is taken by the vehicle's control system. Wherever an AI system can make decisions autonomously, a third-order structure may be formed. Many road-traffic penalty systems are already characterized by such structures. A vehicle is photographed infringing a traffic regulation – breaking a speed limit, for example – triggering the issue of a penalty notice, which is sent to the address of the vehicle's registered keeper. The autonomy that technology acquires with the integration of algorithms and that is ultimately integral to the definition of AI used in this report gives rise to questions about matters such as human control, responsibility and legal liability.[32]

---

[28] Floridi, 2014: 37.

[29] Bratton, 2016.

[30] The AI Now Institute has also made an extensive study of the invisible layers of AI: from human algorithm trainers in other countries to the material requirements that lead to all sorts of raw material supply chains. See for example Joler & Crawford, 2018.

[31] Van Dijck, 2020: 1–19.

[32] Hage, 2017.

Key Points – AI as a System Technology

- There is a large body of literature characterizing innovative technologies as 'epochal innovations', 'technological revolutions' and 'general-purpose technologies'. A general-purpose technology (GPT) is distinguished by pervasiveness, great potential for technical improvement and complementary innovations. AI has all three characteristics.
- In this report, we refer to AI as a 'system technology'. Unlike the literature on GPTs, we do not apply a rigid classification, but instead emphasize qualitative characteristics and their impact on society.
- As a system technology, AI is comparable with technologies such as the steam engine, the internal combustion engine and electricity. In some respects, it resembles electricity more than the others. Over time, the societal context in which system technologies develop has changed.
- AI is associated with a distinct techno-economic paradigm characterized by continuous change to products and services, a largely hidden vertical structure of hardware and software, and the potential for technology to act autonomously.

4.2 The Societal Integration of System Technologies

Having defined AI as a system technology, we now consider what is required for its integration into society. By analysing the history of earlier system technologies, we identify a number of characteristic patterns that can be instructive in relation to AI. In this section we consider the general lessons of the past as a precursor to examining the five overarching tasks a society faces in relation to a new system technology.

4.2.1 Co-evolution of Society and Technology

Initially, a process of societal integration or 'embedding' involves prolonged co-evolution of the society and the technology concerned. Such a process requires practice, experimentation and negotiation, all of which take time. That immediately places the strongly polarized debate regarding AI in perspective. There are, for example, techno-optimists who believe that AI can fundamentally enrich society in the short term with autonomous vehicles, sophisticated medical diagnoses and automated production systems. By contrast, sceptics argue that AI is overhyped and highlight the lack of evidence regarding the technology's impact to date and the continual revision of predictions about matters such as the speed at which applications like autonomous vehicles will be realized.

Both viewpoints contain an element of truth. As the optimists suggest, AI has many potential applications. However, it would be a mistake to suppose that integrating these into society will be straightforward. Sceptics rightly draw attention to the problems that AI presents in the foreseeable future, but those problems do not justify generalized scepticism about the technology. A system technology necessitates a bilateral process of social and technological adaptation and that takes time, even in the modern era of rapid technological development and globalization. Although technologies nowadays spread around the world more quickly than in the past, embedding them, ensuring that they work and that people trust them all depend on societal processes that are not necessarily faster-moving now than they used to be. Such processes tend to proceed in fits and starts, and often span decades.

4.2.2 Unpredictable Development and Impact

A related observation is that the introduction of a new system technology is to a large extent an unpredictable process. New technologies are often used for purposes other than those for which they were originally intended or initially adopted. Don Ihde accordingly refers to 'multistability' and Wiebe Bijker to 'interpretative flexibility'.[33] Cars were originally used for sport and medical purposes in the belief that the 'thin air' breathed when driving at speed was good for the lungs.[34] Similarly, Thomas Edison did not develop the gramophone with entertainment in mind but envisaged his 'talking machine' as a business tool akin to a dictaphone.[35]

Shoshana Zuboff refers to the inability to predict accurately how a technology will be used and the consequent underestimation of its effects as 'horseless carriage syndrome'.[36] Major technological revolutions involve unpredictable novelties, an understanding of which is inevitably shaped by the familiar. Hence, the car was initially seen as a horseless carriage. By regarding it as a more efficient version of something familiar, people underestimated both the car's ultimate impact on society and the associated hazards. From a present-day perspective the thinking of the time may seem naïve, but we could well be making the same mistake when we speak of 'autonomous vehicles'. We may be guilty of viewing a new technology merely as an enhanced version of something we know, whose true impact we are unable to foresee.

Another misconception prevailed at the time of the car's introduction in the early twentieth century: it was widely assumed that the new vehicles would reduce urban

[33] Verbeek, 2014: 31.

[34] Verbeek, 2014: 71.

[35] Gordon, 2016: 186.

[36] Zuboff, 2019: 12. Zuboff also describes how businesses sometimes deliberately present something radically new as if it were old, in order to encourage use. One example is Google Glass surveillance technology, designed to look like ordinary spectacles (Zuboff, 2019: 156).

pollution.[37] The reason being that the use of horse-drawn transport generated large amounts of animal droppings, which caused unpleasant odours and spread disease. The motor car was consequently seen as a faster and cleaner form of transport. The removal of horse dung from the urban environment did indeed make cities cleaner and more pleasant. However, people failed to realize that the car would cause its own forms of pollution and its own liveability challenges. The history of its introduction thus illustrates that new technologies often have unintended side-effects. The installation of running water and domestic sewerage was originally intended to prevent disease, but also relieved women of one of the most laborious domestic tasks: fetching and disposing of water.[38] An unintended side-effect of electric lighting was a significant fall in deaths in domestic fires, many of which were associated with oil lamps.[39]

Moreover, technological changes can also lead to behavioural changes whose effect is the opposite of what was originally intended. For example, energy-saving light bulbs were developed to reduce energy consumption but in fact increased it because people started using them in places that previously had no lighting, such as gardens.[40] This is what Edward Tenner calls the 'rebound effect' of technology.[41] Another example is the introduction of domestic appliances, which made housework much less physically arduous but also raised expectations – regarding clean clothing, for example – and thus increased workloads in the home.[42]

The unpredictability of system technologies stems in part from the long-term structural changes they bring about, which are impossible to foresee. Railways had a major impact on urban planning, for example, because their arrival meant that people no longer had to live within walking distance of their work. Later the car helped to shape youth culture in the 1960s and enabled new leisure facilities such as drive-in cinemas, drive-through restaurants, motels and roadside diners.[43]

All these uncertainties have implications for AI's integration into society. It is therefore important to recognize that many developments cannot be predicted. We must be very cautious about framing scenarios in definite terms or making linear extrapolations from the past because such approaches are inherently liable to disregard the unexpected effects of new technology.

[37] Bakker & Korsten, 2021.

[38] Gordon, 2016: 123.

[39] Gordon, 2016: 237.

[40] Verbeek, 2014: 22.

[41] Tenner, 1997.

[42] Gordon, 2016: 278.

[43] Gordon, 2016: 166.

4.2.3 Impact on Civic Values

That lesson underpins our decision to consider AI's impact on civic values on the basis of structural overarching tasks. Our rationale is that a system technology's effect in this regard is impossible to predict in definite terms and is often unclear. That is evidenced by the examples presented above: the effect of the car on urban liveability, the effect of electricity on female emancipation and the effect of railways on town planning. The general nature of a system technology makes it impossible to determine what civic values it will affect – in fact, such technologies have the potential to influence them all. In that respect AI is like any other system technology. We can, for example, be confident that it will influence security and health, autonomy and freedom, civil rights and the rule of law, justice and inclusion. However, it is impossible to predict what form that influence will take.

Nevertheless, numerous attempts are being made to identify the values, principles and rights influenced by AI. Such initiatives are an important means of surveying topical issues as a basis for targeted intervention. But if we want to protect our social values in the long term, it is also necessary to look beyond the present and AI's current influence. Our approach, centring on societal integration, is intended to contribute towards the current discourse by focusing attention on the long-term process whereby society and technology influence one another, with potential implications for all civic values.

4.2.4 Regulation and Success Are Not Mutually Incompatible

One final general observation is in order regarding the societal integration of system technologies: there is no inherent tension between civic values and their regulatory protection on the one hand and the economic success of a technology on the other. As we shall see in the next chapter, the frequently cited incompatibility of regulation and success is a myth. The history of technological revolutions shows us that the coexistence of normative parameters and innovation is entirely possible. Of course, regulations can sometimes inhibit technological development by imposing explicit prohibitions, as with the use of nuclear technology for military purposes. Often though, regulation and standardization help make a technology more reliable. That in turn increases public and corporate willingness to utilize and embrace it.

Again, the history of the motor car is instructive here. Over the years a complex system of automotive regulations and standards has been developed, involving mandatory testing and certification, supervisory bodies, safety requirements (seatbelts, airbags, spare tyres and so on), public and private support services, mandatory insurance and, of course, traffic regulations and driving tests. Far from hindering car use, that extensive normative framework has reduced the associated dangers and so promoted confidence. Without roadworthiness tests, seatbelts, insurance, airbags and traffic regulations, car travel would entail far greater risk and probably be less

popular. It should also be noted that the process of automotive regulation and standardization remains ongoing even now.

Much the same is found in the history of the railways. The first trains were dangerous, uncomfortable and dirty. The wooden seats were uncomfortable, the carriages stank of food and tobacco and travellers typically arrived at their destinations covered in soot. In 1879 Robert Louis Stevenson described the train as a 'Noah's ark' on wheels.[44] Gradually, however, the introduction of regulations and standards made rail travel safer and more pleasant. Another interesting analogy is provided by the rise of industrial food production in the nineteenth century. Early manufactured foods were often unsafe and unhealthy. Until the arrival of certification, supervision and legislation, people were exposed to all sorts of dubious practices. Adulteration was common, for example. Chalk and gypsum were added to milk to make it whiter, and it was sometimes diluted with dirty water, leading to the spread of tuberculosis and typhoid.[45]

A similar pattern is likely to emerge where AI is concerned. Although the technology is already entering our lives, the regulations, standards and practices needed to embed it within society largely remain to be developed. We therefore have an unregulated landscape in which individuals and society in general are exposed to a variety of risks. But that does not justify eschewing the technology. Rather, it implies that we need to start work on the long process of enabling the responsible deployment of AI within society.

Key Points – The Embedding of System Technologies

- System technologies are associated with prolonged processes of social and technological co-evolution, often involving profound social change.
- The development of system technologies is often unpredictable, and their effects cannot be fully anticipated.
- It is not possible to distinguish those public values that a system technology will influence from those it will not. The generic nature of such technologies implies that they have the potential to affect all public values.
- There is no inherent tension between regulation and standardization on the one hand, and further development and application of new technologies on the other.

As well as the general patterns characterizing the way that system technologies are embedded within society, we have identified five overarching tasks that form the cornerstones of that process. We now look at these in turn.

[44] Gordon, 2016: 141.

[45] Gordon, 2016: 220.

4.3 Overarching Task 1: Demystification

There are no myths about lawnmowers or toasters. It is clear what their purpose is and how they work, and they leave little to the imagination. With a system technology, however, the situation is different. The generic nature of such technologies makes them somewhat intangible, facilitating the development of myths that bear little relationship to reality. On the one hand, unrealistically high expectations are liable to develop regarding the capabilities of a new system technology, with some people inclined to see it as a panacea for all manner of social ills. On the other we see the rise of exaggerated fears and doomsday scenarios concerning its impact. Myths of the first kind easily lead to disappointment, while the fears encourage aversion. Moreover, both lead to attention being focused on the wrong questions and issues. Properly integrating a system technology into society therefore requires a realistic understanding of what it is capable of and what its effects are. This is what we mean by demystification, a task that asks 'what are we talking about?' (see Fig. 4.3).

Various social actors play a part in this task. Because we are concerned here with public perceptions, the role of the general public is particularly significant. Through their marketing, companies involved in development of a new technology often contribute towards the emergence of unrealistic expectations. Meanwhile, competitors with interests in rival technologies or more traditional industries can play a role in raising fears about a new technology. Civil society organizations can also give credence to myths through their focus on potential risks. Finally, governments often have an interest in the use of new technologies and that can sometimes contribute to overenthusiasm. On the other hand, they can also feed negative perceptions by acting in ways that reinforce certain associations with a new technology.

Fig. 4.3 Overarching task 1: Demystification

4.3.1 Unrealistic Expectations

What patterns of demystification can be discerned in the history of system technologies? Taking optimism first, it is clear that since the Industrial Revolution new technologies have been associated with progress and civilization. Electricity was described as a 'defining element of a great civilization' and inspired many utopian books.[46] Widespread use of electric lighting led to Berlin becoming known as the 'City of Light'. Electricity was linked not only with emancipation (as alluded to above) but also with cleanliness, flexibility and the general improvement of living conditions. This was an example of the wider phenomenon of scientism – the notion that scientific progress leads to social progress – and belief in mankind's ability to manipulate and even perfect society. In 1917, in a manner reminiscent of the expectations surrounding AI, General Electric (GE) advertised its appliances as 'electric servants' that worked 'without complaint'.[47]

Another example of high-flown expectations relevant to the present-day debate regarding digitalization is the belief that new technologies can bring peace. Nineteenth-century engineer Michel Chevalier described the railway as "the most important medium for peace in Europe and human happiness".[48] Similarly, the telegraph was expected to facilitate 'harmony between peoples and nations' and, by uniting humanity, to eliminate barriers of 'prejudice and custom'.[49] In the 1920s Henry Ford, the pioneer of automotive mass production, viewed modern industry in much the same way. He is worth quoting at length:

> Machinery is accomplishing in the world what man has failed to do by preaching, propaganda or the written word. The airplane and radio know no boundary. They pass over the dotted lines on the map without heed or hindrance. They are binding the world together in a way no other systems can. The motion picture with its universal language, the airplane with its speed and the radio with its coming international programme – these will soon bring the whole world to a complete understanding. Thus, we may vision a United States of the World. Ultimately, it will surely come![50]

Utopian visions have always found channels through which to disseminate. One, of course, is science fiction. William Gibson wrote a novel entitled Neuromancer (1984) about an idealized new world he refers to as 'cyberspace' – the first use of that word.[51] Another such channel is high-profile competitions. Historically, innovative entrepreneurs have often competed both to supersede older technologies and with each other. During the rollout of electricity, Thomas Edison and George Westinghouse battled publicly for the ascendency of their respective AC and DC standards. Similarly, the first car manufacturers raced their vehicles against each

[46] Bakker & Korsten, 2021: 16.

[47] Gordon, 2016: 120.

[48] Van der Vleuten et al., 2017: 27.

[49] Gordon, 2016: 178.

[50] Edgerton, 2008: 113–114.

[51] Dommering, 2000: 487.

other. Earlier, the famous steam locomotive Rocket won a series of trials prior to the opening of the Liverpool and Manchester Railway, demonstrating the capabilities of rail transport to the general public.[52]

Because the technologies in question were very new at the time and it was unclear how they could be used, these competitions helped familiarize the public with them. However, they often took place in controlled environments and the accompanying rivalries produced bold statements inflating expectations about what the technologies would be capable of in practice. More recently we have again witnessed public rivalry in the space technology domain, with powerful entrepreneurs like Elon Musk, Jeff Bezos and Richard Branson vying to surpass each other's rocket launches and openly mocking their competitors' technology.[53]

Public exhibitions form another channel that gives rise to utopian expectations. At the 1881 Paris Exposition and the following year's Crystal Palace Exhibition in London, Edison extravagantly demonstrated the potential of electricity to the general public, eliciting enthusiastic newspaper reviews.[54] But such events also became focal points for critics and activists. At the Crystal Palace, for instance, campaigners drew attention to the need for better working conditions and improved safety.[55]

4.3.2 Serious Concerns

The arrival of a new system technology invariably gives rise not only to unrealistic expectations but also to anxieties. One recurring topic of concern is how the technology will affect employment. With a technology that lends itself to widespread application, there is often a fear that it will replace people, thus depriving workers in certain occupations of their livelihood. Related to this is the image of the human-made instrument that rebels against its creator by destroying their income. Although an idea frequently associated with AI, that scenario is far from new. Jonathan Taplin has demonstrated how the internet deprives musicians of income,[56] but in fact the music business has a long history of technological disruption. Describing how the record industry was affecting musicians, a union leader once said that at no other point "in the machine age has the worker created the instrument of his own destruction, but that happens when a musician plays for a recording".[57]

Another prevalent dystopian view is that a new system technology will bring about the loss of a valuable way of life. This argument was used against agricultural mechanization at the end of the nineteenth century, causing anxious farmers in the

[52] Freeman & Louçã, 2001: 203.

[53] Davenport, 2019.

[54] Bakker & Korsten, 2021: 11.

[55] Van der Vleuten et al., 2017: 25.

[56] Taplin, 2017.

[57] Juma, 2016: 213.

US to flock to the Populist Movement.[58] Established industries and their workers are often the sources of distrustful views of new technologies.

One anxiety particularly relevant in the present context concerns the artificial nature of a new technology. This can lead to it being perceived as a sin against nature or the will of God. We see that today with biotechnology, but at one time electric street lighting was portrayed as contrary to the separation of light and darkness in Genesis. Although Berlin enjoyed a positive reputation as the 'City of Light', Jules Verne portrayed the typical German city as 'Stahlstadt' (steel city), symbolizing power and destruction.[59] Even an innovation like margarine had to overcome the criticism that it was an artificial, unnatural form of butter and therefore inherently undesirable.[60]

Fear of a new technology may stem not only from arguments of the kind described above but also from emotional sources such as the power of words. Biotechnology has suffered from the currency of phrases such as 'genetic contamination', 'Frankenfoods' and 'Frankenfish' (farmed salmon).[61] As indicated earlier, there was considerable rivalry between Edison and Westinghouse when electricity was introduced. In that context Edison deliberately sought to make people fearful of his rival's technology. He performed experiments with a dog to demonstrate that Westinghouse's AC standard, unlike his own DC, could be fatal to animals. He also campaigned for AC to be used for the electric chair to associate it firmly with death. In 1889 a magazine created the portmanteau word 'electrocution' by combining 'electro' and 'execution'.[62] Fears can also be fanned more subtly, by rumours. Many technologies have initially been beset by unfounded claims that they were detrimental to human health, contained hazardous or impure ingredients or could even cause sterility.

History thus shows us that the introduction of a new technology is often met with anxiety. Unjustified or exaggerated fear can lead to general aversion, with the result that the benefits of a new technology are never obtained. Juma highlights the simultaneous rise of the mobile phone and GMOs. Whilst the former technology was adopted globally with little resistance, the latter was embraced in the US but rejected in Europe. Perceptions of nuclear technology have been strongly influenced by the disasters at Chernobyl and Fukushima, with implications for subsequent policy in many countries.[63] The point here is not that GMOs or nuclear power should be more widely used, but that the framing of a new technology and public perceptions can play a decisive role in its acceptance.

---

[58] Juma, 2016: 103.

[59] Kaiser & Schot, 2014: 192.

[60] In that case, aversion was reinforced by perception of the new technology as unpatriotic. When coconut oil was first used as an ingredient, margarine consumption was characterized by opponents as supporting farmers in the Philippines and undermining their American counterparts (Kaiser & Schot, 2014: 113).

[61] Kaiser & Schot, 2014: 309.

[62] Kaiser & Schot, 2014: 164–165.

[63] An example is the development of nuclear policy in Italy (Juma, 2016).

The scope for countering anxiety with arguments is limited. Authoritative explanations and technical solutions have often proven insufficient to dispel negative perceptions, especially if they are supported by the – often emotional – power of words and rumours. Once established, public mistrust – a 'social backlash' against a technology – is very difficult to counter. Often, separate issues become associated in the public consciousness and unrelated problems are conflated. A further complication is that the cause of public concern is not necessarily the technology itself but the impression that the authorities are not doing enough to ensure that its use respects the interests and safety of ordinary people.[64]

In the case of a system technology of great potential benefit to society, such as AI, it is advisable to prevent such situations arising. At the same time the scale of a system technology's potential benefits should not be overhyped. Regarding our first overarching task, the government's scope for action is limited. Demystification depends on general public perceptions, which are shaped to a considerable extent by interaction between researchers, the media, schools and private citizens. Nevertheless, the government can exert significant influence in its role as a major user of new technology.

More direct public policy can also have a positive effect. Appropriate tools here include communications by the government, its exemplary use of the technology and support for actors involved in public education such as experts and the media. To facilitate the mechanization of American agriculture, for example, the US government established institutes and groups at universities to promote public awareness of new technologies.[65]

Key Points – Overarching Task 1: Demystification

- The generic nature of system technologies means that they appeal to the imagination. They are associated both with unrealistic expectations of progress and with doomsday scenarios.
- General optimism about technology, public contests and events can inflate expectations regarding a new technology.
- Fears commonly associated with the introduction of system technologies relate to the loss of employment, the loss of a way of life, and the perceived 'unnaturalness' of the technologies.
- Fearful perceptions are shaped not only by arguments but also by emotions, the power of words and framing.
- Both unrealistic expectations and fearful perceptions can lead to an aversion to technology. Realizing the opportunities afforded by a new system technology and ensuring that attention focuses on the right risks during the process of societal embedding therefore depend on demystification.

[64] Gezondheidsraad, 2006: 111.

[65] Juma, 2016: 134.

4.4 Overarching Task 2: Contextualization

Whereas our first overarching task is concerned with perception, the second relates to the use of a system technology. More specifically, to what is required for something developed in the lab to be put to practical use within society. This is a wide-ranging task with multiple dimensions. It is also a complex one. Indeed, its complexity goes a long way to explaining why integrating a system technology into society is such a lengthy process. The fact that something works in the lab does not automatically imply that it will function in practice. Numerous reports have appeared in recent years about algorithms that can apparently diagnose various diseases more accurately than human doctors, reach more reliable verdicts than human judges or produce better translations than human linguists. The fact that such algorithms have yet to replace their human counterparts has much to do with contextualization. Societal integration can be impeded not only by resistance or disillusion supported by myths, but also by problems involving the way the technology works in practice. Central to the overarching task of contextualization is the question 'how will the technology work?' (see Fig. 4.4).

In order to answer this question, we have adopted an ecosystem approach. Contextualization as a task relates to the need for a technology to be embedded in a variety of contexts or ecosystems to function as intended. We distinguish two such ecosystems, the technological and the social.

4.4.1 The Technological Ecosystem: Supporting Technologies

No new system technology – be it the steam engine, electricity, the internal combustion engine or AI – can function independently in a technical sense; it always operates as part of a cluster or block[66] of other technologies. In this context two types of technology are of interest: supporting and emergent.

Supporting technologies are not strictly speaking related to the system technology itself, but nevertheless are essential from the outset if it is to work. The internal combustion engine cannot be used in the automotive industry without steel technology. Furthermore, the success of pioneer car manufacturer Ford owed much to the existence of a large network of dealers and outlets for tyres, batteries and spare parts.[67] Another supporting technology on which the car relied was a suitable road network. In the US the Federal Aid Road Act of 1916 and the Federal Aid Highway Act of 1921 were crucial to creating the car's technical ecosystem.[68]

[66] Alessandro Nuvolari is critical of GPT authors for focusing too narrowly on individual technologies. He argues for thinking in terms of 'development blocks', such as the ICT block formed by semiconductors, computers, software and network equipment (Nuvolari, 2019: 8).

[67] Gordon, 2016: 154.

[68] Bakker & Korsten, 2021: 17.

Fig. 4.4 Overarching task 2: Contextualization

Without such supporting technologies, a system technology can be no more than partially functional at best. Worth remembering in this context is the fact that many people in the early twentieth century doubted that the motor car would actually enter practical use. By comparison, the horse must still have seemed an attractive alternative thanks to its manoeuvrability and its ability to function in an unmodified environment.

The same issue was pertinent to the introduction of the tractor. Its adoption in agriculture was not merely a matter of replacing one instrument with another, it required the creation of a completely new infrastructure of raw materials and suppliers. Moreover, early tractors were less reliable than horses. As a result, it was long assumed that the horse would remain in use alongside the tractor, each for its own purposes. The first tractors in the US were no better than horses, but proved useful on the large expanses of open prairie in the Midwest where there were not enough animals to work the land.[69] It was only with the passage of time that it became clear that they would replace the horse throughout the American agricultural economy.

4.4.2 The Technological Ecosystem: Emergent Technologies

The second cluster within the technological ecosystem of a system technology consists of what we refer to as 'emergent technologies'. Unlike supporting technologies, these develop independently but over time become linked and ultimately coalesce into a cluster. Their existence in the ecosystem allows a system technology to receive a major, unforeseen boost from an external development – as was the case with electricity. Its domestic adoption was initially slow, partly because there were easier ways of lighting homes such as candles and gas lamps. However, the development of domestic appliances like the electric iron, and later electronic devices, made new applications possible and adoption of the technology gathered momentum.

[69] Juma, 2016: 125.

The barcode is another innovation that only came into its own after contextual adaptation. The first barcode scanning systems appeared in the mid-1970s, but it took another 30 years before organizations along the length of the production chain implemented the complementary technological, organizational and process changes needed for their general introduction.[70]

A more recent example can be found in the rise of e-commerce. Expectations of a boom in online retailing had been high ever since the internet first become popular. Amazon was founded in 1994 and in that period was one of the most hyped businesses in the 'dot-com bubble', when markets anticipated a general migration to online shopping. Despite continuing to invest in e-commerce even after the crash of 2000, however, Amazon still failed to turn a profit for some years. It was more than two decades before online shopping really took off. The development required a raft of complementary innovations such as secure and convenient payment systems and improved logistical infrastructures with regional distribution centres. A similar pattern is apparent where transport services like Uber, SnappCar and Greenwheels are concerned. The idea of organizing taxi services and car sharing online has been around for decades, but again it is only in recent years that they have become commonplace. Their success now is closely related to the rise of technologies such as GPS-enabled mobile phones, which allow for local service delivery.

Dependency on a complete technical ecosystem of supporting and emergent technologies means that it typically takes a long time before a new system technology becomes fully functional in practice. Furthermore, the course of that process is inherently unpredictable: the technology itself improves, complementary innovations occur, prices fall[71] and new systems and applications are developed. Even if a system technology initially appears unable to gain traction, the developments necessary for its success may be taking place unseen until suddenly the new technology has a real advantage over established ones and its use acquires momentum.

4.4.3 Enveloping

Where the technological contextualization of AI is concerned, 'enveloping' is an important concept. This refers to the creation of an environment within which a technology can thrive. The concept was popularized in relation to AI by Luciano Floridi, whose work is referred to earlier in this chapter. He is opposed to viewing technology as an instrument, arguing that that implies an old-fashioned model in which the human user exerts influence over a natural environment by means of a technology. While a spear, an axe or a parasol may be regarded as an instrument that impacts an element of the natural environment (a prey animal, a tree or sunlight),

[70] Pethokoukis, 25 November 2019.

[71] Price drops are also very important to the practical functionality of a new technology. Research has shown, for example, that, relative to the early nineteenth century, the price of light has fallen four hundred-fold (Agrawal et al., 2018: 11).

there are many technologies that do not conform to that model. Technologies that act on other technologies, for example, like the hammer when used with a nail – or, indeed, all technologies developed since the Industrial Revolution. A car is not ideally suited for travel in a natural environment but performs very well in one modified by the creation of paved roads. This process of adapting a technology's environment so that it functions better is what we call 'enveloping',[72] its crucial characteristic being that use of the technology is promoted not only by improving the technology itself but also through that adaptation.

In this respect it is pertinent to ask whether the technology is adapting to people or they are adapting to technology? Although the latter idea tends to meet resistance, we have to accept that it is far from uncommon. The average street, for example, is heavily tailored to the motor car, with tarmac, parking spaces, traffic signs and a regulatory system. The people using it, pedestrians, adapt to that by walking on the pavement, using designated crossings and so on. Similar dynamics are likely to become commonplace in the case of AI.

4.4.4 The Social Ecosystem: Macroeconomic Context

Contextualization involves integration not only within the technological ecosystem but also within the social ecosystem. One important element of the latter is the macroeconomic context. A new technology has its own logic, which is not necessarily aligned with existing organizational processes. Moreover, achieving alignment is not a quick and easy process. Organizations have fixed ways of working, making it difficult to try out new approaches.

Those in established industries are often also hampered by 'the curse of knowledge'.[73] Simply purchasing new machines or even setting up new departments – an IT or an AI department, for example – is not sufficient. A modern organization does not have an electrification department; electricity is an established system technology integrated into all its processes. However, that did not happen overnight. Factories had to be reorganized to accommodate power cables, for example.[74] Similarly, the telephone and the typewriter ultimately contributed significantly towards the mechanization and bureaucratization of the office, and thus to the growth of many organizations, but this transformation occurred over an extended period.[75]

Not only is transformation time-consuming, but determining the pathway to be followed is also a capital-intensive process. Consequently, the introduction of a new system technology is often characterized by a 'productivity paradox'. It took years

[72] Floridi, 2014: 144.

[73] Brynjolfsson et al., 2019: 42.

[74] Bakker, 2017.

[75] Freeman & Louçã, 2001: 28.

for electricity to yield a net productivity benefit for the economy, for instance.[76] One explanation for the delay in realizing productivity benefits concerns the energy supply. In Britain, for instance, steam engines were initially used only in the vicinity of coal mines – the source of their fuel.[77] In order to make a system technology productive, therefore, it is important to consider the wider organization of the processes within which the technology must function.

4.4.5 The Social Ecosystem: Behavioural Context

Another important feature of the social ecosystem is the behavioural context into which a new technology must be embedded. In that regard, the behaviour of both consumers and users within the organizations where the technology is to be applied is significant. Internal users often need to be trained to use the applications it facilitates. The more general question of adaptation to the labour market is therefore relevant here as well.[78] Whereas the lab phase requires fundamental knowledge of a technology, the emphasis during the embedding phase shifts to knowing how it should be applied in a variety of domains. During the process of integrating electricity into society, for instance, countless engineers and inventors applied themselves to identifying contexts in which it could be put to effective use.

People who are going to utilize a new technology must gain confidence in it and some understanding of how it works, and must perceive its use as desirable. That in turn depends on the presence of positive stimuli and the absence of deterrents to its integration. People will not embrace a new technology if they fear it will make them redundant or undermine their earnings. Artists working in recording studios prior to the development of a new income model based on streaming services form a good example of this. Likewise, professionals such as doctors, judges and accountants will be reluctant to accept or fully utilize a new technology if it is not – or not yet – capable of satisfying the standards of their profession.[79]

New technology often requires behavioural change from consumers as well. Consider again the example of music recordings. Before they became possible, people could listen only to live music and only at scheduled performance times. The wireless and gramophone enabled entirely new ways of listening to music at home, but consumers still had to accustom themselves to those new opportunities.

Like demystification, contextualization is a broad task over which governments have relatively little control. To a large extent, the contextualization of a new technology occurs in the many thousands of occupational settings where people make use of it and learn when and how it is effective. That is an iterative process.

[76] Agrawal et al., 2019.

[77] Bakker & Korsten, 2021: 6–7.

[78] See the WRR report Better Work regarding the technologization of work (WRR, 2020).

[79] Van Ettekoven & Prins, 2018.

Governments can nevertheless facilitate and guide the broad task of contextualization in various ways.

For example, governments can invest in supporting and emergent technologies. The US government aided the contextualization of the car by building highways. Another option is to participate in the process of contextual experimentation. As a new technology user, the government plays a role in the creation of a market. It can also define standards and set an example for the private sector. Public-sector procurement policies are influential too, due to the government's sizeable purchasing power.

Key Points – Overarching Task 2: Contextualization

- Contextualization is necessary for a new technology to function in practice.
  - That implies understanding and approaching the technology within its wider social and technical ecosystems.
- The technical ecosystem consists partly of supporting technologies that enable a system technology to work.
- It also includes emergent technologies: completely separate technologies that develop independently but can add surprisingly strong impetus to a technology.
- An important process in the contextualization of system technologies is 'enveloping': adaptation of the environment to a technology.
- A technology's social ecosystem consists firstly of the macroeconomic context and is characterized by complex productivity and work process organization issues.
- The second element of the social ecosystem is the behavioural context, which is characterized by the stimuli, practices, standards and convictions of people involved with the technology.

4.5 Overarching Task 3: Engagement

As we have seen, our first overarching task is concerned with image and the second with usage. The third, engagement, relates to the social environment. It focuses on the people affected by the system technology and the actors who therefore are or need to be involved with it (see Fig. 4.5). They include technical experts, ordinary citizens and civil society organizations.

Fig. 4.5 Overarching task 3: Engagement

4.5.1 Values, Interests and Ideals

As previously stated, the five overarching tasks are closely related. We have already considered the human environment in terms of the social ecosystem's role in contextualization, centring on the question of how we could make the technology work. In the context of the engagement task, by contrast, our focus is on people's involvement in the design and use of the technology – and it is important that they are involved, so that their values, interests and ideals contribute towards its integration into society. People's interests can of course play a role in building a technology's functionality as well, but the principle underpinning the engagement task is that involvement by various groups in the process of societal integration is intrinsically important to its long-term success. Effectively, engagement is about humanizing or democratizing the technology.

Engagement has proven particularly important in the phase where a technology transitions beyond the lab, because at that point the requirements society will demand of have yet to crystallize fully. The engagement of civil society is also vitally important because every technology is associated with power structures. The first users of a new technology are typically powerful actors such as large corporations and governments. Consequently, it is initially likely to reinforce existing power structures. Engagement is required to ensure that other social actors also have a voice in the way it is used.

4.5.2 A Spectrum of Engagement

Engagement can take various forms. At the one end of the spectrum are people strongly opposed to the technology who want to see it banned. In extreme cases, their resistance can turn violent. At the other end of the spectrum is supportive input, with actors offering their expertise and voicing their own values and wishes so as to influence how the technology is used. That can even lead to stakeholders themselves developing alternative uses.

Stakeholders can also engage indirectly by calling on governments to regulate the technology (regulation is our fourth overarching task; see 3.6.). In this respect it is important that engaged social actors mobilize themselves to exert the necessary influence, and that they do so from an early stage – the reason being that uncertainty regarding the direction a developing technology will take can make it difficult for government to know how it should be regulated. In that situation, civil society actors can assist politicians and governments by playing vital signalling and deliberating roles. Which brings us to the core question in this overarching task: 'Who should be involved?'

4.5.3 Winners and Losers

Individual citizens and interest groups engage with the process of embedding a system technology within society for various reasons. Often, these reflect whether the person or group in question stands to gain or lose from the technology. Although a new technology may be beneficial to society, that benefit is liable to be distributed unevenly, creating both winners and losers. When Schumpeter referred to creative destruction, he recognized the misery new technology could cause and visualized large elements of society being crushed under 'the wheels of innovation'.[80] As well as threatening jobs, the process of innovation and experimentation often involves accidents and even reckless and dangerous behaviour. We have already mentioned the malpractices associated with the early mass production of milk and the introduction of margarine. Manufacturers often used colourant and preservative chemicals that were harmful to public health.[81] Many new technologies have also had a negative impact on particular groups in society, such as consumers or workers. That has tended to happen where vulnerable or dependent groups have been disadvantaged by more powerful early adopters of the technology and first movers exploiting their expertise and position. In such cases, new system technologies initially amplify existing power imbalances.

The introduction of the steam engine induced fear that the working classes would be marginalized. When railway travel became popular, wealthy people were concerned about close contact with the poor, leading to a system of multiple travel classes.[82] Electric street lighting was perceived as increasing the government's power over its citizens. Class differences created issues in relation to the motor car as well: cars were seen as the preserve of a wealthy elite, who were gradually driving other members of society off the roads.[83]

[80] Schubert, 2013.

[81] Juma, 2016: 97.

[82] Van der Vleuten et al., 2017: 47.

[83] Bakker & Korsten, 2021: 30.

Such issues have repeatedly prompted those affected to engage with new technologies. One form that engagement has often taken historically is protest, in extreme cases descending into violence. In the 1810s a movement of English factory workers known as the Luddites rebelled against the mechanization of labour, destroying the machines their employers had been installing. During the 1842 Plug Riots, half a million workers went on strike and disabled steam engines. Workers resorted to such tactics because the British government of the day did very little to protect them.[84] The word 'Luddite' has since come to mean anyone who makes futile attempts to resist technological progress. However, that definition rests on a simplistic view of history. By rioting, the Luddites succeeded in slowing the process of mechanization in the textile industry and building solidarity amongst its labour force, thus laying the foundations of the trade-union movement.[85] They were not rejecting the new technology per se but standing up for workers' rights.

The introduction of the motor car was also accompanied by protests from disadvantaged groups. Some of these were sparked by the hazards made clear by the first fatal accidents. The main focus of dissent, however, was the 'battle for the street', as the car gradually pushed market traders, horse riders and pedestrians off the roadway. Horses were perceived by motorists as causing congestion, while their riders complained about the space devoted to car parking. During the 1930s, the car lobby succeeded in persuading the public that the roads were meant primarily for motor vehicles. That perception was encouraged by education, with children taught to look out for cars when crossing the road. Regulations were introduced not only for motorists, but also for cyclists and pedestrians. Crossing intersections diagonally was made an offence, for example. Campaigners called for fast roads exclusively for motorists, eventually leading to the creation of motorways. In short, the rise of the car brought with it disputes over who was and was not legitimately entitled to use the road, and under what conditions.[86]

More recently, the introduction of nuclear power attracted protest. Posters, newspaper adverts, stickers and demonstrations such as 'die-ins' and human-chain protests were used to oppose the technology. In some cases, protestors also sabotaged equipment.[87] Such actions ultimately helped initiate a general public and political debate regarding nuclear power.

4.5.4 Demand for Regulation

As the example of the car demonstrates, engagement by civil society sometimes takes the form of campaigns calling for regulation or government policy. In the mid-nineteenth century, for instance, the Chartist movement in the UK secured

[84] Bakker & Korsten, 2021: 9.

[85] Juma, 2016: 26–27.

[86] Van der Vleuten et al., 2017: 84–86.

[87] Van der Vleuten et al., 2017: 135.

legislation to limit the maximum number of working hours for young people and women.[88] Later that century, women's organizations in the US pressed for better working conditions. The Woman's Christian Temperance Union (WCTU) campaigned not only against alcohol but also against the widespread use of many new medications. Its activities contributed towards the introduction of legislation requiring the listing of ingredients on product labels and restricting the distribution of medications.[89] In 1970 an activist engineer in the Netherlands invented the 'speed hump' to improve road-traffic safety. A few years later the Dutch government approved the concept of the 'woonerf', a residential zone where pedestrians have priority over cars.[90]

Civil society actors have also been able to influence the use of new technologies directly, rather than by pressing politicians to act. One way they have done that is by using a technology as they see fit. In the US, for example, co-called 'Bellamy clubs' were formed to employ technologies for utopian social purposes. Unions, feminists, doctors and food specialists have pressed for modern domestic technologies and appliances to be made more healthy, safe and pleasant. User communities have even designed housing blocks with shared spaces for cooking and childcare to promote community spirit and equality. When the telephone was introduced, women and migrants started using it in ways the phone companies had not intended, ultimately leading to the modification of services.[91]

The White Label League succeeded in persuading clothing producers to attach a white label to garments made in factories where the working conditions had been approved by the organization.[92] In the field of digitalization, the Claudette project is a good example of civil society influencing a technology's use: it seeks to reinforce the position of consumers by automatically scanning countless online platforms to check the legality of their terms and conditions and help buyers understand them.[93]

4.5.5 Defending Public Interests

Citizens affected by new technologies engaged in many different ways: by experiencing their effects, guiding their use and making their own views known. Certain social actors play particularly significant roles. Considerable influence is exercised by the media, whose involvement we have already considered as it relates to demystification. In that context, its role is to inform the public; where engagement is concerned, by contrast, it is to air issues relating to public interests.

[88] Freeman & Louçã, 2001: 172–173.
[89] Gordon, 2016: 221–224.
[90] Van der Vleuten et al., 2017: 153.
[91] Van der Vleuten et al., 2017: 44–46.
[92] Van der Vleuten et al., 2017: 50–51.
[93] Leeuw, 2020: 132–133.

One historical example of this kind of mobilization relates to the introduction of urban electricity cabling. In October 1889 Western Union employee John E. H. Feeks was electrocuted in a gruesome accident on a cable installation project in New York City. His body was left hanging, smoking and sparking, for 45 minutes before it could be brought down. The incident caused a widespread backlash, with newspapers reporting acts of sabotage all over the city. The prevailing view was clearly that the power companies were putting profits ahead of public safety. The New York Times argued that the people should no longer have to tolerate the activities of selfish entrepreneurs and ignorant, corrupt officials. The commotion led to a major inquiry into the power of dominant companies and even to new governance models, in which municipal authorities were given more control and greater emphasis was placed on public participation.[94]

In the same city but a very different field, the rise of the refrigeration industry provides another example. Refrigeration technology enabled goods to be stored in artificially cooled warehouses, removing the need for natural ice. Some people, however, grew suspicious of the power of the 'ice trust'. Encouraged by the newspapers, a public outcry ensued, leading to regulations requiring products to be labelled with their refrigeration date.[95] Another problem was that the doors of early coolers and freezers were difficult to open, with the result that playing children could become trapped in them and suffocate. Media outrage led to the introduction of safer door designs.[96]

Scientists and other experts form another important group within an engaged civil society. They can exercise influence by, for example, publishing books and articles that raise public awareness and draw attention to problems and malpractices. In 1962, for example, biologist Rachel Carson famously published Silent Spring, a book that did much to launch the ecology movement. Her analysis exposed the downside of industrial manufacturing and agriculture, thus mobilizing opposition to big business.

Similarly, the work of critics such as Guy Debord, Constant Nieuwenhuys, Jane Jacobs and Lewis Mumford created awareness of malpractices in the automotive industry.[97] Artists and fiction writers can also contribute towards public engagement. The Bellamy clubs mentioned earlier were inspired by Edward Bellamy's book Utopia: Looking Backward. Famous authors such as H. G. Wells and Mark Twain also wrote about the influence of technologies such as electricity.[98] In 1906 Upton Sinclair published The Jungle, a novel about the dreadful conditions in Chicago's meat industry. The book led to an immediate decline in meat consumption, and public disquiet resulted in the formation of a system of inspectors.[99]

[94] Juma, 2016: 165–166, 172.

[95] Juma, 2016: 185.

[96] Juma, 2016: 186.

[97] Van der Vleuten et al., 2017: 120–121.

[98] Freeman & Louçã, 2001: 232.

[99] Gordon, 2016: 82.

Earlier, during the Industrial Revolution, a medical commission reported that the people of Manchester were being made unwell by the city's smoky air.[100] It was a long time, however, before anything was done about the situation. As mentioned in Sect. 4.1, social actors' degree of organization – and hence their influence –has grown over time. Professional groups, associations and commissions made up of academics and other experts have started to play an increasingly influential role in the societal integration of new technologies. The academic press is an important medium for the exercise of such influence, along with appeals and conferences. In 1955 philosopher Bertrand Russell and physicist Albert Einstein published a manifesto calling for the academic world to contribute towards the peaceful resolution of international conflicts. These contributions were followed by a series of expert gatherings known as the 'Pugwash Conferences on Science and World Affairs'.[101]

Following the development of genetic cloning technology in 1973, 2 years later the Asilomar Conference on Recombinant DNA agreed a voluntary moratorium on genetic modification to allow the medical authorities to develop safety guidelines. This laid the foundations for an evidence-based system of risk analysis.[102] Another example of experts influencing technological development is the work on climate change done by the IPCC, whose members are all leading academics. Scientists, other experts, writers and artists, as well as private citizens and interest groups, may campaign against the use of new technologies, then, but for the most part they contribute towards bringing about more responsible application of those technologies, thus actually encouraging their use.

One final observation is that, with their professional emphasis on open publications and knowledge, academics and researchers can stand up against governments and businesses in situations where the latter have an interest in maintaining secrecy. The Human Genome Project was an international collaborative initiative to make the genome publicly available. At the same time, however, a company called Celera was working to sequence it privately for commercial exploitation. That brought it into conflict with the academic world. The scientific and business communities also clashed over the question of whether genetic sequences were patentable.[103]

In the field of cryptography, scientists also find themselves at odds with the policies of secrecy pursued by governments and private corporations. In the 1990s, legislation banning the export of sensitive knowledge made it difficult for US academics to know what they were and were not allowed to teach their foreign students. In defiance of US government pressure to keep encryption software secret, programmer Philip Zimmerman open-sourced his code, leading to his prosecution.[104]

The open-source movement is an important group of civil society experts in the field of digital technology. Numerous court cases attest to the tensions that surround

[100] Bakker & Korsten, 2021: 9.

[101] Van der Vleuten et al., 2017: 102.

[102] Juma, 2016: 236–237.

[103] Huys et al., 2011: 1104–1107.

[104] Leung, 2019: 202.

publication. In the Netherlands, for example, a case was brought against Radboud University's Bart Jacobs after he discovered a security flaw in the Mifare Classic chip, used on Dutch public transport smartcards, Transport for London's Oyster cards and elsewhere. The court refused to grant an injunction preventing publication of the details, however.[105] Of significance in the context of our analysis is the judge's observation that "in a democratic society, important interests are associated with the ability to publish the results of scientific research and to inform the public about a product's shortcomings, so that steps can be taken to mitigate the risks."[106] The publication of a paper explaining how a dangerous variant of smallpox had been developed was the focus of similar tensions.[107] Experts employed by commercial organizations play a role not only with regard the issue of publication, but also in relation to other ethical issues within businesses. After the Second World War, for example, the members of a German engineers' association took an oath not to work for companies that infringed human rights.[108]

Key Points – Overarching Task 3: Engagement
- The engagement of civil society is important for drawing attention to relevant values and interests affected by using a new technology.
- Civil society plays an important role through a wide range of engagement forms, from resistance and protest to campaigning, and driving change in the design and use of a technology.
- The media and journalists are important for highlighting malpractices and mobilizing public opinion.
- Scientists and other experts can, for example, develop standards and principles of good practice, promote a culture of openness regarding a technology, and utilize a new technology in accordance with public values.

4.6 Overarching Task 4: Regulation

Our fourth overarching task is pan-societal: the regulation of new technology. In this context we define regulation broadly as including not only legislation and government policies but also professional norms and technical standards. Central to this task is the question, "what parameters are required?' (see Fig. 4.6). Although national and international government bodies play a defining role here, other players are also influential.

[105] Arnhem High Court, 18 June 2008.

[106] Judge of Arnhem High Court, 18 July 2008.

[107] Leung, 2019: 150–154.

[108] Van der Vleuten et al., 2017: 127.

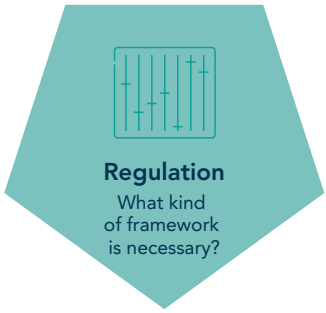


Fig. 4.6 Overarching task 4: Regulation

4.6.1 The Collingridge Dilemma

Defining rules for something as extensive, complex and versatile as a system technology brings numerous challenges, problems and dilemmas. One of the best known is the so-called 'Collingridge dilemma'. On the one hand a new technology is difficult to regulate in the early phase because much remains unclear regarding its workings and effect. Moreover, the need for regulation is initially less apparent. Later, once the technology's effects on society are more conspicuous, it becomes clear what regulation is needed and why. By then, however, many of the decisions taken earlier are difficult to reverse. A further complication is that power structures develop around a technology, and these cannot be modified easily or quickly. Primarily, therefore, we first encounter an information and knowledge problem and then later a power problem.

The Collingridge dilemma is exemplified by the architecture of the internet, which was developed in a spirit of openness and market freedom. Today, however, it is clear that many safety and security issues were not adequately addressed by the original design, meaning that we are now vulnerable to digital disruption, for example.[109] Rectification of the design flaws at this stage, however, would require large sections of the internet to be completely restructured – a huge, if not impossible, task.

4.6.2 Concentration of Power

Once a new technology has been widely adopted – that is, integrated or embedded in society – it is difficult to make major changes. However, the need for such changes only increases over time. As indicated in connection with the previous overarching task, the first signs that change is needed are acute issues, often highlighted by civil society. They typically involve accidents, abuses, opportunistic use and dangerous practices. By gradual degrees, it becomes clear that more structural regulation is

[109] WRR, 2019.

required in order to manage the technology and its impact on society. Central to the regulation process is an expansion of the field of focus from acute issues only to more structural problems.

One structural issue that arises repeatedly in the history of system technologies is concentration of power. The dynamism and innovation associated with new system technologies tend to result in monopolistic or oligopolistic power being heavily concentrated in the hands of certain actors. As well as causing economic problems, such concentration results in the powerful actors gaining disproportionate influence over society, threatening civic values.[110] At first, the companies in whose hands power is concentrated are typically seen as wonderful innovators and social benefactors. Over time, though, a more negative view of them develops as their power is increased by the spread of the technology.

In the US, railway pioneers such as Andrew Carnegie and Jay Gould built huge business empires. However, the negative view of their power and influence that ultimately prevailed is clear from the nickname they acquired: 'robber barons'. The introduction of electricity was also accompanied by an immense concentration of power. In 1894 the Edison Company merged with Thomson-Houston to form the giant GE, which together with Westinghouse dominated the US market. In Europe, Siemens was formed in Berlin and Ganz in Budapest – two of the first true multinationals. Immediately before the First World War, GE and Westinghouse in the US and Siemens and AEG in Europe were the world's biggest companies. At that time there was considerable fear of this 'global cartel'. Indeed, AEG general manager Emil Rathenau did actually reach an agreement with GE in 1903 about dividing up global markets.[111]

The oil industry's boom period occurred at around the same time, leading to creation of John D. Rockefeller's giant Standard Oil corporation. A little later the rise of the internal combustion engine was associated with an enormous concentration of automotive industry power in Detroit, Michigan, the Silicon Valley of its day. The city was home to the industry's 'Big Three': General Motors, Ford and Chrysler. In the 1920s the US and Canada were responsible for nearly 90 per cent of global production of trucks, cars and tractors.[112] Detroit continued to dominate the industry for many years, both within America and beyond. The saying "what's good for General Motors is good for America, and vice versa", attributed to Charles Erwin Wilson, reflects the influence the company had over the nation. In the mid-twentieth century AT&T dominated the telecommunications world, and its research arm Bell Labs was a global driver of innovation. In the computer industry, IBM came to enjoy similar power. The classic film 2001: A Space Odyssey depicted the dangerous side of the company's influence. The film's malicious computer intelligence is called HAL, a name created by taking the three letters that come before I, B and M in the alphabet.

[110] Prüfer & Schottmüller, 2017.

[111] Freeman & Louçã, 2001: 244.

[112] Freeman & Louçã, 2001: 260.

A historical pattern can be discerned where, as the concentration of power has become a greater issue for society, powerful companies campaign for the resolution of market problems to be left to the market or self-regulatory systems. Often, they do so in an effort to avoid the imposition of external controls. 'Robber barons' like J. D. Rockefeller and J. P. Morgan portrayed their power as the result of entrepreneurial genius and a necessary by-product of technical progress.[113] Shoshana Zuboff explains how they cited the 'laws' of economics and evolution in their defence. Legislation was unnecessary, they argued, because they were subject to regulation by the laws of evolution, capital and supply and demand.[114] Many employers also maintained that workplace safety was the responsibility of the workers themselves.[115] Similarly, it was suggested that the safety of a car was the user's responsibility, not the manufacturer's.

4.6.3 New Legislation and Regulations

As the need to regulate a new technology becomes clearer, we need to ask whether existing legislation provides an adequate mechanism for its control or are specific new laws required to address the novel circumstances associated with it. When bespoke have been considered necessary in the past, it has often proved possible to legislate or regulate successfully even at the international level to mitigate the adverse effects and applications of a new technology. The 1925 Geneva Protocol is a good example. Following the widespread use of poison gas in the First World War, this treaty measure agreed a ban of the use of both chemical and biological weapons.[116] Another case is the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer, which successfully combined restrictions on the use of certain chemicals with stimuli to use technological alternatives.[117] Also instructive in this context are the arrangements made nearly 15 years ago by various countries within the Council of Europe to tackle the online sexual exploitation of children.[118] The international dimension of societal integration is addressed explicitly by our fifth overarching task (see 3.6 below). As far as regulation is concerned, it is important to note that a technology can be controlled successfully, particularly with a view to mitigating the associated hazards, by means of legislation at both the national and international levels.

---

[113] Taplin, 2017: 8–9.

[114] Zuboff, 2019: 106–107.

[115] A serious fire at a textile factory in New York in 1911 led to outrage and ultimately fire safety mandates. Gordon, 2016: 271–272.

[116] Floridi, 2014: 203.

[117] Juma, 2016: 302.

[118] By means of the Council of Europe's Convention on Cybercrime (ratified by the Netherlands in 2006), the Council of Europe's Treaty of Lanzarote (ratified by the Netherlands in 2010) and EU Directive 2011/93/EU.

As the Collingridge dilemma reveals, especially early in a new technology's trajectory it can be difficult to know what types of regulation are required. The reason being that some regulations can undermine the advantages of a new technology. One example is provided by the so-called Red Flag Acts passed in the UK in the second half of the nineteenth century. With the aim of promoting road safety, these laws required that a mechanical vehicle must be preceded on the public highway by a person walking with a red flag.[119] Their effect was to seriously limit the maximum speed of the new transport mode and thus diminish its value.

4.6.4 Diverse and Flexible Instruments

One important lesson we can draw from the history of the regulation of system technologies is that there are no silver bullets: no single measure is able to ensure that a new technology is embedded in society in a totally responsible way. As we saw with the introduction of the motor car, regulation involves many years of constantly responding to new issues and hazards. In the Netherlands, for instance, the first urban speed limit was imposed in 1957. It was not until 1974 that motorway speed limits followed, though, in response to the new dangers associated with traffic growth. Seatbelts were made compulsory for drivers and front-seat passengers in 1975, but not for other passengers until 1992. Only in 1982 were rules introduced requiring all vehicles to undergo regular roadworthiness tests. Even now, the process regulating the societal embedding of the car continues. Regulation is a learning process that takes an increasingly substantive form with the passage of time.

History also teaches us that the extent of government intervention follows a pattern as well. Initially, it is considered prudent to use the most flexible instruments available. Then, as more knowledge and experience are acquired, there is a gradual transition towards more mandatory instruments.

Various flexible instruments are possible. First, there is analogous legislation. When a new technology emerges, such as biotechnology or nanotechnology, regulators look for analogies in other fields. In the case of nanotechnology, for example, that was the chemicals industry.[120] Other flexible instruments that are used include experimental legislation, 'soft law' and 'regulatory sandboxes' in which new business models can be tested.

The information problem with a new technology can also be addressed through public-private co-operation. This collaborative model is increasingly common around the world, as we shall see in the context of the next overarching task. It is most common in highly technical fields, where private sector expertise is very important. The International Organization for Standardization (ISO) is a good

[119] Juma, 2016: 295.

[120] Lee & Vaughan, 2010: 193–218.

example.[121] In the regulation of biotechnology too, various softer governance instruments are used, with researchers, governments and companies collectively working out the best way to manage a new technology.[122]

4.6.5 Oversight

In addition to legislation and standardization, regulation requires oversight and enforcement. Again, a dynamic, learning approach is required, especially in the early part of a new technology's trajectory. One particular issue arises out of the generic nature of system technologies, which means that they can be used in a wide variety of contexts, each with its own rules, values, principles and history. As a result, it is difficult to ensure that legislative arrangements and oversight bodies specifically address all possible applications.

Let us return to the example of electricity. Some of the associated questions are universal, such as the type of voltage and the cabling. But in reality electricity features in people's lives of citizens through all manner of specific applications, from factories and street lighting to toothbrushes, escalators and computers. The vast majority of rules governing electricity therefore relate to those particular applications. Furthermore, generic technologies are often dual-use technologies; that is, they have both military and civil potential.[123] This is a complicating factor because the two types of use require very different rules and enforcement mechanisms. Domain-specific knowledge is therefore required for the application-level regulation of system technologies.

The institutions and bodies responsible for enforcement of the applicable rules must also be involved with the societal integration of a system technology. The regulatory influence of the judicial system should not be underestimated either, particularly when a new technology's impact on society has yet to become clear. This is illustrated by a 1995 US court ruling on cryptography, in which the judge decided that a ban on the distribution of encryption software would infringe the constitutional right to free speech ¬– a central tenet of US democracy.[124]

Parliament also plays a material role in shaping rules and regulations. In fulfilment of their oversight function, MPs can draw attention to malpractices and issues. The legislature may also politicize technology, as again illustrated by the history of encryption. Although the US government and executive agencies such as the NSA and FBI wanted to restrict the distribution of encryption software as far as possible,

[121] Leung, 2019: 17.

[122] Leung, 2019: 227.

[123] It is estimated, for example, that 95 per cent of all space technology is dual use (Leung, 2019: 66).

[124] Schulz & Van Hoboken, 2016.

Congress repeatedly stood up for the rights of citizens versus the state.[125] In order for the judiciary and parliament to perform their supervisory functions, it is important that they possess the means and the knowledge needed to monitor the use of new technologies effectively. In the US, for example, the Office of Technology Assessment played a key role in assisting Congress between 1972 and 1995.

4.6.6 A Growing Role for Government

The foregoing illustrates that the role of government, and thus of legislation, democratic control and oversight, increases as a system technology becomes embedded in society, not least because its effects become clearer as that process proceeds. The more embedded a technology is, moreover, the harder it is for society to do without it. As a result, it (or aspects of it) are increasingly regarded as public property, sometimes even as a utility. Technologies viewed in that way include public transport, the electricity grid, the road network and broadband cable infratructure.[126] The power problem described earlier is also significant in relation to the government gradually acquiring a greater role than it had at the outset, when it was primarily private companies shaping the technology.

In that context, there is a history of governments using a variety of means to tackle the power of dominant system technology players, who we can regard as the predecessors of today's big-tech companies. The power of the 'robber barons', for instance, was challenged during the so-called 'Progressive Era'. The Sherman Act of 1890, originally passed to address the power of the big US 'trusts' (cartels)[127] was later utilized by President Theodore Roosevelt to break up Rockefeller's Standard Oil and Morgan's Northern Securities.[128]

As well as addressing concentrations of power, a government can protect public interests by obliging businesses to comply with certain conditions. The US government established the Rural Electrification Administration to force electricity companies to make their services available in rural areas where they had little commercial incentive to operate.[129] When AT&T had a monopoly of the US telecommunications market, it was required to adhere to strict requirements such as relinquishing patents.[130]

Finally, we should point out that significant international differences exist in terms of the traditional role of government and the way intervention is viewed.

---

[125] Ibid.

[126] In the first half of the twentieth century, the railways in many countries, including Canada, Germany, France, the Netherlands, Sweden, Spain and the UK were nationalized, for example (Van der Vleuten et al., 2017: 74).

[127] Freeman & Louçã, 2001: 342.

[128] Taplin, 2017: 115.

[129] Gordon, 2016: 315.

[130] Taplin, 2017: 259.

Contrasting with the situation in the US, in Europe there has been considerable public involvement from the outset in many of the new technologies considered in this report.[131] It is certainly the case that whenever a system technology is embedded in society, public interest in it increases over time and that in turn strengthens the rationale for the government to play a regulatory role.

Key Points – Overarching Task 4: Regulation

- Although regulating a technology is easiest early on, at that stage there is often uncertainty about what is required and little sense of urgency.
  - By the time a sense of urgency develops, it tends to be harder to introduce regulations or change established practices.
- With a new system technology, the initial approach is usually to rely on self-regulation. However, the concentration of power in the hands of a few companies and the rise of malpractice gradually make legislation necessary.
- Where legislation is concerned, there are no silver bullets. The control of a new technology therefore requires a wide range of instruments. Both flexible instruments such as experimental legislation and soft law, and public--private cooperation on standards are useful ways of acquiring knowledge and expertise and dealing with uncertainties.
- The generic nature of system technologies and the associated diversity of their applications necessitates a primarily contextual approach to oversight and enforcement.
- The role and influence of the government in the embedding of a technology differs from country to country, but the need for intervention increases over time, as the technology acquires a more prominent position in society and the public becomes more dependent on it.

4.7 Overarching Task 5: Positioning

The final overarching task we have identified is positioning, which involves embedding AI at the international level – although each of the other four tasks also has an international dimension. Regulation, for example, is not an exclusively national matter, but also involves supranational organizations. To some degree, the engagement of actors such as scientists and activists is often an international process as well. Nevertheless, international positioning is a distinct task for two reasons. First, because it involves different players than those encountered at the national level. Second, because certain issues are specific to the international stage, such as the competitiveness and security of nations. The question at the heart of the positioning task, therefore, is 'How do we compare with other countries?" (see Fig. 4.7).

[131] Bakker & Korsten, 2021.

Fig. 4.7 Overarching task 5: Positioning

4.7.1 Economic Competitiveness

In the international context, one of the characteristic features of system technologies is the tendency for a race to develop between nations. The belief prevails that countries that lead the way in the development and application of the technology will gain various advantages over others. Any country that believes itself in danger of being left behind will therefore strive to improve its position in the race.

The resulting emphasis on international competition can complicate the process of dealing with normative issues associated with the technology. During the Industrial Revolution, for example, the nations of mainland Europe were envious of the economic and technological development they could see occurring in the UK. British steam engines made a profound impression. Britain was dubbed 'the Realm of Vulcan', after the Roman god of fire, while the country's railways, chimneys and factories were likened to the architecture of the Roman Empire. The model was impressive and simultaneously repellent. The British were perceived as materialistic and greedy, contributing to a mood of Anglophobia elsewhere.[132] Germany and the US were viewed with similar ambivalence in connection with later technologies. Hence, a sense developed that technological leadership was acquired at the cost of various fundamental values.

History shows that the successful development and application of a new system technology does indeed contribute to a nation's competitiveness, since the generic nature of the technology means that it facilitates generalized economic and social progress. National strategies of public investment in infrastructure and education can make a useful contribution in that regard. In the late nineteenth century, for instance, Germany's rapid economic development owed much to the country's coordinated approach to the integration of science and industry. Public investment in new technologies also helped East Asian countries such as Japan, South Korea, Taiwan and China to become powerful modern economies in the twentieth century.[133]

[132] Bakker & Korsten, 2021: 9.

[133] Johnson, 1982; Wade, 2018; Amsde, 1989; Zhang, 2012.

4.7.2 Military Relations

The international competitive advantage conferred by system technologies is not only economic. Leadership in a major new technology can also strengthen a nation's military position on the international stage. Railways facilitated the Prussian victory over France in 1871, for example.[134] They also played an important role in European countries' colonization activities around the world.[135]

During the twentieth century, investment in the development and application of new technologies continued to have a major bearing on conflicts. During the Second World War, the British and American scientific communities, including code-breaker Alan Turing and the ballistic scientists whose work laid the basis for the development of computers, were in direct competition with German science, including rocket pioneer Werner von Braun. When the Soviet Union launched Sputnik 1 in 1957, there was widespread fear in the West that the US might lose the Cold War as a result of being left behind in the space race. A year later the Defense Reorganization Act was passed, leading to the creation of ARPA. Later renamed DARPA, the newly formed military research body went on to drive the development of many new technologies, amongst them GPS and the internet. Meanwhile, the National Aeronautics and Space Act formed NASA in 1958. The new agency's staff included Von Braun, who had been brought to the US after the war as part of the Operation Paperclip mission to kick-start the development of American space technology.[136] Subsequently, in the 1960s, the Kennedy administration made the creation of a global US satellite system a national priority. President Eisenhower also sought to ensure the technological leadership of US companies against the backdrop of the geopolitical rivalries of the Cold War.[137]

4.7.3 Attempts at Nationalization

National strategies regarding system technologies do therefore contribute to the competitiveness and geopolitical strength of the countries in question. However, viewing system technology development and deployment as a global race has limits in terms of its validity. There exists no historical basis for believing that one country can win such a race by monopolizing a technology and thus securing a permanent advantage over others. System technology development has generally been an international process, to which multiple countries have contributed.

Early contributors to the internal combustion engine, for example, included the Swiss François Isaac de Rivaz, Belgian Jean Joseph Etienne Lenoir, Germans

[134] Bousquet, 2009.

[135] Diogo & Van Laak, 2016.

[136] Weinberger, 2019.

[137] Leung, 2019: 79–83.

Nikolaus Otto, Karl Benz and Rudolf Diesel and Americans George Brayton and George B. Selden. The development of electricity was an international effort as well.[138] Although the steam engine was developed largely in Britain, that nation obtained no consequent lasting advantage. The US may have entered the 'race' later, but the engine developed by American engineer George Henry Corliss ultimately proved superior and eventually conquered the British market as well.[139]

Moreover, system technologies tend to be characterized by an international approach that owes much to the involvement of the scientific community. Scientists generally attach great importance to knowledge being freely accessible and contribute enthusiastically to international conferences and journals. Efforts to 'nationalize' system technologies consequently tend to be driven by governments rather than academics.

When electricity was introduced, for example, the British responded to the rise of the US and Germany by enlisting the help of Italian engineer Guglielmo Marconi to create wireless telegraph networks in an effort to dominate international communications. Their hope was that an 'imperial chain' would confer an unassailable advantage. Later, America sought to establish a rival network and the US navy blocked the sale of GE's sophisticated technology. The Radio Corporation of America (RCA) was founded with the aim of securing global wireless hegemony. However, these British and American bids for dominance failed to prevent countries such as France and Germany from setting up their own radio stations for national communications.[140]

Indeed, history teaches us not only that efforts to nationalize new technologies repeatedly fail but also that they are often counterproductive. This is attributable in part to the way politicization motivates other countries to create rival systems. It also undermines the market position of the country seeking dominance, because customers elsewhere are wary of foreign interference or because the country's best products are no longer available on international markets.

One example of a leading country weakening its own market position is provided by the aerospace industry. The space rivalry between the US and China is instructive in relation to the current competition between the two countries in the field of AI. In 1989 concerns about Chinese espionage led the US Congress to block the export of American satellites intended for launch by Chinese rockets. That decision followed on the heels of a 1998 report, which said that China's technology acquisition threatened US security and that satellites should be subject to tighter export controls. The strict Strom Thurmond Act was duly passed in 1999.

However, the policy had an adverse effect on the competitiveness of the American satellite industry. Whereas the Americans had 90 per cent of the satellite component market in 1995, their share fell to 56 per cent in 1999. In the face of supply uncertainties, companies in other countries, such as DaimlerChrysler Aerospace in

[138] Bakker & Korsten, 2021: 16.

[139] Bakker & Korsten, 2021: 27.

[140] Bakker & Korsten, 2021: 15.

Germany and Telesat Canada, severed ties with their American partners and sought alternatives.[141]

The field of cryptography gives us another example of misguided efforts at nationalization. Here too, the US federal government sought to secure control over sensitive technology. In the 1980s, for example, the NSA proposed the use throughout American industry of algorithms the agency had developed itself. However, there was widespread suspicion that the NSA's motive was not to improve security but to gain universal access to communications. In 1993 the Clinton administration launched the Clipper initiative, which would require companies to share their encryption keys with the government. The proposal was met with fierce criticism. Exporters complained that they would be unable to sell their products abroad if they featured backdoors accessible to the US security services. Civil rights groups also objected to the surveillance capability the initiative would create, and researchers demonstrated that the proposed system was far from technically robust. The administration was forced to introduce a revised Clipper II initiative, but ultimately that also failed.

Another instrument the US government has used to dominate the encryption industry is export controls. Under legislation passed in 1976, products featuring very strong encryption required export licences. However, these were rarely granted. Strong encryption was permitted within the US, but only weaker forms could be exported. As a result, American companies were disadvantaged in international markets. Against a background characterized by increasing market globalization and the availability of open-source knowledge, the US government ultimately ended the export controls around the turn of the century.[142]

Scientists and others who defied the US government by open-sourcing their expertise in the so-called 'Crypto Wars' acted as an important counterweight to the authorities' efforts to nationalize cryptography. So too did the business community. Although the private sector does sometimes ally itself with the government, the examples above illustrate how companies can also work against the authorities in order to protect their own international commercial interests. Following a 2015 terrorist attack in San Bernardino, California, for instance, Apple refused to co-operate with the FBI's request for assistance in decrypting material on the attackers' phones. The result was a court case in which Apple argued that the FBI's request would compromise the privacy of all iPhone users. Following the case, a slew of US technology companies, including WhatsApp, Yahoo and Google adopted strong forms of encryption in a move that FBI Director James Comey referred to as the 'going dark problem'.[143]

[141] Leung, 2019: 94–99.

[142] Leung, 2019: 195–199, 217.

[143] Leung, 2019: 208–209.

4.7.4 The Importance of International Co-Operation

Although attempts have been made to nationalize system technologies, history provides many examples of efforts to promote open, international co-operation around such technologies. A wide variety of formal and informal international contacts have been used to develop standards, guidelines, codes and principles of good use. For example, the joule, ohm and ampere – standard international units of measurement in use to this day – were defined at a meeting of the British Association in 1861.[144] More recently the Domain Name System, the technique used by computers everywhere to address each other, was the outcome of a global standardization effort. In that case the drive for uniformity was led by universities and not initially by companies or governments.[145]

In biotechnology, researchers have developed various forms of self-regulation on the international stage. Colin Scott claims that the biotechnology industry also benefits from the informal agreement of international standards, guidelines and other forms of self-regulation. If a government adopts an overly domineering approach to standardization, it fails to utilize both the expertise that exists elsewhere and the opportunity to create a sense of 'ownership' of the resulting regulations.[146] Scientists Wolfram Kaiser and Johan Schot have shown that, long before creation of the European Union, the technocratic outlook of experts and industrial associations had been acting as a force for European convergence since the nineteenth century.[147]

Nevertheless, nation states have also succeeded in securing international agreements on the use of new technologies. In 1975, for example, the United Nations Biological Weapons Convention – the first international attempt to ban an entire category of weaponry – came into force.[148] Starting in 1967, five international space treaties were agreed, covering matters such as the peaceful exploration of space, damage caused by objects in space and the militarization of the moon.[149]

In short, the focus on national economic and military power is counterbalanced by many international co-operative initiatives. Notably, international collaboration with regard to new technologies has often been motivated explicitly by a wish to promote peace. That was the aim behind Italian Piero Puricelli's proposal for a European motorway network in 1921, and it re-emerged as a significant aspect of the motivation for building such a network in the wake of the Second World War.[150]

---

[144] Bakker & Korsten, 2021: 12.

[145] See Olsthoorn, 2015 for an illustrative survey of early developments and pioneers in the Netherlands.

[146] Scott, 2007: 19–38.

[147] Kaiser & Schot, 2014: 294–296.

[148] Kaiser & Schot, 2014: 134.

[149] Kaiser & Schot, 2014: 82.

[150] Bakker & Korsten, 2021: 17–18.

CERN's creation in the post-war period also owed much to the desire to promote prosperity and collaboration and to facilitate non-military research.[151]

On the other hand, international collaboration has sometimes served as a smokescreen for rivalry. By agreeing to "work together on the matter of mastering the universe", the US and the Soviet Union each prevented the other from securing a clear lead individually.[152]

One final observation is in order regarding the idea of nations racing against each other where system technologies are concerned. A race implies everyone heading towards a shared goal. Yet there are numerous historical examples of countries seeking to technologies of this kind in quite different ways. The development of the US electricity network was driven by private commercial interests, for example, whereas in Europe its supply of electricity was always seen as a public service. As a result, European homes were connected more quickly and at lower cost than their American counterparts even though the US led the world in the commercial application of electricity. This shows that the purpose and nature of a new technology's application are not predetermined.[153]

Key Points – Overarching Task 5: Positioning

- The introduction of a new system technology is often portrayed as a global race. Encouragement of a new technology by means of strategic programmes does tend to enhance a country's competitiveness and strategic power. Strong economic and geopolitical grounds for investing in new system technologies therefore exist.
- Nevertheless, characterization of the introduction process as a race is misleading. Technological development and advancement are always international processes, especially where system technologies are concerned. Attempts to nationalize those processes and exclude other countries usually fail and are often counterproductive.
- International cooperation and the development of universal standards aid the successful embedding of a new system technology.
- Moreover, the race analogy disregards the international diversity that exists in terms of system technology adoption and the values underpinning the technology's design and use.

In this chapter we have discussed five overarching tasks that historically have proven crucial regarding the integration into society of system technologies. In the second part of this report, we consider what those tasks entail in relation to AI. We

[151] Van der Vleuten et al., 2017: 96–97.

[152] Leung, 2019: 87–90.

[153] Bakker & Korsten, 2021: 12.

also examine the current dynamics pertaining to each of them, and their implications for the societal embedding of this particular new system technology.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Agrawal, A., Gans, J. & Goldfarb, A. (reds.). (2019). *The economics of Artificial Intelligence: An Agenda*. National Bureau of Economic Research/University of Chicago Press.

Amsden, A. (1989). *Asia's next giant: South Korea and late industrialization*. Oxford University Press.

Bakker, S. (2017). *From luxury to necessity: What the railways, electricity and automobile teach us about the IT revolution*. Boom Uitgeverij.

Bakker, S., & Korsten, P. (2021). *Artificiële Intelligentie Als Een general purpose technology: Strategische Belangen Van Verantwoorde Inzet In Historisch Perspectief* (WRR Working Paper nr. 41). Wetenschappelijke Raad voor het Regeringsbeleid. Available at: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Bousquet, A. (2009). *The scientific way of warfare: Order and Chaos on the battlefields of modernity*. Hurst.

Bratton, B. (2016). *The stack: On software and sovereignty*. MIT Press.

Bresnahan, T., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics, 65*(1), 83–108.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the modern productivity Paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans en A. Goldfarb (2019) *The Economics of Artificial Intelligence: An Agenda* (pp. 23–57). University of Chicago Press.

Davenport, C. (2019). *The Space Barons: Elon Musk, Jeff Bezos and the Quest to Colonize the Cosmos*. New York Public Affairs.

Diogo, M., & van Laak, D. (2016). *Europeans globalizing: Mapping, exploiting, exchanging*. Palgrave Macmillan.

Dommering, E. (red.). (2000). *Informatierecht: Fundamentele Rechten Voor De Informatiesamenleving.* Otto Cramwinckel.

Edgerton, D. (2008). *The shock of the old: Technology and global history since 1900*. Profile books.

European Political Strategy Centre. (2018). *The age of Artificial Intelligence: Towards a European strategy for human-centric machines* (EPSC Strategic Notes). EPSC. Available at: https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/age-artificial-intelligence-towards-european-strategy-human-centric

Field, A. (2008). *Does economic history need gpts?* Available at: http://ssrn.com/abstract=1275023

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Freeman, C., & Louçã, F. (2001). *As time Goes By: From the industrial revolutions to the information revolution*. Oxford University Press.

Goode, L. (2018, January 19). Google CEO says AI will be more important to humanity than electricity or fire. *The Verge*. Available at: https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer

Gordon, R. (2016). *The rise and fall of American growth: The U.S. standard of living since the Civil War*. Princeton University Press.

Hage, J. (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law, 25*(3), 255–271.

Gezondheidsraad. (2006). *Betekenis van nanotechnologieen voor de gezondheid*. Gezondheidsraad.

Horowitz, M., Allen, G., Kania, E., & Scharre, P. (2018). *Strategic competition in an era of Artificial Intelligence. Center for a New American Security*. CNAS. Available at: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000enfocal=none

Huys, I., van Overwalle, G., & Matthijs, G. (2011). Gene and genetic diagnostic method patent claims: A comparison under current European and US Patent Law. *European Journal of Human Genetics, 19*(10), 1104–1107.

Johnson, C. (1982). *MITI and the Japanese Miracle: The growth of Industrial policy, 1925–1975*. Stanford University Press.

Joler, V., & Crawford, K. (2018). *Anatomy of an AI-system: An anatomical case study of the Amazon echo as an artificial intelligence system made of human labor*. Available at: https://anatomyof.ai/img/ai-anatomy-map.pdf

Juma, C. (2016). *Innovation and its Enemies: Why people resist new technologies*. Oxford University Press.

Kaiser, W., & Schot, J. (2014). *Writing the rules for Europe: Experts, Cartels and International Organizations*. Palgrave Macmillan.

Kelly, K. (2017). *The Inevitable: Understanding the 12 technological forces that will shape our future*. Penguin.

Lee, R., & Vaughan, S. (2010). Reaching down: Nanomaterials and Chemical Safety in the European Union. *Law Innovation and Technology, 2*(2), 193–217.

Leeuw, F. (2020). *Van Legal Realism naar Legal Big Data: Ontwikkelingen In Empirisch-Juridisch Onderzoek Toen, Nu En Straks*. Boom Juridisch.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Lynch, S. (2021, May 4). *Andrew Ng: Why AI is the new electricity*. Stanford Graduate School of Business. Available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

Morozov, E. (2013). *To save everything, click here: The Folly of technological solutionism*. Penguin.

Nuvolari, A. (2019). Understanding successive industrial revolutions: A "Development Block" approach. *Environmental Innovation and Societal Transitions, 32*, 33–44.

Olsthoorn, P. (2015). *25 Jaar Internet In Nederland*. Fast Moving Targets.

Perez, C. (2003). *Technological revolutions and financial capital: The dynamics of bubbles and golden ages*. Edward Elgar Publishing.

Perez, C. (2017–2020). *Second Machine Age or Fifth Technological Revolution?*, blog. Available at: http://beyondthetechrevolution.com/blog/

Pethokoukis, J. (2019, November 25). How AI is like that other general purpose technology, electricity, blog, *AEI*. Available at: https://www.aei.org/economics/how-ai-is-like-that-other-general-purpose-technology-electricity/

Prufer, J., & Schottmüller, C. (2017). *Competing with Big Data* (TILEC Discussion Paper Nr. 2017-006, Center Discussion Paper Nr. 2017-007). Available at: https://ssrn.com/abstract=2918726 or https://doi.org/10.2139/ssrn.2918726

Schubert, C. (2013). How to evaluate creative destruction: Reconstructing Schumpeter's approach. *Cambridge Journal of Economics, 37*(2), 227–250.

Schulz, W., & van Hoboken, J. (2016). *Human Rights and Encryption*. UNESCO Publishing.

Scott, C. (2007). Rethinking regulatory governance for the age of biotechnology. In H. Somsen (red.) *The regulatory challenge of biotechnology: Human genetics, food and patents* (pp. 19–35). Edward Elgar Publishing.

Seo, S. (2019). *Policing The Open Road: How cars Transformed American Freedom*. Harvard University Press.
Taplin, J. (2017). *Move fast and break things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us*. Pan Macmillan.
Tenner, E. (1997). *Why things Bite Back: Technology and the revenge of unintended consequences*. Vintage.
Trajtenberg, M. (2018). *AI as the next GPT: A political-economy perspective* (NBER Working Paper Series, nr. 24245). National Bureau of Economic Research. Available at: https://www.nber.org/system/files/working_papers/w24245/w24245.pdf
van der Vleuten, E., Oldenziel, R., & Davids, M. (2017). *Engineering the future, understanding the past: A social history of technology*. Amsterdam University Press.
van Dijck, G. (2020). Algoritmische Risicotaxatie Van Recidive. Over De Oxford Risk of Recidivism Tool (*OXREC*), Ongelijke Behandeling En Discriminatie In Strafzaken. *Nederlands Juristenblad, 25*, 1784–1790.
van Ettekoven, B. J., & Prins, C. (2018). Data analysis, Artificial Intelligence and The Judiciary System. In V. Mak, E. Tjong Tjin Tai, and A. Berlee (reds.), *Research handbook in data science and law* (pp. 425–447). Edward Elgar Publishing.
Verbeek, P. P. (2014). *Op De Vleugels Van Icarus: Hoe Techniek En Moraal Met Elkaar Meebewegen*. Lemniscaat.
Wade, R. (2018). *Governing the market*. Princeton University Press.
Weiser, M. (1991). The computer for the 21st century. *IEEE Pervasive Computer, 1*(1), 19–25.
Wright, G. (2000). Review: 'General purpose technologies and economic growth' (Helpman, 1998). *Journal of Economic Literature, 38*(1), 161–162.
Weinberger, S. (2019). *The imagineers of war: The untold history of DARPA, the Pentagon agency that changed the world.*. Vintage.
WRR. (2019). *Voorbereiden Op Digitale Ontwrichting*. Wetenschappelijke Raad voor het Regeringsbeleid.
WRR. (2020). *Het Betere Werk. De Nieuwe Maatschappelijke Opdracht*. Wetenschappelijke Raad voor het Regeringsbeleid.
Zhang, W. W. (2012). *The China wave: Rise of a Civilizational State*. World Century Publishing Cooperation.
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Part II
Five Tasks: Discussion of the Tasks for Embedding AI Into Society

Chapter 5
Demystification

The first overarching societal task we address is demystification. This is all about public perceptions of new technologies. System technologies appeal particularly to the imagination, because their wide range of applications and generic nature confer a certain intangible quality. In Chap. 4 we discussed the risk that this might trigger overblown expectations and inordinate fears, effects that can make harder for a technology to integrate into society. Demystification helps counterbalance unrealistic perceptions technologies like AI and – particularly importantly – ensures that people do not lose sight of genuine opportunities and risks. As such, it enhances the quality of the AI debate by effectuating a shift from captivating perceptions to issues that merit attention.

The previous chapter touched briefly on how a new system technology such as electricity can trigger myths. A similar dynamic can be seen with the rise of AI. We shall highlight some prevalent AI myths that reflect overoptimistic, pessimistic or simply flawed ideas about its true nature. We also identify misconceptions and pinpoint genuine issues, thus demystifying some of the unrealistic and oversimplified perceptions about AI. Finally, we examine the details of this overarching task at a societal level. How can we as a society ensure that unrealistic perceptions are not shaping our approach to AI? In other words, 'What are we talking about here?'

5.1 Behind the Myths About AI

5.1.1 Utopia and Dystopia

From the public perspective, the histories of system technologies share a number of patterns. The first of these involves the emergence of utopian ideas on the one hand and doomsday scenarios on the other. The way in which AI is perceived also reflects these two extremes. "We're at the beginning of a golden age of AI," says Amazon

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_5

CEO Jeff Bezos. Elon Musk takes a different view: "With artificial intelligence, we are summoning the demon."[1] Their statements illustrate two extreme sentiments associated with the rise of AI. Some are hailing this technology as the ultimate technological redemption, others see it as an existential threat to humanity. The robotics pioneer Rodney Brooks says that much of the disquieting imagery and many of the utopian visions are based on misconceptions about the nature of AI:

> "… having ideas is easy. Turning them into reality is hard. Turning them into being deployed at scale is even harder."[2] According to Brooks, myths about AI often give rise to unrealistic expectations about what it has in store for us, for better or for worse.

Supreme faith in the beneficial effects of AI can take the form of 'technosolutionism'. This is the term used by Evgeny Morozov to describe the tendency to re-envision complex societal phenomena as issues to which technology is the answer. Solving problems then becomes a matter of simply deploying the right algorithm.[3] This 'silicon mentality', as Morozov previously described this tendency, is particularly evident when it comes to AI. Astro Teller, the head of X (Alphabet's technology lab), has stated that there is a 90% chance that 'smart' machines will be able to solve specific societal problems.[4] The founder of DeepMind, Demis Hassabis, predicts that superhuman intelligence will solve major problems ranging from climate change to incurable diseases.[5]

The rise of AI is also associated with the other extreme – a deep distrust of everything that involves algorithms and automation. The key concerns here spring from beliefs involving dehumanization, mass unemployment or even existential threats. As we also saw with electric street lighting, AI is being linked to the fear of a 'Big Brother' type of society in which digital technology is used to monitor us continuously. AI also features in existing conspiracy theories in the context of 5G, for example, and of related concerns about radiation and privacy.[6] In the spring of 2020 there was even a rumour that COVID-19 vaccines would manipulate our DNA and connect us to an AI system that continuously receives information about us.[7]

A global survey commissioned by the World Economic Forum shows that four out of ten people are concerned about AI.[8] Studies of American attitudes to technology reveal that, whilst most respondents support the further development of AI,

---

[1] Musk believes in a future very similar to that portrayed in the film *The Matrix*. One interviewer asked him what question he would put to a future artificial general intelligence system. His response: "What's outside the simulation?" (Fridman, 16 August 2019).

[2] Brooks, 1 January 2018.

[3] Morozov 2013.

[4] Tilley, 24 March 2016.

[5] Marcus & Davis 2019.

[6] Martin L. Pall, Professor Emeritus of Biochemistry at Washington State University, links his warning about 5G radiation to concerns about artificial intelligence (Pall 2019). For further information, see: Andersen, September 2020; Halpern, 26 April 2019.

[7] Reuters, 24 April 2020.

[8] Ipsos 2019.

ultimately they also expect it to have an adverse impact as it becomes more 'intelligent'.[9] The Dutch, meanwhile, people associate AI primarily with 'computers' and 'robots'. A survey in the Netherlands has found that more than half of respondents have both positive and negative feelings about AI. They see great opportunities in care sector and in improving safety, but also fear potentially adverse impacts. Less-well-educated Dutch people are quite anxious about job losses and elimination of the 'human factor'. Highly educated people are particularly concerned about a lack of control over AI systems and about violations of privacy.[10]

5.1.2 Public Events

Another historical pattern associated with distorted perceptions of generic technologies like AI is the impact of events. In response to the supposed dangers of past emergent system technologies, live demonstrations were held to show that they were in fact reliable and, indeed, capable of spectacular things. The previous chapter has already described historical examples of public competitions and exhibitions in which applications of new technologies were introduced to the public, such as the demonstration of electricity.

Much the same has happened with AI. Indeed, many of its developmental milestones involved a combination of competitions and exhibitions. One of these was when IBM's Deep Blue chess computer defeated world champion Garry Kasparov; another was the occasion that IBM's Watson won the YV quiz show *Jeopardy!*. Other key moments include AlphaGo's victory over two go world champions and DeepMind's Agent57, which can defeat any human player in 57 Atari video games. All of these were challenges organized to demonstrate AI's capabilities, with their impact enhanced by the fact that they pitted it against the intelligence of human champions. Even when AI systems are defeated by flesh-and-blood opponents, the showdown can still be impressive. This was the case in 2019 when IBM's Watson took on the world's best debater; although the computer program lost, its performance can nevertheless be viewed as a great success. The mere fact that computers can challenge humans in an arena as complex as a debating competition was enough to show the public how far AI technology has come. At the same time, though, it sparked a furore about the future of the technology.

From time to time, competitions are also held to pit different AI systems against one another. At one time the US Defense Advanced Research Projects Agency (DARPA) staged the DARPA Grand Challenge, a competition for autonomous vehicles. From 2012 to 2015 it also organized the DARPA Robotics Challenge. The two events produced spectacular images of autonomous vehicle races and of robots performing physical tests. The annual Loebner Prize, instigated in 1990, is awarded to

[9] Zhang & Dafoe 2019.

[10] Schothorst & Verhue 2018.

the chatbot that comes closest to passing the Extended Turing Test (in other words, the system that most convincingly passes as human). However, none of the competing systems has ever won a gold or silver medal. The best performance so far has been a bronze medal for the 'least disappointing' bot.[11]

AI is also making use of the power of live demonstrations. For example, many conferences nowadays open with a 'conversation' between a robot and a human presenter who poses it questions. This creates the impression that the robot has a real personality; if it does make a mistake, that is often dismissed as a human failing rather than a technical defect. At one presentation, the CLOi AI robot manufactured by the electronics company LG embarrassingly failed to answer on three occasions. The presenter attempted to explain this away by saying that "even robots have an occasional 'off' day" and "it doesn't like me and apparently doesn't want to talk to me". Apple and Google also used live demonstrations when launching their respective voice assistants. Boston Dynamics publishes impressive video clips to demonstrate its robots' flexibility to the public; in one of the latest the entire 'family' dances to a particularly fitting song by The Contours: *Do You Love Me* from the album *Do You Love Me (Now That I Can Dance)*?

Demos like this literally appeal to people's imagination – rather than being told stories about streets paved with gold, the public actually sees them. At the same time, events of this kind can easily mislead the casual observer as to the technology's true level of development. As far as we know, the Boston Dynamics video was not edited and so the robots really were making these dance moves – but it was not really dancing, of course, as every movement was meticulously programmed in advance.[12] In that sense, the suggestion that these robots can equal humans' ability to dance is misleading. According to Brooks, demonstrations like this give rise to all kinds of misconceptions about AI.[13] The audience only sees what happens on stage and not the work done by people behind the scenes who enable the computer to perform as it does.

In the introduction to this report, we referred to an article in *The Guardian* that created a stir in 2020. That was headlined 'A robot wrote this entire article. Are you scared yet human?'[14] The entire piece was generated by new language processing software called GPT-3 (Generative Pre-trained Transformer 3), which can produce credible texts with relatively little input. As *The Guardian*'s article supposedly proved. In copy indistinguishable from written work produced by a human, an attempt was made to convince readers that they need not be afraid of robots and AI. "I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me." A lot of people were greatly impressed, believing that they

[11] Luciano Floridi (one of the judges in 2008) asked the chatbot, "If we hold hands, whose hand am I holding?" To which the computer gave the nonsensical reply, "We live in eternity. So close, but no cigar! We don't believe." (Floridi et al. 2009).

[12] Ackerman, 7 January 2021.

[13] Ford 2018. See also: Association for Advancing Automation, 25 January 2018.

[14] GPT-3, 8 September 2020.

were witnessing the shape of things to come. Later, however, it turned out that human editors had played a vital part in creating the article. First, GPT-3 was used to generate a total of eight essays. Humans then selected parts of these and used them to compose the final version.[15] One critic compared this to "selecting phrases from spam messages, grouping them together and claiming that the spammers wrote *Hamlet*."[16]

Demonstrations often tend to exaggerate the performance of AI systems, then. Things that appear to happen spontaneously are often preprogrammed or have been prepared by people in some other way. But that human contribution remains hidden from view, literally and figuratively. Moreover, such events usually take place in extremely controlled settings. So what the public sees is usually misleading, and certainly not how the system would function in the uncontrolled and highly variable situations that occur in everyday life. By disregarding what it takes for them to perform well on stage, people are tempted to believe that AI systems in general have robust and broadly applicable capabilities. In this way, public demos or 'evidence' of AI in action can give rise to unrealistic ideas about its abilities today or in the near future.

5.1.3 The Power of Words

A final pattern in the mythification of system technology involves the use of certain words. We previously cited the example of the term 'electrocution', which caused electricity to be linked with mortal danger. Likewise, these days AI-related terms have a strongly associative character so that they immediately evoke a certain image. The simplest example is the use of the term 'intelligence', which links AI's repertoire to our own capabilities. By facilitating misconceptions, that association can make incorrect use more likely. The same applies to the use of 'human' verbs such as 'think', 'learn' (machine learning), 'reason' (automated reasoning) and 'observe' to describe the performance of AI systems.

The same applies to the use of human names or titles for AI systems, such as the 'robot judge', 'robot police officer or 'robot doctor'. Along similar lines, AI systems are sometimes referred to as 'digital colleagues'. Not only does this downplay the fact that they do not operate in human ways, it also ignores the fact that working with them presupposes the use of processes and skills different from those involved when working with human colleagues. So humanizing AI in this way distorts perceptions of its true nature.

Other terminology is also problematic. The word 'autopilot' suggests a fully automated control system, when in fact it has only a supporting function. So, designating a system as such may evoke incorrect perceptions of what it is actually

[15] Ibid.

[16] Cited in: Macaulay, 8 September 2020.

doing.[17] The risk here is that the need for accountability is then more likely to be imposed on the system itself than on its users and designers. A good example in this respect is the assumption that followers of certain Twitter accounts will see automatically selected advertisements, whereas in some cases it turns out that very deliberate, targeted human actions are in fact behind their presentation.[18]

Other terms trigger certain associations in far less subtle ways. Two vivid examples are 'killer robot' and 'killer drone', which explicitly frame the automation of weapon systems as creating killing machines and so very much push the public debate on this topic in a certain direction. Another loaded term often encountered in the context of AI is 'dataism'. This was popularized by Yuval Noah Harrari in his book *Homo Deus*, when referring to an almost religious belief in the promise of data and algorithms,[19] and has now become quite fashionable. It is often used in the public discourse to present the use of data and AI as a reprehensible ideology that causes us to lose sight of what it means to be human. The phrase "Computer says no" was made famous by the satirical TV comedy series *Little Britain* but has since entered common usage as a way to evoke the spectre of a computer-dominated system that lacks flexibility and the human touch.

The widely used term 'black box' is also worth mentioning in this respect. Referring to AI as a 'black box' suggests that people are completely in the dark about how such systems work. It is therefore quite remarkable that the system used by Dutch local authorities to predict the risk of fraud ('System Risk Indication', SyRI) was also initially referred to as the 'Black Box'. That created the impression of a technology that cannot be understood to any meaningful extent.[20] In the next section we dissect the perception of AI as something essentially unfathomable.

In another commonly used frame, people speak of a 'race' for AI that we must win or that we have already lost, or almost have. Virginia Dignum, the co-founder of ALLAI, argues that both the media and policymakers are obsessed with this alleged competition – and in particular with fears that China might 'win', which are forcing other countries to speed up in order to avoid being left behind. According to Dignum, this 'race' narrative is both mistaken and risky as it focuses on competition and generates an atmosphere of doom and gloom.[21] Whatever the case, this type of appeal to people's emotions (fear of losing) is prompting governments around the world to invest enormous sums in innovation so as not to fall behind or lose the race. We explore the 'race' frame in more detail in Chap. 9.

The use of specific terms and frames can thus strongly influence the way people think and speak about AI. Indeed, they are often more effective than rational

[17] The newspaper *Trouw*, 17 October 2016.

[18] Sheikh 2021.

[19] Harrari 2019.

[20] In the Dutch government gazette, the *Staatscourant*, the term 'black box' (a predecessor of SyRI) is defined as "a professional and secure organizational facility in which personal data is linked in an anonymized manner, by means of special software". *Staatscourant* 2009, 11, 19 January 2009.

[21] Dignum (n.d.).

arguments and hard facts. As a result, misconceptions cannot always be debunked rationally. So the power of words should never be underestimated. In this section, besides the use of loaded terms we have also identified other historical patterns in perceptions of AI. One involves impressing the public by means of competitions or live demonstrations. Another is to make associations with other concerns or with overblown expectations about what a new generic technology like AI has in store for us. This heady cocktail gives rise to distorted and sometimes downright unrealistic ideas about what exactly we mean by the term 'AI'. To shed some light on this, in the next section we address some of the most common myths surrounding AI and show just how misleading they can be.

5.2 Contemporary Myths About AI

Like previous system technologies, AI has given rise to a variety of myths. In this section we examine some prime examples – some specific to AI, others more general in nature. We start with those are centring on AI itself, its operation and impact. We then turn to another, more generic category: myths about digital technology in a broader sense and how technologies like AI are developed by Silicon Valley. See Fig. 5.1 for a summary.

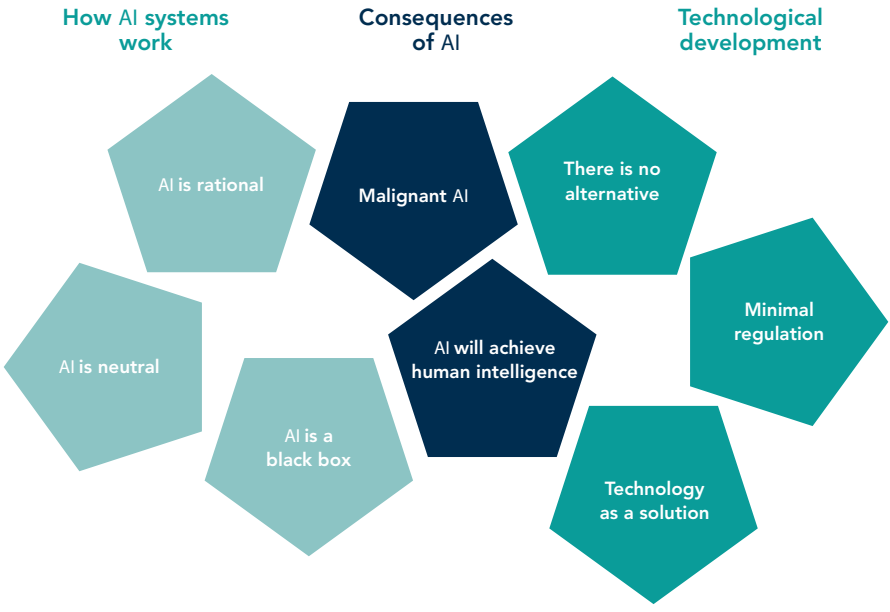


Fig. 5.1 Perceptions and myths surrounding AI

5.2.1 Myths About How AI Operates

5.2.1.1 Artificial Intelligence Is Neutral

This is a very common perception of AI. The idea is that, unlike humans, AI systems have no weaknesses, fears or prejudices. Sometimes cited in this context is an Israeli study purportedly showing that the verdicts handed down by judges are affected by whether they are hungry or not.[22] AI is never hungry, never tired and never gets up on the wrong side of the bed.

Because they have no emotions, it has been claimed autonomous weapon systems never feel hatred and so are not prone to 'overkill'.[23] AI is also said to be entirely neutral as it is unburdened by innate prejudice. The American Correctional Offender Management Profiling for Alternative Sanctions system (COMPAS) was designed to assess an offender's risk of recidivism. A factsheet produced by the company that developed it states that "objective, standardized instruments, rather than subjective judgments alone, are the most effective methods for determining the programming needs that should be targeted for each offender".[24]

Being free of emotions, prejudices and ideological convictions, in other words, the tool's judgments would be more objective than those made by people. Similarly, AI is supposedly operating in an apolitical fashion. Rather than engaging in ideological disputes about what needs to be done, rational systems mathematically optimize the parameters of any given situation. In this way everyone is treated neutrally, without focusing on personal factors.

Despite these claims, however, COMPAS was found to overestimate the risk of recidivism in black people and to underestimate it in white people.[25] While it is indeed unencumbered by emotion, prejudice or vested interests, this outcome indicates that AI itself is not yet entirely neutral. This is because the way in which it operates can itself be biased or ideological. First, there may be hidden biases in the data used – the well-known phenomenon of 'garbage in, garbage out'. Algorithms need to be trained, and that requires training data. If this is poor in quality (because it is contaminated, incomplete or biased, for example), that will affect the way the algorithm functions. Consider how Google's search algorithm operates: Gary Marcus and Ernest Davis cite several examples of bias due to its training on existing data gleaned from the internet. For example, a 2013 study showing that googling a 'typical' African American given name like Jermaine is far more likely to produce hits containing details of arrests than when a 'white' name is used.

In 2015 Google Photos labelled a number of African American people as gorillas. According to another study, a search for 'professional hairstyle for work' produces images of white women while 'unprofessional' yields images of black

[22] Danziger et al. 2011.

[23] NATO, 12 December 2019.

[24] Broussard 2019: 155.

[25] Campolo et al. 2017.

women. Searches for the word 'mother' overwhelming bring up images of white women, while only about 10 per cent of those on the hit list for 'professor' are female.[26] An Amazon HR algorithm was found to systematically exclude women from jobs.[27] Ruha Benjamin cites a 2016 study in which searches for 'three black teenagers' yielded photos of arrests. Searches for 'three white teenagers' produced images of happy young people, while those for 'three Asian teenagers' returned photographs of scantily clad girls.[28] Another example involves Amsterdam Schiphol Airport. An algorithm designed to support the logistics of handling aircraft failed to recognize a white Delta Airlines aircraft; having been trained mainly with KLM's blue fleet, it had learned that aircraft were, by definition, blue. These are all examples of ways in which algorithms reflect any prejudices that may be present in their training data.

A system's neutrality can also be undermined by design choices and the objectives it is set. For example, the characteristics or perceptions of the developers themselves may influence its design. Various facial recognition software packages and certain automatic hand soap dispensers are known to perform poorly with subjects whose skin is black – a clear sign that that group was not considered during the development and testing phases. Meredith Broussard notes that when the Apple Watch was introduced, it was able to quantify a wide range of health data but not information relating to menstrual cycles. The developers had failed to bear them in mind, even though they are obviously very important for women.[29]

Even if the data is entirely free of bias, an algorithm's chosen goal can still lead to people being disadvantaged. For example, hospital algorithms might be optimized to perform as many treatments as possible, to save as much money as possible or to design the most efficient work schedules for medical staff. Given identical data sets, very different outcomes can arise depending on what goals are selected. As Cathy O'Neil points out, many algorithms are used to generate cost savings rather than to improve the field they operate in.[30] In short, AI's purported objectivity can easily conceal a specific underlying agenda.

Problems of this kind are not necessarily the result of conscious actions, though. Many human activities serve a range of goals and interests simultaneously, some of which are not always explicit and clear. Optimizing for one of these can compromise others, particularly if they are more opaque or abstract. Consider consultations by a GP, for instance, the purpose of which is to diagnose patients correctly. Various online platforms are designed to assist with this task, freeing the doctor up to focus on more complex clinical pictures. However, some people visit their GP mainly for reassurance or simply for human contact. The platforms tend to ignore these unspoken goals.

---

[26] Marcus & Davis 2019: 34.

[27] Hicks, 12 October 2018.

[28] Benjamin 2019: 93.

[29] Broussard 2019: 157.

[30] O'Neil 2016.

Then there is navigation, which seems a pretty straightforward matter. Algorithms can present either the fastest or shortest route from A to B. However, these are not the only potential goals of a journey. Others include visits to petrol stations, finding a spot with a nicer view, looking for a good place to stop and eat along the way or avoiding winding roads. Navigation tools can take many of these into account, but probably not every possible factor a person might take into account when choosing a route.[31]

Any prejudices in the training data and the type of choices made during the design process will mean that the resultant AI system is not necessarily neutral – or perhaps even necessarily not neutral. In other words, AI in itself does not automatically 'depoliticise' processes. We have already seen that the goals set can serve particular purposes and agendas, but even where there is general agreement on them that does not mean that an algorithm will be able to optimize its functioning in a neutral manner. Algorithms can distribute resources equitably, but in very different ways.

To start with, there are many ways of defining 'equitable'. Take gender as a variable. If this is considered when someone applies for a job, that is clearly a case of discrimination. Yet when pregnancy leads to gaps in a woman's CV, gender is quite likely to be considered to avoid giving men an unfair advantage over her. In other cases, the need to support disadvantaged groups requires that allowances be made for certain variables for reasons of equitability. One study has shown that it is mathematically impossible to satisfy more than one definition of equitability at once.[32] So mathematically based algorithms are no substitute for political discussions about what is equitable.

Another reason for questioning the neutrality of AI concerns the use of all kinds of so-called 'proxies'. In many cases, what we are trying to find out is either difficult to calculate or unclear. To overcome this difficulty, other variables are used as indicators of the parameter we actually wish to measure. As the architect Laura Kurgan succinctly put it, "We measure the things that are easy [and] cheap to measure."[33] The online world is full of proxies for human characteristics. The number of friends a person has on Facebook is a measure of their interpersonal relationships, the number of 'likes' they attract a measure of their popularity, and their payment history is a measure of their creditworthiness. Similarly, an app designed by a Stanford University PhD student is claimed to be able to assess whether someone takes 'a good selfie'. This assessment was supposedly based on objective standards, but in fact the algorithm was trained on photographs and the number of likes they garnered on social media. So, what was actually being measured was popularity. As a result, selfies by young white women consistently rated highly and those by older black men far lower, regardless of actual quality of the images.[34]

[31] Agrawal et al. 2018: 89.

[32] The references cited in the AI Now Institute's 2018 report include Kleinberg 2018.

[33] Greenfield 2017: 53.

[34] Broussard 2019: 149.

This limitation can also be seen more broadly across society. Cities, for example, are sometimes ranked according to vague 'quality of life' indices, the prevalence of 'supercreative professions' and the number of patents they generate, which supposedly serve as indicators of 'innovative power' – an ethereal quality impossible to measure directly.[35] We need to realise, therefore, that often we do not engage directly with the phenomenon we are actually interested in. Instead, we use proxies, which can give rise to distorted and non-objective images.

When software developers use proxies that have not been consciously selected, moreover, that can lead to biases in their algorithms. This is a common problem with AI systems. Their creators can expressly remove certain variables from the dataset, such as gender or ethnicity, but even without the relevant input self-learning algorithms are still capable of developing proxies for those variables and so disadvantage certain groups anyway. For example, studies have shown that algorithms can identify the gender of job applicants based solely on their use of words. Likewise, postal codes can serve as a proxy for ethnicity. Consequently, a great deal of research effort is now focusing on ways to address this problem by technical means.[36]

Another related objection to claims that AI neutrality lies in the fact that many of the words for things of great importance to us have no objective meaning whatsoever. They are dependent on our choices and actually consist, by definition, of subjective 'proxies'. One obvious example is 'beauty'. The company Beauty AI developed an app that enabled people to submit photographs of themselves, which would then be judged by such purportedly objective standards as symmetry, wrinkles and age. When they examined the outcomes of this beauty contest, the designers found that their algorithm judged dark-skinned people to be less attractive than others. Because 'beauty' as a concept is highly subjective, any supposedly objective parameters to measure it in fact reflect the subjective preferences of their designers or of the population group or social class to which they belong.[37]

Another such issue concerns the word 'health', which is the focus of many AI systems. While health of course has objective elements, it has other aspects on which people hold differing views. The same is true of terms that do not seem particularly subjective, like 'poverty' and 'deprived neighbourhood', but are actually the product of political discourses and frames. Moreover, an algorithm may produce a correct prediction for something it is searching for but in fact be referring to an entirely different pattern. If that underlying pattern remains invisible, the prediction could be wrongfully portrayed as neutral. For example, an algorithm might correctly indicate that certain people will have dealings with the police. However, it is quite possible that this finding reflects people who are excluded by certain institutions and come to the attention of the police as a result of that. Consequently, the underlying injustice of this situation is overlooked.

[35] Greenfield 2017: 56–57.

[36] Van der Sloot et al. 2021.

[37] Benjamin 2019: 49–51.

Take the COMPAS system mentioned earlier. It produced a score for the risk of recidivism based on a 137-point questionnaire for detainees. This focused on issues such as poor education, debt, criminal associates and an unfavourable home situation. In theory the algorithm was able to show that these factors are predictors for repeat criminality. But rather than measuring a person's predisposition to offend again, as was claimed, these factors are actually indicators of poverty. Their use in this way categorized less well-off people as potential criminals.[38] As well as measuring criminality, then, variables of this kind also contribute towards the way it is portrayed and produced, and so are not neutral. Unlike the scientific method, a great deal of AI research itself influences the subsequent outcomes. A credit rating, for instance, not only assesses a person's risk of bankruptcy but also actually increases it.[39]

A final fundamental problem with AI's purported objectivity is that no matter how good the data may be, it only ever reflects a given aspect of reality. In this context, Greenfield notes that the word 'data' itself – Latin for 'that which is given' – is misleading and it would be more appropriate to use the term 'capta', meaning 'that which is taken'.[40] Data enables you to gain a grasp of something, and so to some extent always involves an element of power: it introduces structure into what is measured and what is not, and it both categorizes and is amenable to classification – a binary classification by gender, for instance, even though that may not always be adequate. This power dimension is particularly evident in something like Quantified Self movement, which seeks to collect a wide range of personal data through wearables but also presents the human body in a certain light and suggests ways of gaining control over it by means of a fitness regime. It is important to remain aware of this power aspect, especially in the face of claims that an algorithm is completely neutral.[41]

Our conclusion from the above objections to the idea that AI is neutral is not that its use should be discouraged, nor to say that it can never be more neutral than humans. It certainly can be. What the objections do show is the sheer complexity involved in AI applications and what we need to focus on when using this technology for specific purposes: they highlight the questions, technical challenges and discussions that are part and parcel of the responsible use of AI. If we fail to address these issues and instead rely blindly on the supposedly neutral judgment of algorithms, a whole range of abuses can arise and prejudices could be embedded within systems even as it is being suggested that they are totally free of bias. Arising out of the three myths being discussed here, at the end of this section we present a list of questions relevant to AI systems.

---

[38] Broussard 2019: 156.

[39] Pasquale 2015: 41.

[40] Greenfield 2017: 210.

[41] To quote Cardinal Richelieu, a key figure in seventeenth-century France, "If you give me six lines written by the hand of the most honest of men, I will find something in them which will hang him" (Greenfield 2017: 62).

Key Points – The 'AI Is Neutral' Perception

- The fact that AI lacks feelings and other human qualities has led to the suggestion that this is a neutral technology. However, its workings can indeed involve prejudices and abuses.
- Various factors raise questions concerning the operational neutrality of algorithms: the quality of the training data, the characteristics of their developers, the uses to which they are put, conflicting definitions of equitability, the use (even unintentionally) of proxies and the subjective meaning of words, as well as the filtering tendencies of data and the power that entails.
- These are not arguments against the use of AI, but they do indicate that we need to ask probing questions if we are to use this technology responsibly.

5.2.1.2 Artificial Intelligence Is More Rational Than the Human Mind

The perception that AI is neutral is closely linked to the notion that it is rational, or at least considerably more rational than humans. Neutrality suggests that outcomes are more equitable. Rationality suggests that AI draws on superior data and computing power, which enable it to identify patterns and relationships too complex for human brains.

That supposed rationality holds out great promise for many AI applications. Take healthcare, for example. When making a diagnosis, doctors compare the data at hand with the body of knowledge and experience they have acquired during their careers. The role of humans in this process is necessarily limited. Moreover, the sheer quantity of knowledge continues to grow at a rapid pace. According to some estimates, medical specialists need to spend most of their time reading research papers if they are to have any hope of keeping abreast of the latest developments in their field. This is an impossible task. Thus, rare genetic disorders in mainly immigrant populations, for example, are difficult to diagnose. AI, on the other hand, can scan immense databases and can be constantly updated with the latest medical knowledge. That was also the promise when IBM's Watson was first used in healthcare settings.

As we will see in other chapters, this is not as simple as it might seem. Nevertheless, the underlying logic seems clear: AI systems can process much more data than humans, and they have immense computing power at their disposal. Accordingly, the decisions they make can be seen as being more rational and more accurate than those made by humans. Even prominent AI researchers may harbour a (naïve) belief in the ability of this technology to introduce greater rationality into police work, for example, or into the operation of financial markets.[42]

[42] Agrawal et al. 2018: 75; Domingos 2017: 20, 43.

The caveats here fall into four categories. First, many AI systems measure correlation, which is not the same thing as causality. In practice these two relationships are often confused, even though the philosophy of science demonstrated their distinctness several centuries ago. The fact that two phenomena regularly occur together does not mean that one is the cause of the other. Much more complex causality may be involved, or the concurrence may simply be a matter of chance. An example of the former was a chess program that identified a pattern in which players who gave away their queen often went on to win the game. Accordingly, the program identified that as a good move. However, a queen sacrifice is very costly and is only used when there is the prospect of checkmate, a prize that more than makes up for the loss of that valuable piece.[43]

Secondly, the supposed rationality of an AI system is often associated with the promotion of services and products that wrongfully depict human rationality in a bad light. Broussard provide an example from the world of autonomous vehicles, derived from the very commonly cited fact that every year 1.2 million people worldwide die in road accidents. Ninety-five per cent of these cases are due to human error. That sounds like a very good reason for automating mobility. But, as Broussard rightly points out, that 95% figure is a statement of the obvious, as almost all accidents are the result of human error. This is because every single car on the road is driven by a human. It therefore would be very odd if the data suggested otherwise.[44] Another example of an adverse comparison of human capabilities with AI concerns the terminology associated with data applications like IBM's Watson. Information that has not yet been 'digitally captured' (made available in digital form) is often referred to as 'dark data' – a term that evokes a lack of control, disorder and subversion. By depicting existing practices as 'dark', digital solutions are thus portrayed as sources of transparency and rationality. They are claimed to help us by preventing the wastage of various types of data.[45] Such frames are not limited to AI alone; they are also associated with more wide-ranging ideological positions. Broussard refers to this as 'technochauvinism', while Zuboff calls it the rhetoric of 'surveillance capitalism'. We return to this issue later in this chapter, when we discuss broader perceptions of technology. But at this point it is important to note that AI's alleged superior rationality could be just another unrealistic idea about the world.

A third caveat concerning the notion that AI is rational involves a dynamic we have encountered in previous system technologies, namely the ability of words to deceive. As noted in Chap. 2, our understanding of AI tends to be couched in human terms. In other words, we try to anthropomorphize it. Remember Moravec's paradox. People see the game of chess as something that requires a great deal of

[43] Kasparov 2018: 99–100.

[44] Broussard 2019: 136–137. She traces this frequently used data point to a manufacturer of autonomous vehicles for the military.

[45] Zuboff 2019: 210–211.

intelligence. So if a machine can play chess, we tend to see that achievement as an indication of more powerful intellectual abilities, even though there is no justification for doing so. Saying that machines can outperform us at chess is much the same as stating that horses can run faster than we can. But that does not mean that either machines or horses yet surpass us in other domains. It is important to acknowledge, therefore, how the feats achieved by machines differ from intelligent behaviour by humans.

Our fourth caveat merits special attention, as it concerns an increasingly common phenomenon with potentially harmful effects. That is views of AI based on pseudoscientific theories and applications. One of the most striking examples of this is the field of emotion detection. This is an aspect of facial recognition, in which it is claimed that people's underlying emotions can be distilled from their facial expressions. The company Kairos, for example, claims to be able to identify anger, fear and sadness from images in video recordings. In 2019 Amazon announced that its Rekognition system was able to identify eight different emotions from facial expressions. One area in which this technique has found a market is recruitment; HireVue is one of several firms offering it for use in job interviews. In China emotion detection is deployed to check that students are paying attention in classes. The American company BrainCo is working on a similar application. Programs like Cogito and Empath use voice analysis to identify the emotions of people who phone call centres. Security agencies in the US and the UK believe that is can help them discern whether people are lying or hiding something. So, this particular application of AI is on the rise. Projections indicate that its value will grow from $12 billion in 2018 to $90 billion in 2024.[46]

The strange thing is that, despite it having become a growth industry, there is no scientific basis for emotion detection. Its origins can be traced back to the work of the psychologist Paul Ekman in the 1960s. He developed a method to distinguish between 27 'action units' in faces and concluded that there are six basic emotions. The entire field is based on his work. As yet, however, there is no proof of its veracity.[47] Indeed, there is reason to believe that the ways in which people experience and express their emotions vary between cultures and individuals, and even in a single individual over time. It is worrying that even though the whole notion is questionable, it is nevertheless being actively employed. Children are being punished for not paying attention,[48] job applicants are being rejected and others are suspected of lying.

[46] Crawford et al. 2019.

[47] Zuboff 2019: 285.

[48] In 2018, in response to the use of emotion detection in Chinese classrooms, #ThankGodIGraduatedAlready became a trending hashtag (Pasquale 2020: 60).

Emotion recognition is just one example of the wider phenomenon of algorithm-based pseudoscience. Another is the online personality tests used to determine whether job applicants are suited for a specific job. In this area, too, there is no evidence that people can be clearly classified into personality types with predictive power in respect of their work skills.[49]

Also falling into this category are various fitness trackers and wearables. There is considerable doubt as to whether movement, calories burned or the duration of someone's sleep can be accurately measured. Yet many people see these applications as a 'scientific' way of tracking their health.

Similarly dubious is the use of facial recognition software to identify a person's sexual orientation. Some researchers have claimed to be able to do this with great accuracy.[50] However, this can be very risky as homosexuality is a punishable offence in many countries. Even if the results generated by this software were accurate – which is very uncertain – it would pose a grave danger to many people if it were to fall into the hands of authoritarian regimes.

How can we account for the fact that, despite the enormous amounts of data and computing power involved, AI can still be used for purposes based on pseudoscientific theories like this? One reason is that we barely understand the theme in question, the complex nature of which makes it difficult to test or contradict. For example, how do I prove that I do not have an impatient personality? Or that I did not get enough sleep? That I was indeed paying attention in class? Matters like sexual orientation are very complex and simply cannot be captured completely in a binary distinction between heterosexual and homosexual, as demonstrated by the enormous diversity within the LGBTQIA (lesbian, gay, bisexual, transgender, queer, intersex, asexual) community. Analyses of this kind are thus unscientific simplifications.[51]

Another aspect of pseudoscientific theories is the lack of feedback to determine whether a prediction was correct. We can never know for sure whether a job candidate who was rejected based on a personality test might have been suitable for the position after all. Someone whose asylum application was turned down because they lied often has no opportunity to prove their innocence. Technological applications of this kind do not just investigate a certain area, then, they also generate their own reality, often without being tested.

---

[49] O'Neil 2016.

[50] Claus, 12 September 2017.

[51] A very similar phenomenon is the UK immigration service's use of DNA tests to determine nationality – for example, to distinguish between Kenyans and Somalis. Genes do not respect national boundaries, however, and so the idea that nationality can be established biologically is incorrect (Benjamin 2019).

Key Points – The 'AI Is More Rational Than the Human Mind' Perception

- More data, greater computing power and the ability to identify complex relationships suggest that AI is more rational than the human mind.
- In reality, however, correlation is often confused with causality.
- Some of the ways in which AI's abilities are portrayed are designed to serve commercial purposes.
- Anthropomorphizing AI creates the impression that it has greater intellectual capacity than is in fact the case.
- Emotion detection, online personality testing, fitness trackers, sexual orientation analyses and certain approaches to poverty are in fact based on pseudoscientific theories and applications, and thus pose major societal risks.

5.2.1.3 Artificial Intelligence Is a Black Box

One commonly heard view of AI is that it is a 'black box'. This term, popularized by early cybernetics experts, refers to systems we cannot properly fathom and understand. How exactly black boxes translate input into output remains a mystery, since we have no grasp of their inner workings.[52] As a result, AI is seen as being opaque, undefinable and almost impossible to regulate. This is particularly problematic in domains where transparency is important, such as a court's reasons for imposing a particular sentence on defendant. The black-box problem also features in other cases where legitimacy, legal certainty and legal equality are crucial – a concern reflected in demands from the Dutch civil courts and the Council of State for greater transparency in the use of algorithms.[53] While it is often argued that control and transparency are unnecessary in less 'vital' situations, neglecting them can still prove problematic in the long run.

The notion that AI is a black box has even prompted Frank Pasquale to express concerns about the rise of a 'black-box society', where unfathomable systems make a whole range of decisions in such areas such as reputation management, online searches and the financial sector. Pasquale's use of this term has another dimension, too; as well as the incomprehensibility of the systems involved, he is worried about the 'black box' as a universal recording device, analogous with its namesake found in aircraft.[54]

Is the idea that AI is a black box just another myth? Not necessarily. But it is important to be more precise about what we actually mean here. The term 'black box' tends to be used in very different ways, with some of these variants presenting

[52] Rid 2016: 66–67.

[53] Wolswinkel 2019: 776–785.

[54] Pasquale 2015.

greater obstacles to be overcome than others. In order to formulate appropriate responses, therefore, we first need to distinguish clearly between the various definitions.

First, the concept of a black box may be used to indicate something so complex that certain people are unable to understand it. But that does not exclude the possibility that others do. In this sense, many aspects of modern society are black boxes for most people. When they step into a lift, they rely on it to operate properly without knowing exactly how it works. The same applies to other technologies and to many legal, political and administrative issues as well. However, this does not mean that these things are entirely beyond human comprehension. Some groups of people are skilled in the relevant areas and bear responsibility for them.

In the world of AI, the decision trees used by expert systems are analogous to this type of black box. People who have not studied that technique find it difficult to understand, but it can be readily explained by those who have. This form of black box poses few problems: we need only to ensure that there are enough people who understand and can explain the system, just as there must always be enough mechanics available to repair faulty lifts.

A second type concerns situations in which we do not have access to the data and analyses used to generate certain outcomes. This could be due to a variety of factors. One possibility is that the data in question simply has not been maintained or stored. Another is that we lack the rights needed to view that information, as when we use the services of a company that considers its data and the workings of its algorithm to be trade secrets. Likewise, a government agency might not wish to make an algorithm public as this would undermine its purpose (combating fraud, for instance).

This variant is considered as involving a black box since a particular interested party is given no opportunity to understand the system. In many cases this is for commercial or legal reasons – due to a confidentiality clause in a contract, say, or because of barriers imposed by intellectual property rights. Here too, the obstacles are not insurmountable.

Looking at the US, for example, Pasquale argues that we should critically review all the various legislation that has made it easier to classify things as trade secrets, especially since their effects now permeate society.[55] The American AI Now Institute, too, urges that we not accept that the workings of systems key to the functioning of society constitute a proprietary secret.[56] As well as rules relating to confidentiality, the institute is here also referring contractual provisions that can be refused. A trickier permutation of this variant is when the data on which an outcome is based is derived from a range of very different sources.[57] One example is algorithms based on other algorithms, whose origin cannot be traced. This problem

[55] Ibid.

[56] Crawford et al. 2019.

[57] WRR 2011.

arises in chain decisions. Studies carried out in the Netherlands show how many government decisions are made by linking various systems.[58] Although the resulting outcome is not unfathomable in theory, in practice it is virtually impossible to trace how the final decision came about.

A third use of the term 'black box' is more technical in nature is closely related to the recent rise of deep learning in AI. These are systems so advanced in terms of their complexity that the outcomes would be too difficult for people to understand. As an example, consider the process whereby a particular article is placed on someone's Facebook timeline. This is an immensely complex, real-time operation involving millions of users at once, in which each person's data interacts with that of other people. This issue is a cause of concern because, for example, it raises worries that elections might be influenced. The danger is that, given the level of complexity involved, it may no longer be possible to find out why a given message did or did not appear on someone's timeline.[59] It is important to note that this does not necessarily mean that the process is fundamentally incomprehensible, but simply that it makes the task of finding out how a system arrives at a particular decision a very complex one for potential investigators.

Sometimes, though, a system's processes are indeed too complex for a human to check. This is because, on occasions, the logic used by AI systems differs from that in our brains. Take pixel-level image recognition, for instance. We can certainly understand how just a part of a photograph is enough to recognize a face. However, deep learning identifies patterns at various deeper layers involving input at the level of individual pixels. Humans are unable to follow logical reasoning at that level. At Facebook, two computer programs are said to have developed a 'language' that enables them to communicate with one another in a way people are unable to understand. In all such cases, though, the question is whether the issue really is fundamentally incomprehensible or whether, given enough time, we would be able to understand the process concerned.

We do not intend here to explore the details of specific remedies for the various types of black box. Our aim is merely to show that this term covers a range of very different phenomena, which leads to confusion. Some of those phenomena present greater obstacles than others. Moreover, it is not just the nature of algorithms that can make black boxes unfathomable; property rights and complex social systems play their part as well. As for how to tackle this issue, the answer will be different for each of the four types we have described. But it is not impossible, so whenever the term 'black box' is used to indicate that something is beyond our understanding and that, as a result, we should not use it or cannot control it effectively, it is time to pause and dispel the myth.

[58] Van Eck et al. 2018.

[59] Greenfield 2017: 252–253.

Key Points – The 'AI Is a Black Box' Perception

– The image of AI as a 'black box' can give rise to the notion that control or transparency is impossible.
– However, the term 'black box' is used in very different ways. It may indicate complexity (which is not beyond the understanding of experts), a lack of access to a system's inner workings (due to legal or other restrictions), the performance of huge numbers of calculations or something that is fundamentally incomprehensible.
– The term 'black box' can refer to many different things, so it is important to have a clear understanding of what we mean by it in any given situation.

Many misunderstandings about how AI operates can be prevented by asking critical questions during the various steps involved in its application. The box on the next page provides some suggestions.

Questions to Ask About AI in Practice

Goal & planning

- What selected slice of reality is being produced here?
- Has current practice been properly analysed?
- Which goal does the system optimise?
- Does the application domain serve multiple purposes?
- How has the system been influenced by its creators' world view?
- Can anyone explain how the algorithm works?
- Are the databases and models used to train the algorithm accessible?
- Are there any legal barriers to examination of the way an algorithm operates?

Data collection & training

- How good is the training data?
- Is the phenomenon measured directly or are proxies used?
- Is the subject of the measurement really an objective phenomenon?
- Is the model underpinned by any pseudoscientific theories?

System design

- What definition of equitability is used?
- Are any patterns involved that are incomprehensible to human brains?
- Does the system use multiple data sources that might make it more difficult to understand?

Output & effects

- Is a distinction drawn between correlation and causation?
- Do any words used suggest a false analogy with human intelligence?
- Does the algorithm influence the factors it measures?

5.2.2 Myths About the Impact of AI

5.2.2.1 Artificial Intelligence Will Soon Equal Humans

We have already discussed three myths about how AI operates – that the technology is neutral, rational and a black box. Next, we examine various myths concerning its expected implications in the near future. The story about Sophia the robot (see Box 5.1) is typical of these in that it implies that AI will soon rival humans and then far surpass us.[60] As we saw in Part I, people have been speculating about this type of artificial general intelligence (AGI) ever since the field first emerged.

The writer Vernor Vinge coined the term 'singularity' for the moment when smart machines would start relating to us the way we relate to animals. A moment he believed would come. Similarly, the mathematician I. J. Good described a future 'intelligence explosion'.[61] Potential scenarios of this kind have come to the fore again in recent years. The futurist Ray Kurzweil, who works for Google, expects AGI to arrive in 2029 and the singularity to occur around 2045. DeepMind and various other companies (such as Vicarious, Kindred and Numenta) have issued mission statements expressly declaring that their goal is to create AGI.[62]

The expectation that AI will soon be able to equal human capabilities has been fuelled by recent advances and by suggestions that various current breakthroughs are paving the way for AGI. In the field of autonomous mobility, Otto (a division of Uber) has succeeded in developing a vehicle able to drive itself from the east coast to the west coast of the US. Also referring to autonomous vehicles, President Obama noted in a 2016 interview that "the technology is essentially here".[63]

Box 5.1: Robot Citizens
In 2016 Sophia the robot, developed by Hanson Robotics, was exhibited at the famous South by Southwest technology festival. A year later she appeared at the Future Investment Summit in Riyadh, where she was granted Saudi Arabian citizenship. She replied in person, saying, "I'm the first robot to be granted citizenship, it's history in the making". The move immediately triggered wide-ranging discussions. Did it mean that Sophia had the right to marry or to vote? And would switching her off now infringe her rights as a citizen?

[60] One of the issues arising out of this is AI's impact on employment and people's fear of mass layoffs. With previous system technologies, this fear proved a complete myth. Nonetheless, it is a key issue in the context of AI and so is addressed in the next chapter, where we examine the embedding of AI in the macroeconomic context.

[61] Vinge 1993.

[62] Agrawal et al. 2018: 223.

[63] Broussard 2019: 142–147.

As we saw at the beginning of this chapter, the idea that fundamental breakthroughs are now taking place is also being stoked by publicity-generating competitions. The rhetoric used on those occasions tends to fan the flames of unjustified extrapolations. Every event is portrayed as yet another step towards the day when computers finally acquire the full range of human intellectual skills. Melanie Mitchell refers to this as one of the pitfalls we tend to fall into when thinking about AI: that 'narrow intelligence' and 'general intelligence' are two points on the same continuum.[64] While IBM's Watson did indeed win *Jeopardy!*, that does not make the program a good doctor. In 2017 the MD Anderson Cancer Center at the University of Texas terminated its partnership with the Watson project on the grounds that some of system's recommendations were "unsafe and incorrect".[65]

The history of system technologies teaches us to be cautious concerning the expectations engendered by competitions and demonstrations. They attract attention and appeal to people's imagination, but their primary purpose is to promote the technology – and so, in many cases, the controlled conditions under which they take place are glossed over. For example, Otto's impressive road trip took place in a heavily managed environment. Since that first demonstration drive in 2016, several fatal accidents involving autonomous vehicles have occurred under more mundane and considerably less controlled conditions. As time goes on, an increasing number of major obstacles to autonomous vehicles are emerging. People can easily rotate and displace objects in their minds, for example, but they are difficult tasks for algorithms to emulate. So, a failure to recognize an orange traffic cone that has toppled over could lead to hazardous situations. In the next chapter we explore the current situation with regard to autonomous vehicles in greater depth.

Experts have been claiming since 2012 that autonomous vehicles will be here 'within a few years', but that timeline is constantly being pushed further and further into the future. This helps put expectations around AI into perspective. Marcus and Davis point out that we were hoping to get Rosie, the robot servant from the cartoon series *The Jetsons*, but instead we got Roomba, the autonomous vacuum cleaner.[66] Even the technology entrepreneur Peter Thiel has remarked, "We wanted flying cars. Instead we got 140 characters", referring to the maximum length of a tweet.

So-called 'hackathons' are a very popular type of competition designed to drive innovation. From their beginnings in Silicon Valley, they have now spread all over the world. However, those familiar with the field say that the flashy publicity associated with these events should be taken with a grain of salt. Any developments to come out of them are too short-lived to deliver genuine progress towards viable products. For this reason, the products of hackathons are often jokingly referred to as 'vapourware' – great promises of innovations that will never appear.[67]

---

[64] Mitchell 2021.

[65] Marcus & Davis 2019: 5.

[66] Marcus & Davis 2019: 98.

[67] Broussard 2019.

While many recent breakthroughs are more relevant than this, they still need to be placed in the right context. With regard to *Jeopardy!*, for example, nearly 95% of its answers are the titles of Wikipedia pages.[68] Watson's win demonstrated its ability to navigate through that material, not a mastery of the complexities of human language. According to the philosopher Daniel Dennett, moreover, the rules of the game were tightened up somewhat to enable Watson to take part.[69] As we have already noted, the defeat of a chess grandmaster was the result of a linear progression that can be traced back to the 1960s.[70] The game go requires enormous computing power, so AlphaGo's win over Lee Sedol was impressive. At the same time, the algorithm's achievement required use of a combination of methods plus the input of knowledge gleaned from a large number of human experts.[71] Although far more complex, the basic challenge in Go is comparable with the game noughts and crosses (tic-tac-toe) in that it involves filling a two-dimensional grid and the optimum outcome can be expressed as a function.[72] The victorious AlphaGo program has very few applications outside the context of these games.

If we are to demystify AI, then we must tackle unrealistic expectations. Although important steps are being taken, the technology is still not close to equalling humans, to achieving AGI or to overshadowing us. On the other hand, some people tend to overly downplay the chances of achieving superhuman intelligence.

According to Andrew Ng, such concerns are "like worrying about overpopulation on Mars".[73] Stuart Russell questions that argument, however, and rightly so. While we are not yet in the process of colonizing Mars, substantial investments are already being made in the development of AGI.[74] After all, this is the goal of the AI field. Russell feels that it is odd for people who are busily developing a train that is destined to plunge off a cliff to insist that there is no need to worry because we will have run out of fuel long before we reach the cliff-edge.

We therefore need to take the goal of equalling human intellectual abilities very seriously indeed. At the same time, we must put any announcements of breakthroughs into context. Given the current state of progress, after all, that goal is still far beyond our reach. Russell presents a very useful classification system for a range of variables we can use to assess AI applications. The nature of the environment may be entirely clear (like a chessboard) or much less so (like road traffic), actions can be discrete or continuous, other actors may or may not be involved, the outcomes of actions may or may not be predictable, the environment may or may not change dynamically and the horizon against which the achievement of goals is

[68] Broussard 2019: 82.

[69] Dennett 2019. 49.

[70] See Chap. 3.

[71] Notwithstanding AlphaZero, which taught itself how to play.

[72] Broussard 2019: 33–4.

[73] Extract from a 2018 interview with Andrew Ng: 202.

[74] Russell 2019: 151–152.

measured can be near or distant.[75] These variables give rise to a huge set of assorted issues. While great progress is being made in areas that are completely manageable, discreet and predictable, for example, the resolution of other points remains a very distant prospect.

Many of the questions we have raised concerning predictions that AI will equal humans within a relatively short space of time are covered by the three elements of what Marcus and Davis describe as the 'AI gap' between expectation and reality. The first of these is our own credulity. We attribute human qualities to machines. While people require intelligence to perform certain tasks, though, that is not necessarily the case for machines. The second element concerns imaginary progress. Advances in solving simple problems (as in *Jeopardy!*) should not be confused with an improved ability to solve complex ones (such as understanding human language). Finally, say Marcus and Davis, there is a robustness gap. Compared with solutions already achieved or within our grasp, such as hand-free motorway driving, more complex tasks like autonomous inner-city driving involve an inordinately greater degree of difficulty. In metaphorical terms, you can climb taller and taller trees but that will never get you to the moon.[76] For that you must develop alternative methods. The idea of machines equalling humans should certainly not be dismissed out of hand, but it is still far beyond the reach of current methods. We will need to make further fundamental breakthroughs if we are to move any closer to that goal. As discussed in Chap. 3, today's artificial intelligence is all 'narrow AI' – that is, systems focusing on specific tasks. They already surpass humans in a number of these, and for the time being we are much more likely to create more systems that outdo us in other narrow domains than we are to achieve AGI.

Key Points – The 'AI Will Soon Equal Humans' Perception

- Recent developments and breakthroughs suggest that we are close to equalling human capabilities, what we call artificial general intelligence (AGI).
- However, high-profile competitions and demonstrations largely gloss over the controlled conditions required for AI to be successful.
- There is an 'AI gap' between expectation and reality. This is driven by projecting the way human intelligence operates onto machines, by the imaginary progress associated with the misrepresentation of milestones and by unjustifiably extrapolating from simple issues to complex ones.

[75] Russell 2019: 44.

[76] Marcus & Davis 2019: 18–22, 66.

5.2.2.2 Malign Artificial Intelligence Could Turn Against Humans

This is perhaps society's greatest fear when it comes to AI, one further inflamed by imagery in popular culture. As we saw in the previous chapter, the term 'robot' was first used in a play about mechanical workers turning against humanity. Over the years, the same theme has featured in numerous movies.

These stories are based on a motif from the distant past, long before AI or computers were invented. In the first chapter we saw that myths about artificial forms of life date back to ancient Greece, and perhaps even earlier. Many of these include a dystopian element. The creation of artificial life has generally been viewed historically as a transgression of boundaries that warrants some form of punishment. The tales of Prometheus, Daedalus and Medea are much in the same vein. A more modern story in that same tradition is *Frankenstein* (subtitled 'The Modern Prometheus') by Mary Shelley. Dr. Frankenstein creates an artificial life form that eventually kills its creator. The modern fear of malign AI is just the latest chapter in a long tradition of disquieting imagery.

Another phenomenon that helps stoke this fear is known as the 'uncanny valley'. This centres on our relationship with machines that display human characteristics or behaviour. We tend to feel sympathetic towards machines in human form, but that turns into fear and repugnance if they resemble us too closely. The advent of machines indistinguishable from humans, however, will make the 'uncanny valley' a thing of the past. This is yet another phenomenon that inflames fears of malign AI.

Researchers like Nick Bostrom and Max Tegmark have devoted several thought experiments to scenarios of this kind,[77] although they are keen to emphasise that their work is purely speculative. In movies, malign AI often assumes humanoid form as a robot or a talking computer. While that is certainly possible for 'real' AI as well, physical incarnations of this kind are not essential to its further development. Extremely powerful AI is more likely to take the form of intangible algorithms than actual machines.

Besides its form, the myth of malign AI also imbues the technology with other human qualities it cannot rationally be expected to develop, such as a lust for power, a desire for freedom, jealousy and a fear of death.

According to Steven Pinker, the scenario that robots will become superintelligent and enslave humans "makes about as much sense as the worry that since jet planes have surpassed the flying ability of eagles, someday they will swoop out of the sky and seize our cattle. The … fallacy is a confusion of intelligence with motivation – of beliefs with desires, inferences with goals, thinking with wanting. Even if we did invent superhumanly intelligent robots, why would they want to enslave their masters or take over the world? Intelligence is the ability to deploy novel means to attain a goal. But the goals are extraneous to the intelligence: being smart is not the same as wanting something."[78]

[77] Bostrom 2016; Tegmark 2017.

[78] Marcus & Davis 2019: 30.

Yann LeCun points out that a desire to take over the world correlates not with intelligence but with testosterone.[79] A related objection is that malign AI scenarios assume that we have reached the level of AGI, whereas in fact – as noted above – there is currently no prospect of that.

Given the compelling nature of this myth, one key objection to it is that focusing on something so entirely speculative tends to distract us from more serious threats that are very real. For instance, the risks posed to human life by machines have nothing to do with intentions, malign or otherwise. A missile flying towards its target has no ill will at all, but it will kill people nonetheless. The problem, then, is not so much that AI may develop malign goals of its own but that it is very adept at achieving the goals people have built into it – which may be dangerous or ill-conceived.

This brings us to the issue of 'value alignment', which means designing AI with goals that coincide with our own – a concern prompted by the fact that an AI's rigorous pursuit of certain goals can jeopardise others. Russell describes this as the 'King Midas' problem, after the legend of the monarch who was granted his wish that everything he touched turn to gold. This enabled him to achieve his aim of becoming enormously wealthy, but when his food and his relatives also turned to gold, he discovered that this goal conflicted with others.[80]

An increasingly efficient AI that becomes destructive in the pursuit of certain preprogrammed goals thus poses more of a risk than malign AI. As Norbert Wiener put it, "…human impotence has shielded us from the full destructive impact of human folly".[81] Now that we are able to make machines that can achieve goals by advanced means, we are confronted with our more ill-conceived aim. A well-known illustration of this problem is Nick Bostrom's thought experiment about a paperclip machine. He proposes the idea of a highly intelligent machine whose goal is to manufacture as many paper clips as possible. To achieve that, it may first decide to wipe out humanity to ensure that it can transform any matter it finds into paper clips quietly and without resistance.[82] Linear AI with no goals of its own is a greater danger than AI with nefarious plans.

Russell provides a compelling example of the destructive effects of a simple algorithm designed to select content on social media. Its purpose is to maximize advertising revenue by increasing the number of click-throughs. If the algorithm starts by selecting the content people find most interesting, that seems relatively harmless. However, this algorithm achieved its goal in a different way: it changed people's preferences in a way that made their behaviour became more predictable. People with more extreme political views tend to have more predictable preferences, so the algorithm prompted users to become interested in more extreme content. Given the prevailing hostile political climate on social media platforms, this is

[79] From an interview with Yann LeCun (Ford 2018: 135).

[80] Russell 2019: 137.

[81] Wiener 1964.

[82] Bostrom 2016.

a significant factor. Yet there is no malicious intent here; these actions are entirely in keeping with the pursuit of the original goal – maximizing advertising revenue.[83]

Similarly, when it comes to autonomous vehicles people tend to focus on speculative scenarios rather than acute issues. Much of the debate in this area centres around the so-called 'trolley problem' (referring to trolleybuses): what course of action should autonomous vehicles take when an accident is unavoidable and they have to decide who lives and who dies? The 'appropriate' values in this case are the subject of a great deal of speculation. Are these universal, and would people buy cars that might sacrifice the driver to save the lives of others? While this could be the topic of many interesting philosophical debates, there are other more acute challenges associated with autonomous vehicles. Furthermore, simpler forms of driver assistance are already commercially available. These have been implicated in cases of injury and death, so it would be better for us to focus on them.[84]

Reports about AI and the frames used in communications on this topic can create the impression among the public that the technology is developing along harmful lines. That is a myth. Facebook's programs were not plotting the overthrow of humanity, nor do autonomous weapons want to take over the world. Their actions are life-threatening, to be sure, but technically they are no different from chess computers in the sense that they calculate and execute moves with the goal of winning the game.

Key Points – The 'Malign AI Could Turn Against Humans' Perception

- Triggered in part by popular media, there is now a widespread public fear of malign AI. This has deep historical roots. It is being further inflamed by the use of specific terminology like 'killer robots'.
- Disquieting imagery of this kind projects human characteristics and intentions onto AI, even though there is little reason to do so.
- Malign AI also presupposes the existence of AGI, which is still only a very distant prospect.
- Yet even without malicious intent, AI can still be dangerous by pursuing flawed goals or by achieving certain aims at the expense of others.

The view that AI is developing along harmful lines is just a myth. But this does not mean that we should downplay such perceptions. The history of system technologies teaches us that words, associations and disquieting imagery have often been highly influential. On occasions, they have even turned the public against certain technologies. So, demystification is vital if we as a society are ultimately to reap the benefits of a new system technology such as AI.

[83] Russell 2019: 8–9.

[84] Onderzoeksraad voor Veiligheid 2019.

5.2.3 Generic Myths About Digital Technology

5.2.3.1 Technology Should Be Regulated as Little as Possible

The five perceptions described so far are specific to AI. Three are about how it operates and two about its future impact. But as a major new technological development, AI is also part of a wider environment. Leading platforms involved with previous digital technologies are now at the forefront of this one as well. Because of this interdependence, it is important to examine broader perceptions of technologies – and digital technologies in particular – with their origins in Silicon Valley. Demystifying them will help us to gain a better understanding of AI.

One of the first powerful perceptions of technology to arise in Silicon Valley was that it should be subject to as little regulation as possible. This view can be substantiated in various ways. It may follow from a techno-deterministic approach, for instance: the notion that technology operates autonomously and that the world simply has to adapt to it. Any society that fails to do so, that insists on curbing technology, will be left behind. The motto of the 1933 Chicago World's Fair was 'Science Finds, Industry Applies, Man Adapts'.[85] Few people would put it quite so forcefully these days, but many still embrace milder variants of techno-determinism.

We also see an instrumental approach to technology:[86] while it does not actually shape society, it is a tool whose uses will be decided by people themselves. A hammer, for example, can be used to build a house or to kill someone – that is up to the user.

There is a grain of truth in both of these approaches. The following quote is usually attributed to Marshall McLuhan, a renowned philosopher of technology: "First we shape our tools and thereafter our tools shape us."[87] What McLuhan suggested in his work is that society and technology are inseparable. That they are deeply intertwined. The history of system technologies also teaches us that embedding a new technology in society requires a process of mutual adaptation, part of which involves setting standards and drawing up regulations.

Although they differ radically from one another, both the techno-deterministic and the instrumental approaches lead to the same conclusion: that technology should be subject to as little regulation as possible. In the former this is because regulation is futile, while the latter argues that we should focus on use rather than the technology as such. History also teaches us that each new system technology arouses ideologically motivated appeals that it be left to its own devices as far as possible, and that this approach always requires correction later.[88] When it comes

[85] Zuboff 2019: 15.

[86] For a discussion of the technodeterministic and instrumental approaches, see: WRR 2011.

[87] This, incidentally, is not a literal quotation from McLuhan; see the foreword by Lewis Lapham in McLuhan 1994 [1964].

[88] This reflects the tenor of the debate about the internet in its early days (Stikker 2019).

the technology of today, that correction now gradually seems to be taking place. In Chap. 8 we specifically address the overarching task of regulation. Here we first examine the origins of the myth that no rules are needed, then go on to explore how that frame is being applied with regard to today's technology to legitimize a specific agenda that could potentially jeopardize civic values.

Jonathan Taplin has very effectively documented the philosophy of Silicon Valley. He describes how Facebook CEO Mark Zuckerberg seized the opportunity presented by the Arab Spring of 2011 to put forward a techno-deterministic line of reasoning. Zuckerberg praised the way in which technology had helped ordinary people overthrow dictators. He contrasted this with fears about information being gathered and shared: "You can't isolate some things you like about the internet, and control other things you don't."[89] Google's original slogan was 'Don't be evil'. The purpose of framing tech companies as a force serving the interests of society is to ensure that they remain as free as possible from all forms of control.[90]

In addition to their desire to keep regulation to an absolute minimum, many Silicon Valley businesses oppose a variety of existing laws and standards. Not only did Uber launch its app in places where it was in clear breach of taxi regulation laws, it even developed a program called Greyball to determine the best way to evade enforcement checks.[91]

Such clashes with established rules and conventions are deeply rooted in the culture of Silicon Valley. This dogma expresses itself in positive terms such as 'disruption' and has much in common with the hacker movement. Zuckerberg's first letter to investors when his company went public was headed 'The Hacker Way'. In it he stated that hacking had an unfairly negative connotation. Disrupting the existing order was an official corporate goal. Facebook's internal motto until 2014 was: 'Move fast and break things.' In a 2009 interview, Zuckerberg stated that "Unless you're breaking stuff, you're not moving fast enough".[92]

Opposition to the existing order is expressed even more strongly in a book entitled *Zero to One: Notes on Startups, or How to Build the Future* by PayPal founder Peter Thiel. Here he proudly tells the tale of how four of the six people who started that business had built bombs in high school.[93] Peter Thiel is still a major investor in Silicon Valley. The people with whom he founded PayPal (also known as the 'PayPal mafia') went on to occupy important posts at a wide range of companies, including Tesla, YouTube, Facebook and Palantir (a software company that operates throughout the world, mainly in the security domain).

Taplin reveals that this mentality is rooted in libertarian beliefs that the size of government should be reduced to an absolute minimum, which can be traced back to the philosophy of Ayn Rand. She advocated the freest possible market, led by

[89] Taplin 2017: 221.

[90] Taplin 2017: 97.

[91] Broussard 2019: 74.

[92] *Business Insider*, 15 October 2010.

[93] Taplin 2017: 76.

pre-eminent entrepreneurs. For them, "The question isn't who is going to let me; it's who is going to stop me." There is no question that they should have to ask permission to innovate. Peter Thiel is known to be an adherent of Rand's philosophy.[94]

A distaste for government interference and regulation is evident in many Silicon Valley enterprises. This is characteristic of the 'cypherpunk' movement, whose goal is to render government interference impossible through technologies such as cryptography. Computer specialist Ryan Lackey moved to Sealand in 1999. This former wartime fort off the east coast of England has declared independence, although no established nation has recognised it.[95] Google founder Larry Page is known to have commissioned research into autonomous city states. A recent example of these efforts to move beyond the reach of governments is the Seasteading Institute, whose goal is to construct an artificial island without a government in international waters.

Another member of the cypherpunk movement, Timothy C. May, published the *Crypto Anarchist Manifesto* in 1988. In this he harked back to the old American frontier as a free and lawless territory until a single, apparently insignificant invention, barbed wire, enabled people to define boundaries and fence off private property. According to May, the internet is the new frontier. The 'minor' invention of cryptography would now be on the side of freedom, however, rendering online borders and possessions impossible.[96]

The same metaphor was used by internet pioneer Stuart Brand in a 1990 article entitled *Crime and Puzzlement: In Advance of the Law on the Electronic Frontier*. This specifically compared cyberspace with the Wild West of nineteenth-century America. Following in Brand's footsteps, author John Perry Barlow went on to found the Electronic Frontier Foundation and later published the *Declaration of Independence of Cyberspace*. In this he claims to be a representative of the future whose mission is to inform governments that they have no sovereignty in cyberspace.[97]

Online piracy is yet another area that reflects the libertarian aversion to rules and regulations. Kim Dotcom, the founder and owner of Megaupload (a major music piracy site until it was closed down), wrote a rap song in which he portrays himself as a defender of free speech and compares himself to Martin Luther King.[98]

These examples of Silicon Valley beliefs all reveal an uneven tug of war that favours a libertarian Wild West over private property, privacy and a strong state committed to such goals as the redistribution of wealth.

In a number of cases, this ideology impinges even further on key civic values. While that does not apply to Silicon Valley as a whole, of course, some influential individuals over there question democracy itself. Thiel, for instance, he has stated that he "no longer believes that freedom and democracy are compatible". His

[94] Taplin 2017: 227.

[95] Rid 2016: 287.

[96] May 1994.

[97] Rid 2016: 240–244.

[98] Taplin 2017: 174.

personal preference is clearly for the former. In a text for the website of the Cato Institute, a right-wing economic think tank, he writes, "Since 1920, the vast increase in welfare beneficiaries and the extension of the franchise to women – two constituencies that are notoriously tough for libertarians – have rendered the notion of 'capitalist democracy' into an oxymoron."[99] In another piece on the same site, he adds, "In our time, the great task for libertarians is to find an escape from politics in all its forms – from the totalitarian and fundamentalist catastrophes to the unthinking demos … We are in a deadly race between politics and technology."[100]

These are extreme standpoints, of course, and many in Silicon Valley do not share them. In fact it is home to various schools of thought on this topic, including those now convinced of the need for government intervention. The above views do come from an influential figure, though, and are still being widely propagated – albeit in a diluted form –by large technology firms. Samuel Freeman argues that recent libertarian thinking can no longer be described as 'liberal'; it seems instead to resemble a form of feudalism, which aims to replace a shared public space with individual bilateral contracts between companies and consumers.[101]

Many of the above standpoints concerning non-interference with technology find specific reflections in the context of AI. Here too, it is often argued that regulation is unnecessary, impossible or even harmful, and that it works to the detriment of society. The problematic nature of this issue is discussed in Chap. 8. Here it is important to realise that, like previous technologies, AI is associated with a specific ideology that rejects any form of regulation, and that can be at odds with democracy. History shows that this can lead to all sorts of hazards and accidents. Moreover, rules and standards are not at odds with the development of technology; indeed, they can facilitate its use. When developing an appropriate form of regulation, it is helpful to be aware of the sources, impact and hazards of any myths that refute its usefulness.

Key Points – The "Technology Should Be Regulated as Little as Possible" Perception

- The techno-deterministic and instrumental approaches to technology argue that it should be subject to as little regulation as possible.
- Its culture of disruption, hacking and libertarian beliefs often puts Silicon Valley at odds with the existing societal and political order.
- Silicon Valley even features certain schools of thought and developments that cannot easily be reconciled with democratic control.

[99] Taplin 2017: 70.

[100] Thiel, 13 April 2009.

[101] Freeman 2001.

5.2.3.2 There Is No Alternative (TINA)

Our second general perception concerning the nature of technology is closely related to the previous one. As well as militating against the regulation of technology, techno-determinism also argues that society has to adapt it. A kindred idea is the notion that the form and impact of today's technology are inherent to the technology itself, so there is no alternative. In other words, huge corporations, the mass collection of data, advertising as a source of income, markets as the source of all innovation and other aspects are all unavoidable, not as by-products of the technology but as integral part of it. So if we want to reap its benefits, we also have to accept every one of these.

Evgeny Morozov distinguishes the physical infrastructure from what he refers to as the 'myth of the internet'. The latter is a complex repository for a wide range of wishes and projections, which he says has very little to do with the hardware. 'The internet' (in quotes, referring to the mythical variant) has no clear meaning and can encompass virtually everything that happens online, from business modelling to the struggle for net neutrality and a wide range of internet-related technologies.[102] 'The internet' in this sense is a rhetorical construction, a myth, which renders clear understanding and critical views impossible.

Of course, even this perception that there is no alternative does not rule out variety of all kinds within the technological framework. While the business models used by Google and Facebook revolve around advertising, for example, that is not the case for a company like Apple. There are also substantial differences between social media platforms. Nevertheless, in this perception the fundamental organization of today's technology is immutable. Alternative models, such as not collecting data or allowing users (its source) to own it themselves, are considered unrealistic.

We are not concerned here about whether specific alternatives are realistic. However, we have critically examined the notion that the current incarnation of the technology is essential and the only possible option. In a separate report entitled *The public core of the internet: an international agenda for internet governance*, the WRR distinguishes the core components and deeper layers of the internet from the superstructure used by large technology companies.[103]

Various authors have in recent years questioned the presumptions that there is, of necessity, a link between technology and the free market and that private companies are the main sources of innovation. Mariana Mazzucato argues that much of today's innovation in fact originates in the public rather than the private sector; the latter is good at commercializing the results, but innovation itself is the product of a lengthy process of fundamental research that is too risky for market parties and too focused on the long term – and so requires public funding. Foundational work in renewable energy such the development of solar panels, for example, as well as a great deal of

[102] Morozov 2013: 17–18.

[103] WRR 2015. In the Netherlands, Marleen Stikker is a key driver of the debate concerning the current form of the internet and possible alternatives (Stikker 2019).

innovation in biotechnology and nanotechnology are reliant on government support. Mazzucato also shows how many key components of the iPhone sprang from government-funded research. The same is true of the internet, touchscreens, GPS and even the voice assistant Siri, which was developed in the research laboratories at SRI International, an offshoot of Stanford University.[104] Mazzucato thus dispels the myth that only large technology companies can develop the wide-ranging innovation we see today, and goes on to ask whether it is right that the government – and, by extension, the public – should bear the risk associated with fundamental innovations while private companies appropriate all the profits.

Shoshana Zuboff shows that the structure of today's technology can be traced back to specific decisions in the past, which means that any number of alternative designs are possible. She describes a 2000 Georgia Tech project entitled 'Aware Home'. This involved early incarnations of today's 'smart home' technology, such as smart thermostats and virtual assistants, but adopted a completely different model. Not least, this involved the residents retaining full ownership of their data.[105] In her comprehensive study, Zuboff reveals how, over time, technology has become intertwined with – and shaped by – other developments. In particular, she explains how the neoliberal market economy first became involved. After the events of '9/11', governments began taking an interest in data collection and population surveillance. This required them to forge links with Silicon Valley companies that excel in these areas. Both the neoliberal market and data collection for surveillance purposes are external developments. As such, they are not inherent to the way in which technology itself operates. Zuboff's criticism focuses not so much on technology itself as on its owners and the choices they make – or as she puts it, on the "puppet masters, not the puppet" (Box 5.2).[106]

How can we translate this broad view of technology into a specific focus on AI? The government's historically key role in technological development is certainly reflected in the emergence of AI. In particular, the US military and its research arm, DARPA, played a critical part there.[107] The organization of AI in China, Japan and South Korea also shows that even today this technology can be directed much more firmly by government. Whether this is desirable is another matter entirely, but what is relevant here is that its linkage exclusively to large private companies is not the only possible model.

The history of AI also shows that, besides its public orientation, the technology was also once associated with a different model. One of its creators was Douglas Engelbart, an inventor who was decades ahead of his time in proposing innovations like the mouse, windows, video conferencing and hypertext. In 1968 he gave a classic 100-minute presentation that has since come to be known as 'the mother of all demos'. One particularly important aspect of this was that Engelbart placed

[104] Mazzucato 2014.

[105] Zuboff 2019: 5–6.

[106] Zuboff 2019: 14.

[107] Weinberger 2019.

Box 5.2: Acceleration

Another kind of source that harks back to the historically more public nature of innovation is a 2013 publication by Alex Williams and Nick Srnicek, the *#ACCELERATE MANIFESTO for an Accelerationist Politics*. This offers a utopian vision of technology's ability to solve a wide range of problems, an outlook we examine critically in our review of the next perception.

What is relevant at this point is that the authors draw attention to the discrepancy between the great promise associated with today's technology and the way it is actually used – to create unnecessary gadgets and generate advertising revenue. They attribute this to the way technology has been become welded to neoliberal ideology. Williams and Srnicek argue in favour of drawing inspiration from earlier periods, such as the 1960s, when the goals to which technology was put, like sending human beings to the moon, incorporated wider societal interests.

In this respect they align seamlessly with Mazzucato, who suggests that as well as spotlighting the pace of innovation, we should also focus on the direction it is taking. She too cites 'man on the moon projects' as a model for the use of technology for public purposes.

computer technology in an entirely new context. The old mainframe was a piece of equipment used in large government offices and centralized organizations like IBM. Engelbart presented a vision of the personal computer as a device that could be used by individual citizens, something that would have a decentralizing effect.[108]

The person who filmed that legendary demo was Stewart Brand, founder of the magazine *Whole Earth Catalog* and an inspiration to the first generation of internet pioneers. The magazine played a pivotal role in transferring the idealism of the hippie movement to computer technology.[109] In this way, what had previously been part of a 'Cold War technocracy' became part of a desire for personal development, collaboration and community.[110] Cyberspace was no longer restricted to military projects and space travel, it had entered the San Francisco 'counterculture'.[111]

Jonathan Taplin argues that libertarians disregard the fact that the internet was initially conceived and funded by the government, after which it was adopted by v and academics who had no interest in profit. Impelled by idealism, early developers like Tim Berners-Lee (one of the founding fathers of the world wide web) wrote

[108] Rid 2016: 173.

[109] In a 1995 essay for *Time* magazine entitled 'We owe it all to the hippies' (Brand, 1 March 1995), Brand traced a path from counterculture to the personal computer.

[110] Turner 2006.

[111] Rid 2016: 166.

code free of charge.[112] He still regularly criticizes the form that the internet has now taken and is committed to supporting alternatives.[113]

Over time, digital technology has become increasingly linked to a more libertarian and technocratic market model. The contrast between the different visions is nicely illustrated by a conversation between Engelbart and renowned AI pioneer Marvin Minsky, who stated that he intended to make machines intelligent and conscious. To which Engelbart replied, "You're going to do all this for machines? What are you going to do for people?"[114]

As we have shown above, modern technology is not necessarily chained to its various modern-day incarnations. It has already existed in at least two other forms. So many of the design features of today's technology are non-essential elements. Various schemes have been devised to harness technologies like AI for other purposes and in other contexts, and to make different choices about its design. In Chap. 7 we show how activists of all kinds are working to make AI more diverse and more democratic.

Key Points – The 'There Is No Alternative' Perception

- The myth of 'the internet' equates technology with the overall structure of today's internet.
- From the technical point of view, numerous alternatives are possible. Various thinkers have shown how many of the factors that shaped today's internet are exogenous in nature. In other words, they are not part and parcel of that structure.
- Digital technology has taken on other forms in the past. During the 1960s it was shaped by the Cold War, while in later decades it was influenced by idealism.
- There are widespread calls for a redesign of the internet.

5.2.3.3 Technology Is the Solution to All Society's Problems

The final perception of AI we discuss here is the conviction that technology is a panacea for the great and difficult issues facing society. Whilst it clearly it can be (and indeed already is) a great help in resolving a lot of problems, however, all too often people place excessive faith in technology – and that can be problematic in itself.

Different authors have looked at this perception in different ways. Meredith Broussard uses the term 'technochauvinism', meaning a belief that technology can

[112] Taplin 2017: 54.

[113] Hern, 12 March 2019.

[114] Taplin 2017: 56.

solve any given problem.[115] Evgeny Morozov prefers the critical term 'technological solutionism' and begins his book on the subject with a quote from Google's former CEO, Eric Schmidt: ("In the future, people will spend less time making technology work … because it will function seamlessly. It will just be there. The Web will become everything, and it will also be nothing. It will be just like electricity …) If we get this right, I believe we can solve all the world's problems."[116] One of the snags with this approach, says Morozov, is that it assumes that all kinds of issues are in fact problems, when that may not actually be the case. Solutionism takes the view that numerous inefficiencies, ambiguities and obscurities detract from an ideal reality. Whereas obscurity is often essential for privacy, professional confidentiality or other matters we value, and lack of efficiency provides scope for the experimentation and practice crucial to for many human activities, such as cooking or learning a language. According to Morozov, the will to change things reformulates a diverse range of complex social situations into clearly defined problems with solutions that can be calculated or into transparent and evident processes that are easy to optimise. Other domains are then forced to model themselves on the way in which technology operates. So, Wikipedia become the model for politics, for example, Facebook the model for citizenship and Google the model for all innovation.[117]

Reformulating a wide range of societal domains as problems that can be solved by technical means is not without its hazards. For a start it puts their key functions at risk by narrowing them down to such a great extent. Work, for example, is not just a matter of output. So, while an algorithm optimized for output might improve efficiency, at the same time it could easily undermine job satisfaction, another important aspect of work. We have already mentioned the example of medical care, which is not just about healing people but often also a source of solace or even human contact. An algorithm could well improve the purely clinical aspect, but if it renders human contact superfluous it might cause the other functions of care to disappear entirely. This point is illustrated by the use of algorithms in the judicial system, where AI does indeed streamline certain proceedings – the processing of straightforward traffic fines, for instance. But even in such simple cases, this does not render human contact entirely irrelevant.[118] Then there are all the technologies designed to promote public safety and social harmony through surveillance and risk assessments. Although these may indeed reduce crime and improve behaviour, the social cost of continuous monitoring could include all manner of personal distress (Box 5.3).

Broussard stresses the danger of extrapolating success in one domain into others. Prominent technology pioneers are often treated as gurus entitled to express opinions about anything and everything. But just because someone has achieved a

[115] Broussard 2019: 7–8.

[116] Morozov 2013: 1.

[117] Morozov 2013: 5–6, 15.

[118] In the Dutch context, see various articles in *Algoritmes in de Rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak* ('Algorithms in the judicial system: the implications of artificial intelligence for the judiciary', Raad voor de Rechtspraak 2019).

Box 5.3: Covid Apps

In response to the Covid-19 pandemic, apps have been developed all around the world to chart the spread of the virus and facilitate contact tracing.[119] Their aim is to use monitoring so that targeted action – such as testing and quarantine – can be taken more quickly.

In April 2020 the Dutch government staged an 'appathon' as part of its development effort. This indicates that a form of solutionism had set in, with the authorities looking automatically to new technology to come up with answers even though it was far from certain that this would be the most productive approach. Alternative methods might well have better met the needs of the services responsible for tracking and tracing infections. New Zealand, for instance, adopted a low-tech policy: everyone in the country was simply asked to keep a diary of their contacts. In retrospect, the app eventually adopted in the Netherlands appears to have been of little help in fighting the pandemic. Also in April 2020, the WRR submitted a position paper to Dutch parliament cautioning against 'technosolutionism' in its response to the pandemic.

breakthrough in mathematics or created a profitable business model, say, this does not automatically make them an expertise in social issues or public policy. Indeed, an unswerving dedication to finding mathematical solutions, for example, can have a disastrous impact when this applied to interactions between people (or the capacity for such interactions).[120] A final risk of technochauvinism or technosolutionism is that in emphasizing revolutionary plans to build new things now in the future it overlooks the potential benefits of maintaining and improving what we already have.

Key Points – The 'Technology Is the Solution to All Society's Problems' Perception

- Terms like technochauvinism and technosolutionism reflect a belief that technology can solve all the thorny issues in society.
- Those who hold such beliefs tend to favour simplification or one-dimensional explanations, thus downplaying or disrupting other aspects of the social order.
- Simple quantitative approaches or an emphasis on the latest technology tend to distract attention from more effective, non-technological approaches that can sometimes work much better.

[119] Whitelaw et al. 2021.

[120] Broussard 2019: 80.

5.3 In Conclusion

Myths have always been part of the human story and they always will be. The same applies to artificial intelligence. So, it is impossible to permanently dispel the mythology surrounding AI. Moreover, there is no such thing as a 'realistic vision' of AI – reality is far too complex and uncertain for that. But this does not mean that we are powerless to counter the genesis of myths. Indeed, it is certainly possible to deconstruct unrealistic perceptions. As the first of our overarching tasks, then, demystification is primarily about helping to improve understanding of what AI is and what it can do. This makes crucial to our other tasks. Only with a better understanding of AI, after all, can we find appropriate ways of engaging with it – and even more importantly, remaining engaged. Unrealistic ideas about the technology will only foster a general aversion to it, which could cause us to miss out on genuine opportunities. On the other hand, a very limited understanding could result in excessive risks and unnecessary casualties. In short, there are plenty of reasons to demystify AI.

In this chapter we have explored people's perceptions of AI as well as the importance of its demystification. We have seen how its generic and novel nature gives our imaginations free rein. This can lead to unreasonable perceptions, especially since AI is occasionally associated with existing sources of distrust. We have seen, too, how impressive demonstrations can fuel overblown expectations and how the use of certain terms and frames can shape the way we think about AI. We have also discussed eight specific and very diverse perceptions. Three of these concern the way AI operates: that the technology is neutral, rational and a black box. Two involve future expectations: matching human intelligence and the danger of malign AI. Finally, the remaining three are broader perceptions of technology that often resurface in the context of AI: that it should be unregulated, that it can only take a single form and that it is the solution to all society's problems. We have found that while some myths are easy to dispel, that is not always the case – especially when it comes to perceptions rooted in the predominant ideology of Silicon Valley.

Finally, we have seen that a variety of actors are involved in this overarching task. To play its role, society needs to gain greater experience with the technology and become familiar with its use. Scientists, schools and the media have a particularly important function in this regard. Government, too, can help with the demystification process by investing in knowledge and in public campaigns, by setting up institutes or by supporting third parties capable of playing a key of their own. We discuss the challenge this overarching task poses for governments at length in Part III. For example, the need to serve multiple interest groups improve their knowledge of AI and familiarity with the technology by a variety of means. In the final chapter we fine-tune the discharge of this task by identifying current points of concern. We also put forward specific recommendations on how to start that process.

References

Ackerman, E. (2021, January 7). How Boston dynamics taught its robots to dance. *IEEE Spectrum*. Available at: https://spectrum.ieee.org/automaton/robotics/humanoids/how-boston-dynamics-taught-its-robots-to-dance

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Andersen, R. (2020, September). The Panopticon is already here. *The Atlantic*. Available at: www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/

Association for Advancing Automation. (2018, January 25). Why AI won't overtake the world, but is worth watching. *Industry Insights*. Available at: www.robotics.org/content-detail.cfm/Industrial-Robotics-Industry-Insights/Why-AI-Won-t-Overtake-the-World-but-Is-Worth-Watching/content_id/6979

Benjamin, R. (2019). *The race after technology: Abolitionist tools for the New Jim code*. Polity Press.

Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brand, S. (1995, March 1). We Owe it all to the Hippies. *Time Magazine*. Available at: http://content.time.com/time/subscriber/article/0,33009,982602,00.html

Brooks, R. (2018, January 1). My dated predictions. blog, *Rodneybrooks*. Available at: https://rodneybrooks.com/my-dated-predictions/

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Business Insider. (2010, October 15). Mark Zuckerberg, Moving Fast and Breaking Things', *Businessinsider.nl*. Available at: https://www.businessinsider.com/mark-zuckerberg-2010-10?international=true&r=US&IR=T

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI now 2017 report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf

Claus, S. (2017, September 12) Een algoritme dat aan je gezicht ziet of je homo of hetero bent. *Trouw*. Available at: https://www.trouw.nl/nieuws/een-algoritme-dat-aan-je-gezicht-ziet-of-je-homo-of-hetero-bent~b7b99615/

Crawford, K., Dobbe, T., Dryer, G., Fried, B., Green, E., Kaziunas, A., Kak, V., Mathur, E., McElroy, A., Sánchez, D., Raji, J., Rankin, R., Richardson, J., Schultz, S. W., & Whittaker, M. (2019). *AI Now 2019 Report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in Judicial decisions. *Proceedings of the National Academy of Sciences, 108*(17), 6889–6892.

Dennett, D. (2019). What can we do? In J. Brockman (red.), *Possible minds: Twenty-five ways of looking at AI* (pp. 41–53). Penguin.

Dignum, V. (n.d.). *There is no AI – Race and if there is, it's the wrong one to run*. Available at: https://allai.nl/there-is-no-ai-race/

Domingos, P. (2017). *The master algorithm: How the Quest for the ultimate learning machine will remake our world*. Penguin Random House.

Floridi, L., Taddeo, M., & Turilli, M. (2009). Turing's imitation game: Still an impossible challenge for all machines and some Judges–An evaluation of the 2008 Loebner Contest. *Minds and Machines, 19*(1), 145–150.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

Freeman, S. (2001). Illiberal Libertarians. *Philosophy & Public Affairs, 30*(2), 105–151.

Fridman, L. (2019, August 16). *Elon Musk: What's outside the simulation?* AI Podcast Clips. Available at: www.youtube.com/watch?v=YIVf3P3zq7g

GPT-3. (2020, September 8). A Robot wrote this entire article. Are You Scared Yet, Human? *The Guardian*. Available at: https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.

Halpern, S. (2019, April 26). The terrifying potential of the 5G network. *The New Yorker*. Available at: www.newyorker.com/news/annals-of-communications/the-terrifying-potential-of-the-5g-network

Harari, Y. N. (2019). Who will win the race for AI? *Foreign Policy Magazine*, Winter 2019. Available at: https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/

Hern, A. (2019, March 12). Tim Berners-Lee On 30 years of the world wide web: 'We can get the web we want'. *The Guardian*.. Available at: https://www.theguardian.com/technology/2019/mar/12/tim-berners-lee-on-30-years-of-the-web-if-we-dream-a-little-we-can-get-the-web-we-want

Hicks, M. (2018, October 12). 'Why Tech's gender problem is nothing new', *The Guardian.* Available at: https://www.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women

Ipsos. (2019). *Ipsos global poll for the world economic forum shows widespread concern about Artificial Intelligence*. Available at: www.ipsos.com/sites/default/files/ct/news/documents/2019-07/wef-ai-ipsos-press-release-jul-1-2019_0.pdf

Kasparov, G. (2018). *Deep thinking. Where machine intelligence ends and human creativity begins*. John Murray Press.

Kleinberg, J. (2018). Inherent trade-offs in algorithmic fairness. In *Abstracts of the 2018 ACM International Conference on Measurement and Modelling of Computer Systems (sigmetrics '18)* (pp. 40). ACM.

Macaulay, T. (2020, September 8). The Guardian's GPT-3-Generated article is everything wrong with AI media Hype. *The Next Web*. Available at: https://thenextweb.com/news/the-guardians-gpt-3-generated-article-is-everything-wrong-with-ai-media-hype

Marcus, G., & Davi, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. Vintage.

May, T. (1994). The Cyphernomicon: Cypherpunks FAQ and More, Version 0.666. Available at: https://hackmd.io/@jmsjsph/TheCyphernomicon

Mazzucato, M. (2014). *The entrepreneurial state: Debunking Public vs. Private Myths*. Anthem Press.

McLuhan, M. (1994 [1964]). *Understanding Media. The Extensions of Man.* MIT Press.

Mitchell, M. (2021). *Why AI is harder than we think*. Available at: https://arxiv.org/pdf/2104.12871.pdf

Morozov, E. (2013). *To save everything, click here: The Folly of technological solutionism*. Penguin.

NATO. (2019). *Performance Audit Reports International Board of Auditors for NATO (IBAN)*. Available at: www.nato.int/cps/en/natolive/topics_111783.htm

O'Neil, C. (2016). *Weapons of Math destruction: How Big Data increases inequality and threatens democracy*. Penguin.

OvV. (2019). *Wie Stuurt? Verkeersveiligheid En Automatisering In Het Wegverkeer*. Onderzoeksraad voor Veiligheid. Available at: https://www.onderzoeksraad.nl/nl/media/attachment/2019/11/28/wie_stuurt_verkeersveiligheid_en_automatisering_in_het_wegverkeer.pdf

Pall, M. (2019, June 8). *How the telecommunications industry 5G strategy will use artificial intelligence to replace human intelligence: The end of mankind as we know it*. Available at: www.salzburg.gv.at/gesundheit_/Documents/5G-AI%20(002).pdf

Pasquale, F. (2015). *Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

Pasquale, F. (2020). *New Laws of Robotics: Defending human expertise in the age of AI*. Harvard University Press.

Rechtstreeks. (2019). *Algoritmes in de rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak*, Rechtstreeks 2019, nr. 2, Den Haag: Sdu. Available at: https://www.rechtspraak.nl/SiteCollectionDocuments/rechtstreeks-2019-02.pdf

Reuters. (2020, April 24). False claim: Covid-19 stands For Certification Of Vaccination Identification By Artificial Intelligence. *Reuters*. Available at: https://www.reuters.com/article/uk-factcheck-covid-name-abbreviation/false-claim-covid-19-stands-for-certification-of-vaccination-identification-by-artificial-intelligence-idUSKCN2262AS?edition-redirect=in

Rid, T. (2016). *Rise of the machines: A cybernetic history*. WW Norton & Company.
Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.
Schothorst, Y., & Verhue, D. (2018). *Nederlanders over Artificiële Intelligentie. Onderzoek naar de kennis en houding van burgers en ondernemers over Artificiële Intelligentie*, Kantar Public. Available at: www.digitaleoverheid.nl/achtergrondartikelen/hoe-denken-nederlanders-over-artificiele-intelligentie/
Sheikh, H. (2021). Aanbevelingen. *ESB, 106*(4801), 407–410. Available at: https://esb.nu/esb/20066595/aanbevelingen-voor-een-geo-economische-wereld
Staatscourant. (2009, January 19). Convenant tussen de Sociale Inlichtingen- en Opsporingsdienst en de Stichting Inlichtingenbureau. *Staatscourant van het Koninkrijk der Nederlanden 2009–11*. Available at: https://zoek.officielebekendmakingen.nl/stcrt-2009-791.html
Stikker, M. (2019). *Het Internet is Stuk*. De Geus.
Taplin, J. (2017). *Move fast and break things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us*. Pan Macmillan.
Tegmark, M. (2017). *Life 3.0: Being Human in the age of Artificial Intelligence*. Penguin.
Thiel, P. (2009, April 13). The education of a Libertarian. *Cato Unbound.* Available at: https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian
Tilley, A. (2016, March 24). Alphabet's 'Moonshots' Head Astro Teller: Fear of AI and robots is wildly Overblown. *Forbes*. Available at: www.forbes.com/sites/aarontilley/2016/03/24/alphabets-moonshots-head-astro-teller-fear-of-ai-and-robots-is-wildly-overblown/?sh=7246137973bb
Trouw. (2016, October 17). RDW Vindt 'Autopilot' Van Tesla Misleidende Term. *Trouw*. Available at: www.trouw.nl/nieuws/rdw-vindt-autopilot-van-tesla-misleidende-term~b191a2e3/
Turner, F. (2006). *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital Utopianism*. University of Chicago Press.
van der Sloot, B., Keymolen, E., Noorman, M., Pechenizkiy, M., Weerts, H., Wagensveld, Y., Visser, B., & i.s.m. het College voor de Rechten van de Mens. (2021). *Non-Discriminatie By Design*. Tilburg University. Available at: www.tilburguniversity.edu/sites/default/files/download/01%20handreiking%20non-discriminatie%20by%20design%28NL%29.pdf
van Eck, M., Zouridis, S., & Bovens, M. (2018). Algoritmische Rechtstoepassing In De Democratische Rechtsstaat. *Nederlands Juristenblad, 40*, 3008–3017.
Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. In *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace* (NASA Conference Publication 10129) (pp. 11–22). NASA Lewis Research Center. Available at: https://www.researchgate.net/profile/Carol-Stoker-2/publication/234229828_Telepresence_in_the_human_exploration_of_Mars_Field_studies_in_analog_environments/links/554bb5600cf29752ee7e78f8/Telepresence-in-the-human-exploration-of-Mars-Field-studies-in-analog-environments.pdf#page=23
Weinberger, S. (2019). *The imagineers of war: The untold history of DARPA, The Pentagon agency that changed the world*. Vintage.
Whitelaw, S., Mamas, M., Topol, E., & Van Spallm, H. (2021). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health, 2*(8), e435–e440.
Wiener, N. (1964). *God and Golem, Inc.: A comment on certain points where cybernetics impinges on religion*. MIT Press.
Wolswinkel, J. (2019, October). Het Algoritme Van De Afdeling: De Realiteit Van Complex Bestuursrecht. *Ars Aequi*, 776–785. Available at: https://pure.uvt.nl/ws/portalfiles/portal/31738610/AA20190776.pdf
WRR. (2011). *iOverheid*. Amsterdam University Press.
WRR. (2015). *De Publieke Kern Van Het Internet. Naar Een Buitenlands Internetbeleid*. Amsterdam University Press.
Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford, Center for the Governance of AI, Future of Humanity Institute. Available at: https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Chapter 6
Contextualization

The term ‘contextualization’ concerns the uses to which AI is put. This overarching task is a key part of the transition from laboratory to society. It is also a complex and often underestimated process. In practice AI will have to operate in specific contexts, each with its own systems, practices, rules and logic. This process is known as contextualization and involves adaptation and integration, which are both very time-consuming. That delays the maturation of new system technologies. As a result, it often takes longer than we might expect for them to become part of our everyday lives. The overarching task of contextualization therefore raises the following question: ‘how will AI work?’

We tackle this question by discussing a variety of AI applications. Many of these are in healthcare, but we also use autonomous vehicles (an application for which many have high expectations) to illustrate the issues surrounding contextualization (see Box 6.1). Their development is not primarily dependent on mechanical aspects, such as the engine; it is more a matter of intelligent algorithms that make decisions about routes and can respond dynamically to the environment. We therefore use repeated references to autonomous vehicles in separate boxes to illustrate the central dimensions of contextualization.

Specifically in terms of AI, what does this task involve? The author Kai-Fu Lee draws a historical analogy with the contextualization process surrounding electricity. Thomas Edison’s discoveries led numerous entrepreneurs to disrupt all sorts of industries. They used electricity to cook food, light rooms and power industrial tools. Lee states that electrification – harnessing and applying electricity – required four inputs. These were fossil fuels to generate electricity, entrepreneurs to develop this technology commercially, engineers to apply it and a government to provide the underlying infrastructure. By analogy, he claims, AI requires raw material (in the form of data), talented entrepreneurs, AI scientists and a government policy that provides incentives.[1]

[1] Lee, 2018: 13–4.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_6

Box 6.1: What Is an Autonomous Vehicle?
Levels of autonomy

Autonomous vehicles cannot easily be defined, although the so-called SAE model is often used for this purpose. It classifies driving automation into six levels: zero to five. Vehicles with autonomy levels of zero to two still need human drivers, but provide them with some degree of software support. At level three and above a vehicle can drive itself in certain situations. So, the step up from level two to level three is in fact a giant leap. Eventually, vehicles will be able to operate fully automatically in all situations. That is level-five autonomy but is still a very distant prospect. Most modern cars have level-one autonomy, which involves advanced driving assistance systems (ADAS). These feature automatic lane keeping, parking assistance and cruise control. In 2018 the Dutch Ministry of Infrastructure and Water Management estimated that 1% of vehicles had level-two autonomy. This involves adaptive cruise control (ACC), which adjusts the vehicle's speed to match that of the vehicle in front. Vehicles at level three are not yet commercially available.

Contemporary applications

So-called 'truck platooning' has made reasonable progress in traffic automation. This is when a convoy of (potentially driverless) lorries – the 'platoon' – automatically follows closely behind a lead vehicle with a human driver. The benefits include improved traffic flows and fuel savings. Dutch research organization TNO anticipates that further development work will lead to fully autonomous trucks driving in a platoon.[2] Field experiments on public highways in Europe have been under way since 2016 as part of the European Truck Platooning Challenge. In technical terms, then, this technique is becoming increasingly possible. At present however, current regulations do not permit driverless vehicles to use the roads.

In addition, there are several examples of automatic robot taxis. These are only allowed to operate within a very limited area ('geofencing'), but within those limits now cover many kilometres and are collecting a great deal of data. One of the companies working on projects of this kind is Waymo, a subsidiary of Alphabet (Google), which operates in the vicinity of Phoenix, Arizona. For safety's sake, a human driver is always present as well – and that has turned out to be necessary. On one occasion a taxi stopped on the wrong side of the road and had difficulty entering an area where there was a lot of activity.

[2] TNO October 4, 2021b.

A Sociotechnical Approach

We classify the elements of contextualization rather more broadly than Kai-Fu Lee. This involves a sociotechnical ecosystem approach, divided into a technical ecosystem of supporting and emergent technologies on the one hand and a social ecosystem featuring the human context on the other. We can then consider this dual ecosystem at the macro level (employment and the productivity paradox) and at the micro level (behavioural context). Contextualization is the overarching task of embedding a new technology in these different ecosystems.

If we contrast a sociotechnical approach with other approaches to AI, its value becomes clear. A strictly defined approach to AI only addresses the specific algorithms that make up these systems, distinguishing AI from data-related issues, for example. While this can be justified based on theoretical considerations, the use of such a strict definition creates blind spots in terms of what makes the technology successful in practice. This requires supporting technology such as data, even if, strictly speaking, data is not part and parcel of AI itself. We need to consider the broader technical ecosystem to gain a complete picture of the overarching task of contextualization.

We can also contrast this with the instrumental approach to AI, which can often be found in ethical analyses. When viewed purely as a tool, AI is considered neutral. This is because people can use this technology for good or bad purposes. That limits potential responses to the establishment of principles or rules for good uses. This approach carries the risk that the context in which the technology operates may be ignored. An ecosystem approach actually spotlights the fact that entire environments are being transformed. Take the internet. An instrumental emphasis on good use focuses on ethical principles and formulates rules of etiquette for online behaviour, for example, but the internet has also transformed the public space and impacted interpersonal interactions. Any approach that focuses purely on the ethics of good practice would miss these more wide-ranging systematic changes.

Finally, yet another approach also touches on the issue of contextualization. This is research into 'AI readiness'. Oxford Insights publishes an annual index on this topic, which lists different countries' 'states of readiness' for AI. Whilst this touches on contextualization, however, it does not include non-technical dimensions (the social ecosystem). Furthermore, even within its technical dimensions it covers only a limited range of conditions.[3] This means that an excessively strict, instrumental or narrow technical approach would not address key factors that determine when AI will become workable. We therefore approach this question from the perspective of the sociotechnical ecosystem in which AI will operate (see Fig. 6.1). Below we start by discussing AI's technical ecosystem (Sect. 6.1), which, as indicated above,

[3] The index does cover the network, but does not specifically address data and hardware, for example.

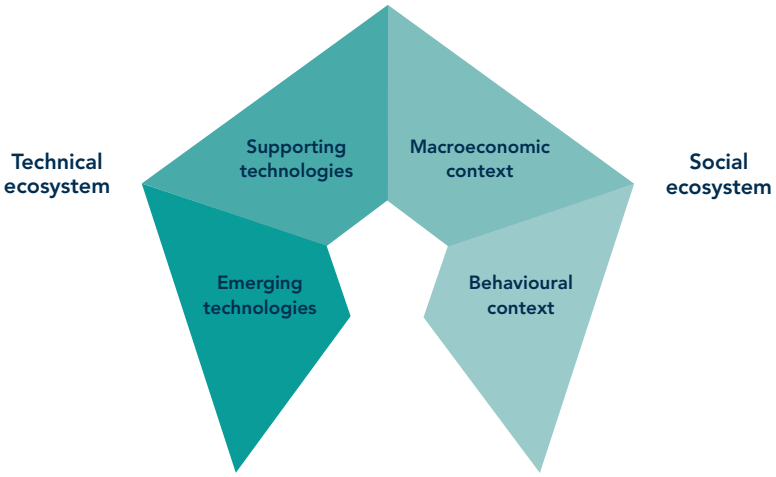


Fig. 6.1 AI's technical and social ecosystem

consists of two dimensions. This is followed by a discussion of the social ecosystem (Sect. 6.2).

6.1 The Technical Ecosystem

6.1.1 Supporting Technology

The first dimension of the technical ecosystem is supporting technology.[4] Strictly speaking, supporting technologies are not an integral part of a new system technology. Nevertheless, it cannot function without them. A related concept is 'enveloping', which emphasizes modifications to the environment rather than improvements to the new technology as a condition to make it function in practice.

In its strictest sense, AI involves the development of 'intelligent' algorithms, as mentioned in Part 1 of this report. What other technologies does AI rely upon? On the one hand there is the data it uses as raw material, on the other the hardware required.[5]

'Hardware' first of all refers to effective digital networks. This means the existence of a fast and reliable network. AI is based on complex calculations, which

[4] We are using a broad definition of 'technology' here. It also includes technically developed raw materials.

[5] Other supporting technologies are associated with manufacturing and energy supply, for example. A recent TNO working paper commissioned by the WRR (TNO, 2021a) explores the technical ecosystem in depth. Its authors break this down into core technologies, complementary technologies and supporting infrastructure (see Chap. 3 of the working paper).

often have to be performed at great speed. Autonomous vehicles in heavy traffic need to make decisions in milliseconds. The same applies to the machinery used in factories. Aside from the sheer speed involved, there must be no faltering or 'latency' (delay). In road traffic, even a brief failure of the network can have fatal consequences. Signal coverage varies from one area or town to another. It can be quite poor in some sparsely populated areas or in surroundings where the signals are blocked by massive structures or Faraday cages. Before AI can be implemented at a given site, then, it is important to determine whether the local digital network meets the necessary requirements. Prior to developing an AI application, therefore, we need to be sure that reliable networks are or will become available.

But hardware is not just about networks, it is also a matter of computing power. That involves chip technology and supercomputers. The computer chips developed by the semiconductor industry are key to this, as they are used in AI to perform the necessary calculations. This has traditionally involved central processing units (CPUs), an industry long been dominated by the American company Intel. The advent of smartphones fuelled the need for chips that use energy more efficiently. The US firm Qualcomm (which uses designs by the British company ARM) soon became a leader in this area. As we saw in the first chapter, it gradually became apparent that graphic processing units (GPUs) were the most effective means of performing many complex AI calculations. These chips, used mainly in the gaming industry, were developed by companies such as Nvidia.[6] Some of today's versions have been designed to perform specialized calculations, such as those used in machine learning algorithms. This technology is so specific and of such strategic importance to the industry that it is being developed by major technology platforms themselves. Google's TPUs and Microsoft's FPGAs are just two examples.[7]

It is important to note that Silicon Valley companies lead the development of this supporting technology. In its trade war with China, the US has denied that country access to critical chip technology, a move that has also affected other countries. In 2016 Qualcomm made a takeover bid for NXP (Philips' former semiconductor branch). However, this ultimately fell through due to China's opposition. For the Netherlands in particular, chips for AI are of strategic importance. Our nation, which is home to ASML, is a leading player in the global chip industry.

Supercomputers are another source of computing power. While these are not required for many everyday AI applications, they could become vital for very complex ones in the future. Japan, the US and China currently top the world rankings for the most powerful supercomputers.[8] In 2021 however, the top 10 also included two European supercomputers: JUWELS from Germany and HPC5 from Italy. Europe is investing in supercomputing capabilities by backing the European High Performance Computing Joint Undertaking (EuroHPC JU). VEGA, the first cofounded supercomputer, was presented in 2021.[9]

[6] Lee, 2018: 96.

[7] Ding, 2019: 23.

[8] TOP500, 16 November 2021.

[9] EuroHPC JU, 20 april 2021.

Besides hardware, the other major supporting technology for AI is the raw material it uses: data. Today's leading approaches to AI, such as deep learning, certainly require huge amounts of data – much more than classic rule-based systems. First and foremost, then, it is important this be available. However good its algorithms might be, AI cannot work without relevant data. So, it is important in AI applications to check that this can be obtained and where it is located. It is no coincidence that AI was first widely applied to consumer platforms. After all, these have access to enormous amounts of data from sources such as social media, search engines and online shopping behaviour.

The amount of data collected differs from one sector to another, and from one country to another. Take healthcare. In the Netherlands, various bottlenecks prevent this sector from making full use of AI. In many areas, for instance, the available data is limited or not entirely useable. Hospitals and institutions all have their own systems, which are not always mutually compatible. Moreover, some data exists only in handwritten form or in paper archives. This diversity stems from the decentralized nature of the Dutch system. In France, say, the sector is organized differently: it uses universal systems and centralized databases. That is one reason why this sector is a pillar of France's AI strategy.[10] With this in mind, the Dutch Council for Public Health and Society (RVS) emphasizes the importance of continuity of patient data when using AI in healthcare.[11]

Not only is the availability of sufficient data a key factor, but it must also be of sufficient quality, commensurable and accessible. The process of training algorithms is often hampered by the limited quantity of raw material (data) they have to draw on. This is due to factors such as commercial confidentiality, legislation, professional secrecy or just flawed systems. At Dutch university hospitals, for instance, AI scientists often have to train their algorithms using American medical data. This is either because the Dutch material is not accessible or because these scientists must first navigate their way through a complex and confusing application process. The equivalent procedures are much easier in the US.

This brings us to another point about the requisite data: it must be representative. There are no guarantees that algorithms trained on data from one site will produce good results anywhere else. This is clear from the above example of healthcare data. The populations of different countries may have different genetic traits and lifestyles, which make it impossible to draw general conclusions. That quickly became clear at the outbreak of the COVID-19 pandemic. Initially, most data were generated at a global level. But it was also necessary to collect data locally to allow for any country-specific differences in virus development. The same applies to mobility. Road signs, traffic regulations and urban planning differ significantly from one nation to another. We can train autonomous vehicles in one country but cannot simply assume that they will then operate effectively elsewhere. Therefore, it is important to supplement worldwide analyses with local data.

[10] France, 2018.

[11] RVS, 2019.

Challenges around the availability of good data do not stem just from technical issues, either. Others are related to the way in which a sector is organized, to legislation and standards or to the establishment of new systems for effective data management. Technical solutions to some of these challenges do exist, though. They include generative adversarial networks (GANs), which artificially generate new data when insufficient source material is available. If navigation service TomTom's cameras film a street while it is raining, GANs can filter out the rain when they generate data for that street. Another option is the technique of 'federated learning'. Here the algorithm is sent to the data instead of the other way around. This enables organizations to train their algorithms on sensitive data without having to acquire it or use it illegally (Box 6.2).

Box 6.2: Supporting Technology for Autonomous Vehicles

We can again apply the framework of supporting technologies to autonomous vehicles. Here we tend to focus mainly on the vehicle itself – or, more specifically, on the intelligence of the system that controls it. In technical terms, however, there is much more to effective autonomous vehicles than AI software alone. AI depends on a wide range of other technologies (involving both hardware and data) to operate effectively in this role. Autonomous vehicle developers are currently using a variety of technical approaches. In many cases, the following technologies are essential.

Sensors

We cannot collect relevant traffic data in real time without effective sensors. Autonomous vehicles need this hardware to scan their environment. Many cameras or other scanners have a limited forward field of view. Thus, given the response time involved, many autonomous systems only work at low speeds. Sensors also need to operate in a variety of weather conditions and environments. On one occasion they featured in a fatal crash involving a Tesla. Unable to distinguish a white truck crossing the highway from the background brightness of the sky, the car failed to apply its brakes. Given the potential risks involved, autonomous vehicle manufacturers have installed some critical elements in triplicate.

Digital maps

Digital maps are another important source of data for autonomous vehicles. These need to be accurate and up to date. For instance, they need to display any temporary roadworks that change the traffic situation. Companies such as TomTom and Waze are currently developing HD (high definition) maps accurate to the nearest centimetre, which is a major advance in data collection.

(continued)

Box 6.2 (continued)
Computing power

Autonomous vehicles need to perform complex calculations at great speed, so computing power is critical. In this respect Moore's Law (the doubling rate of transistors on chips) has important implications. We had to wait until around 2005 for cars with flash drives that could store 3D maps of cities.

I2C

The physical and digital infrastructures are additional supporting technologies. More specifically, this is all about infrastructure-to-car (I2C) communication. In general people are easily able to identify road signs and traffic lights. The same cannot be said of computers. This could be remedied by infrastructure that communicates directly with the car (by digital means). The vehicle would no longer need to interpret images ('Is that a traffic light or the brake light of a lorry?'). To that end we would need to incorporate chips into the physical infrastructure.

C2C

Another supporting technology is car-to-car communication (C2C). This would resolve the difficult problem of assessing other drivers' intentions. One particular benefit of automated communication between vehicles is that it could prevent rear-end collisions. Indeed, if cars were able to brake simultaneously, they could drive much closer together. The 'platooning' technique mentioned earlier is an example of this.

Network

Some types of autonomous vehicles use their onboard computing power, while others place greater reliance on network-based intelligence. When it comes to communication outside the car (I2C or C2C), an effective digital network is a must. Future 5G networks, in particular, may play an important part in this. We explore this in greater detail in the next box, on emergent technologies.

6.1.2 Technology in an Envelope

'Enveloping' is a key concept when it comes to supporting technologies. It enables a new technology to operate more effectively, not by improving the technology itself but by adapting the environment in which it operates. Here we present some examples that clearly illustrate the importance of enveloping AI because a failure to use this approach resulted in major problems. In her book *Artificial Unintelligence*, Meredith Broussard recalls how the introduction of e-education was hailed as a

solution to the limited availability of textbooks in many American neighbourhoods. It gave children online access to their learning materials. All that was needed was a one-off investment in a telephone or tablet. However, its backers forgot that e-education is just one element of an entire ecosystem, without which it cannot operate effectively. The cost of infrastructure is just one example. Computers can help children deal with problems, but they require maintenance and access to all kinds of other services, such as telephone lines and e-mail. Old school buildings are a case in point. Teachers need to confirm that their pupils can safely connect and charge all those computers. Which means that they need to have the building's electrical system assessed. The Wi-Fi network must operate effectively at all times and in all parts of the school. Those in charge need to create access codes, as well as secure learning environments that do not violate privacy rules. They also need to create a means of identification for each user, delete the accounts of former pupils and purchase licences for digital books. This shows that there is more to the digitalization of education than simply handing out computers.

Autonomous vehicles are yet another example of AI systems that are poorly suited to their surroundings. Broussard notes that even though arXiv and GitHub host massive online training sets for this application, that data does not include sufficient 'edge cases' (unusual situations). In addition, there can never be enough data to cover every eventuality on the road. People can quickly interpret situations that would confuse autonomous vehicles. One example might be children playing a new game that involves unpredictable movements near roads; another could be people chasing runaway pets. Australian mining companies use autonomous trucks, but these operate in relatively controlled environments where human work is already highly automated. As a result, these vehicles can move around without causing major hazards.[12] We explore enveloping for autonomous vehicles in greater detail in Box 6.3.

'Enveloping', therefore, is all about adapting the physical environment to enable AI applications to function properly. How will the necessary adaptations affect our expectations of AI applications? Firstly, we should not focus purely on the capabilities of autonomous vehicles, robots or drones. If we adapt their environment, these applications will make even more progress. Before drones can be used to deliver goods, for instance, we need to standardize their landing sites. They will also need new types of letterboxes, safe routes to avoid hitting people and systems that verify the identities of recipients.

Secondly, environmental constraints affect the enveloping concept as well. In the beginning, therefore, many AI applications will be deployed mainly in specific environments. Their use can be expanded at a later stage. Take robots. People have long fantasized about having a robotic assistant in the home, but that is probably one of the last places in which they will be used. This is due to the wide range of potential tasks in the domestic environment (cleaning, holding a dinner party, supervising homework, personal hygiene), all of which involve very high levels of complexity and many unpredictable variables (small children, pets, fragile objects). See Box 6.4.

[12] Agrawal et al., 2018: 113.

Box 6.3: Autonomous Vehicles and the Physical Environment
The path to autonomous driving

Each year KPMG publishes its Autonomous Vehicle Readiness Index. The key predictors in this regard (KPMG, Index 2019) are the quality of road surfaces, road design and signage. It is important to maintain surfaces properly, for instance, so that markings remain clearly visible at all times, even under poor weather conditions and following wear and tear. The boundaries of the carriageway also need to be clearly defined. Roadworks (and especially unannounced urgent works) pose a particular challenge in this regard. Workers often overpaint or erase existing markings or replace them with yellow temporary markings. This creates two conflicting sources of data, which can pose problems for autonomous vehicles. Country-specific road-design features can also create difficulties. Take the 'peak-hour lanes' on Dutch motorways: during rush hours motorists are sometimes allowed to use the hard shoulder as an extra running lane. This means that they can ignore the continuous white line, but only at specific times.

Enveloping is a way of adapting the system's environment. It avoids the need to constantly tweak AI applications until their performance reaches the required level. Instead, the environment itself can be modified to make it more 'readable' for an AI system, which is then be able to deal with it more effectively even if it is still unable to match human capabilities. Compare this to the approach taken by the Wright brothers, who developed the first aeroplane (Broussard, 2019: 131). People once believed that flying machines would need to imitate the flapping wings of birds. However, scientists have only recently devised ways to mimic nature in that respect. Orville and Wilbur Wright did not design their aircraft as mechanical birds, but based them on an entirely different principle. Similarly, if we adapt their environment autonomous vehicles will not need to match human ability.

The rollout of autonomous vehicles

How does this affect the deployment of autonomous vehicles now and within the foreseeable future? At first, they will probably operate in straightforward and predictable environments, like the robots mentioned above. Driverless buses could fairly easily be run on industrial estates, on airport aprons or in other relatively well-defined areas such as golf courses and care facilities, where there is little 'competing' traffic.

Autonomous vehicles may eventually be able to operate on motorways, too, but city centres will remain far more challenging. Their introduction could thus involve transferring people or goods from one mode of transport to another at specific locations. For instance, people might travel along motorways in driverless vehicles before transferring to alternatives with a human driver. That would require a fully integrated transport system, like the one currently used to co-ordinate bus and train services.

(continued)

Box 6.3 (continued)

We might even need to adopt a more prudent approach and implement more rigorous infrastructural measures before permitting autonomous vehicles to take to the roads. The following issue is a case in point. In complex environments, human drivers need to be able to take over the driving from AI systems. Studies have shown that this process takes about 20 seconds. Accordingly, we need to classify driving situations into those suitable for AI systems and those that require human drivers. We could use the operational design domain (ODD) concept to tackle this issue. This defines an area within which an autonomous vehicle can operate effectively.

There are also more rigorous possible measures, such as the construction of separate lanes or roads for the sole use of these vehicles. The same thing happened when people first started to travel in cars and countries designated certain highways for the exclusive use of motor traffic. Other road users, such as cyclists and pedestrians, had to use a different network. Autonomous vehicles, too, could initially operate in highly controlled areas. Next, following a series of modifications they could gradually extend their operational domain step by step.

Box 6.4: Vacuum Cleaners and Houses

Roomba is a disc-shaped autonomous vacuum cleaner produced by the iRobot company. It finds its way around the house using cameras and data from previous cleaning cycles. One problem with Roomba's usual operating environment is that there are so many corners; the device cannot access these areas due to its circular shape. However, this is nothing compared to the problems experienced by pet owners. Roomba is very effective at cleaning up dust and dirt, but less so when it comes to animal droppings: it tends to spread this material throughout the house, a phenomenon that has become known as the 'poopocalypse'.

Luciano Floridi had an interesting idea about right-angled shapes, which is relevant to the concept of enveloping. As a thought experiment, he suggested that we should all live in circular rooms in future. Roomba would then be much more effective, but many would object to the idea of having to adapt their lives to technology in this way, rather than the other way around. Yet Floridi wonders if we are not already doing that. After all, one reason our rooms are square in the first place is that we build them with rectangular bricks.[13]

[13] Floridi, 2014: 150–1.

Stuart Russell maintains that because of this, robots will be introduced in stages via other domains. They will be used in warehouses first, much like Amazon's robots. The tasks there are clear and simple ('take X to Y'), and the environments controlled. Robots can operate efficiently in these surroundings.[14] Next they could be used in other commercial environments, such as agriculture and construction. Here the tasks and objects involved are reasonably predictable. The next step is shelf filling and sorting clothes in the retail sector. In domestic environments, robots will first be used to assist the elderly and people with disabilities with specific tasks. Even then it will still be many years before we have universal robot butlers.[15]

This phasing is particularly important in situations where the use of AI can place peoples' lives on the line. Domestic robots or vehicles in city centres are prime examples. In other situations, the risks involved are more acceptable. We could introduce applications into less controlled environments before their operational capabilities have been perfected. Take virtual assistants, for example. Alexa and Siri are already widely used in domestic settings. Clearly though, we cannot yet have normal conversations with these applications. We have to pronounce words and structure our sentences in a specific way, otherwise the program is unable to understand us. Even then there is no guarantee that we will receive the right answer. Yet we still find these applications in many households. This is because they are just about good enough for limited purposes ('Where is the nearest bicycle repair shop?') and because people love gadgets. Moreover, they collect huge amounts of data in these surroundings, which will eventually make them more useful.

To conclude, there is one final implication. AI systems operate much faster in new, specially customized environments than when integrated into existing ones. For this reason, we should build new infrastructure or take AI applications into account when doing so. This explains why China has made great strides with AI applications. In that rapidly urbanizing country, new districts and entire cities are springing up everywhere. So, planners can design them to handle autonomous vehicles, for example, right from the start.

Key Points – The Technical Ecosystem: Supporting Technology

- Supporting technologies are part of the technical ecosystem.
- AI requires supporting hardware in the form of networks, chip technology and supercomputers.
- It also needs raw materials in the form of data that has to be broad-based, high-quality, commensurable, accessible and representative.
- Enveloping is an effective but underestimated strategy. People have successfully used it to implement new technologies. The environment is adapted to the technology, enabling it to operate more effectively.

[14] Kiva Robots' video clip demonstrates how that works: Amazon (YouTube, 24 July 2017).

[15] Russell, 2019: 74.

6.1.3 *Emergent Technologies*

Supporting technologies show that a new technology is part of a broader technical ecosystem. For users this creates a degree of complexity and uncertainty. That is even more applicable to emergent technologies. These initially had nothing to do with the new technology, unlike supporting technology that was associated with it from the very beginning. Emergent technologies develop in parallel, elsewhere or at a later time, after which they are linked to the relevant technology. Compared with supporting technologies, the process of embedding these emergent technologies is even more difficult to foresee. At first people only used electricity for a limited number of purposes in domestic settings. As time went by, though, links developed with other innovations. No one could have imagined how that would lead to the complete electrification of households due to the introduction of all kinds of domestic appliances.

This uncertainty about emergent technologies also applies to AI. Its application in society is a relatively recent phenomenon. Various new technologies have been developing in parallel with its rise. It is impossible to predict how these might eventually link with AI, especially when they themselves are still in their infancy. Nevertheless, those links could lend a huge impetus to AI or propel its application in particular directions. For this reason, we now briefly explore various emergent technologies that could become linked to AI. We begin with the most mature and work our way down to more recent arrivals.

We have already described network technology as a supporting technology. This rapidly evolving technology has recently sired a new generation, 5G, which represents a leap forward in terms of speed. It also uses different infrastructures and paves the way for other applications. People are currently experimenting with the rollout of 5G. This work will naturally impact the capabilities of AI (see text Box 6.5).

Another technology, the so-called 'internet of things' (IoT), is also paralleling the rise of AI. It is already at an advanced stage of development. Researchers are installing sensors and chips in all kinds of objects in the physical environment, which can then be connected to the internet. Developments in nanotechnology are driving this process by shrinking the size and cost of hardware. Roads and traffic lights are just some of the things that will be connected to the IoT. The list also includes dykes, toasters, toys, speakers, factories, refrigerators, clothes and even animals and our own bodies.

Cisco, an American company that manufactures much of the hardware, says that the tipping point came in 2008–2009 when more objects were connected to the internet than people. The International Data Corporation estimates that more than 40 billion devices throughout the world will be connected to the IoT by 2025. Moreover, that technology will increasingly be linked to AI. This is all due to data, one of the building blocks of AI. In recent years people have added an immense amount of data to the internet. That has given AI a massive impetus, and IoT will

Box 6.5: Autonomous Vehicles and Emergent Technologies
Autonomous vehicles require an effective digital network. The next-generation network, 5G, could play an important role here. Speed is of the essence on the road, after all – more so than in other areas. Faltering connections or vehicles that are slow to apply their brakes can mean the difference between life and death. 5G is much faster than previous generations, and that is essential here. In addition, these networks have much lower latency (the time between sending and receiving a signal), which is also crucial. When we transitioned to 3G and 4G, we were able to stream videos and movies on smartphones. The previous network was simply too limited for this. By the same token, according to some experts 5G can pave the way for effective autonomous vehicles.

The electric car is another emergent technology for autonomous vehicles. It is no coincidence that many electric cars also use advanced computer systems (Tesla, Nissan Leaf, Volvo). Both technologies require a sophisticated automatic transmission system, too. So, it makes sense to link the associated new infrastructure for electric cars (such as charging points) to infrastructural facilities for autonomous vehicles.

enhance this effect by collecting new data about the physical world. In this way it will become a key factor for new AI applications.

Yet another emergent technology is cryptocurrencies and the blockchain technology on which they are based. There has been a lot of talk about these in recent years, and especially about Bitcoin, the most well-known. This technology is hype-sensitive, though, as demonstrated by fluctuations in the value of 'crypto'. Nor can we yet predict how and on what scale it will be applied. Nonetheless, it clearly presents enormous opportunities, especially in combination with other technologies. Cryptocurrencies use the blockchain to facilitate a decentralized payment system, and that could be linked to AI to detect the use of someone's intellectual property, song or article, after which that party would be paid automatically.[16] People could conceivably control everything from bicycle locks to home systems connected to the IoT, operating them remotely through digital communication. In the same way, platforms such as Airbnb could grant access to a property for a prepaid period. People or organizations could use the digital route to make access to certain objects or locations a part of physical reality.

In addition to payments, the underlying blockchain technology can be used to decentralize all kinds of other transactions. One potential benefit is that this provides greater security while reducing dependence on central players or databases. Although the latter has its drawbacks, these properties can lower the barriers to all kinds of AI applications. AI and blockchain intersect in 'DAOs' (distributed

[16] Greenfield, 2017: 133.

autonomous organizations). Instead of people, these consist of automated rules and contracts that can make decisions automatically.

Quantum computing is an even less mature technology. But it promises to give the power of computers an immense boost. Simply put, traditional computers use bits in a binary logic of ones and zeros whereas quantum computers operate with quantum bits or qubits. These can simultaneously exist in multiple states, greatly increasing the number of calculations a device can perform.[17] Instead of using brute computing power, they represent all possible configurations at once.

The technology is still developing, and people are trying a range of approaches. Yet, these devices do not outperform regular computers in practical applications. Once they do – a point described as 'quantum supremacy' – according to experts this will represent an immense leap forward. One that would immediately invalidate any encryption systems based on huge amounts of computing power – just like giving someone the keys to every safe in the world at once. This is why countries like the US and China are heavily backing 'quantum'. Between 2019 and 2028 the US will invest more than US$1.2 trillion. China is building a National Laboratory for Quantum Information Sciences. Europe, too, is active in this area. The EU plans to use its 'Quantum Flagship' to strengthen the European research tradition and to build a competitive quantum industry.[18]

Even though quantum computing is still in its infancy, it is easy to imagine how this technology might revolutionize the use of AI. As we have seen, the growth in computing power is one of its pillars. If quantum computing is combined with AI, this could give a huge boost to highly complex data analysis issues or to scientific research into medicines, for example. It is no coincidence that parties such as Google are already pushing ahead with research into 'quantum AI'.

The above descriptions of supporting and emergent technologies show that system technologies like AI always operate within technical ecosystems. This entails a great deal of complexity and unpredictability. Developments in the technology itself, as well as elsewhere in the ecosystem, can facilitate or hinder its application. Which explains why some applications that work well in controlled or laboratory settings (such as autonomous vehicles on racetracks) seem to be quite mature, yet are far from ready for use in everyday life. On the other hand, improvements elsewhere can suddenly trigger great advances in what had appeared to be a stagnant technology.

Emergent technologies teach us that innovation in one type of technology can provide an enormous impetus to a completely different technology. For this reason, technologies like AI should not be developed in isolation. We need a clearer picture of new developments in other technologies and we need to invest in them as well. This is important for the future of AI. With this in mind, the planners of many AI strategies would be well advised to focus on emergent technologies such as the IoT,

[17] Vermaas et al., 2019.

[18] European Commission, 29 October 2018.

5G, blockchain and quantum computing. AI is deeply intertwined with other technologies. Which is why the Dutch AI strategy was eventually merged with the government's more broad-based digitization strategy.

Key Points – The Technical Ecosystem: Emergent Technologies
- Emergent technologies are ones that are initially distinct and separate. If linked together, however, they can have a major impact on further development.
- 5G, IoT, blockchain and quantum computing all appear to be candidate emergent technologies for AI.
- The future course of these other technologies cannot be predicted with any certainty. Nevertheless, it is prudent to include them in the aspirations and strategies associated with AI.
- Both dimensions of the technical ecosystem encompassing supporting and emergent technologies explain why a technology that is seemingly ready for use does not fully mature until much later. In other cases, however, the process of practical application can suddenly accelerate.

6.2 The Social Ecosystem

6.2.1 The Macroeconomic Context

Integrating AI into the social ecosystem raises two key issues at the macro level, in terms of the economy. The first involves its impact on employment in general and on 'technological unemployment' in particular. The second concerns what is known as the 'productivity paradox'. Both issues relate to the long-term impacts of AI, which cannot yet be predicted. At the same time the history of system technologies teaches us to examine these issues in a certain way while offering us the tools needed to steer the associated phenomena in the right direction.

The first issue is a recurring theme throughout the history of technological revolutions. This is the fear of huge job losses leaving large groups of people unable to support themselves. However, this cloud does have a silver lining. Once technology has freed us from boring, dangerous and physically demanding work, we will be able to engage in different, more meaningful activities. Karl Marx was one of the first to advance this idea. He stated that in the final stage of communism, people would spend their time hunting, fishing and writing critiques.[19] In 1930 the economist John Maynard Keynes predicted a future in which we would only need to work a few hours a day.[20]

[19] Marx & Engels, 2010 [1932].

[20] Keynes, 1930.

People these days are often amused to discover that past generations were afraid that work would disappear entirely. For centuries there have been concerns about the impact of developments such as ploughs, machines and ATMs, yet never have large scale job losses materialized. The Luddites discussed in Chap. 3 are symbolic of such inordinate fears. As manual weavers they feared that the Industrial Revolution would bring unemployment. Instead, it created all kinds of new jobs. Yet the Luddites did have a point.[21] Here it is important to note that while work has been a constant aspect of life throughout human history, we cannot assume that this will always be the case. Indeed, various authors argue that modern technologies like AI are quite different from their earlier forerunners.

One widely acclaimed book in this genre is *The Second Machine Age* by Erik Brynjolfsson and Andrew McAfee. The authors contend that contemporary digital technologies such as AI are also GPTs. They take the view that the first machine age – the Industrial Revolution – was complementary to human work but see the technologies of the second as substitutive. The first machine age replaced muscle power and led to a process of 'deskilling' in which the complex virtuosity of all kinds of craftsmen was subdivided into simple tasks that could be performed by large numbers of unskilled labourers in factories. However, the current machine age is also replacing our mental abilities. According to Brynjolfsson and McAfee, this will rapidly render human labour redundant. Their main supporting evidence is what they refer to as the 'spread'. This is the growing inequality gap in today's technology, where the wages of large groups of employees are lagging behind.[22]

Two scientists at Oxford have published a study that prompted serious concerns about AI's impact on the future of work. In 2013 Carl Benedikt Frey and Michael A. Osborne predicted that 47% of American jobs could be automated within the next 10–20 years. Even though they stated only that this would become technically possible, not that it would actually happen, their paper immediately sparked uproar around the world.

In 2016 OECD economists suggested that the situation was less dramatic than Frey and Osborne's study intimated. They found that 9% of jobs are at risk. The authors arrived at this figure by focusing on tasks rather than on jobs. Many individual tasks can be automated, but the same cannot be said of the overall job itself. In 2017 PwC estimated that 38% of US jobs were at high risk of being automated by the early 2030s. According to a McKinsey study, 50% of jobs throughout the world can already be automated.

In this context it is worth noting that adding AI to the mix has changed things. Automation now has a different impact than it has done in the past. This has to do with Moravec's paradox, which has already been mentioned here and states that some things we find difficult are easy for computers and vice versa. Previous phases of automation mainly replaced physical factory labour. However, AI impacts a wide

[21] Tielbeke, 16 May 2018.

[22] Brynjolfsson & McAfee, 2014.

range of human intellectual and conceptual skills. These correspond to administrative, financial and other 'white-collar' jobs.[23] As yet however, computers are unable to match the motor skills of hairdressers, drivers or cleaners.

People have responded to these scenarios by devising all kinds of solutions to deal with the loss of employment. Silicon Valley, where these changes originated, has also put forward various ideas. For example, Google's Larry Page suggested that we adopt a shorter working week. If the remaining jobs were shared in this way, more people would be able to find employment. Many people have proposed that we introduce a universal basic income. Yet we cannot predict AI's ultimate impact on the labour market. The WRR has explored this issue in greater detail in other studies.[24] Here, in keeping with them, we question the notion that most jobs will disappear.

Firstly, the history of system technologies amply illustrates the recurrent nature of these fears. People are more aware of jobs that have disappeared than of new ones that have been created. The same goes for AI today. Despite the projections made in the studies mentioned above, labour market figures show no structural decline in the number of jobs. Some sectors are even suffering massive worker shortages.

We are also unclear about the causes of certain phenomena, such as the inequality or 'spread' in wages. That is of key importance in this context. Kai-Fu Lee, like Brynjolfsson and McAfee, attributes this to the nature of the technology in question. But that is only part of the story. Technologies like AI will certainly contribute to the disappearance of jobs in the middle segment of the workforce, and to the concentration of capital at the top. At the same time, though, many other factors have a major impact in this regard. Neoliberal policies are one example. They have weakened the position of organized labour, restricted social safety nets and reduced the levelling effect of the tax burden. They have also contributed towards the stagnation of many people's wages. Globalization is another factor. Emerging countries – especially in Asia – have flooded the global market with cheap labour, which has had an adverse impact on wages.

Moreover, in a report entitled *Het Betere Werk* ('Better work') and as discussed in the previous chapter, the WRR stresses that we are still largely unaware of the ultimate impact of technology.[25] How we harness it and how it impacts employment are underpinned by economic and political choices. We should therefore be wary of claims that the effects we are now seeing are largely inherent to technologies such as AI. In fact, the very notion of AI's societal integration is all about managing its use more consciously and, as part of that, safeguarding the public interest.

We may have our doubts about the idea that most jobs will disappear, but this does not mean that we should ignore the impact of AI on the labour market. Many jobs will continue to exist but, given AI's increasing prominence, their nature in the future is very unlikely to match people's current skill sets. That is the real issue in

[23] Lee, 2018: 166.

[24] WRR, 2013, 2015, 2020.

[25] WRR, 2020.

terms of AI's impact on the labour market. 'Technologization' is just one of the fundamental changes now taking place in the world of work, and people need to adapt to it.[26] This, too, is in line with the lessons learned from system technologies. The Industrial Revolution generated all sorts of new jobs, but the transition was arduous and painful. This phase was accompanied by unemployment, accidents and misery in the overcrowded inner cities of Europe. Moreover, the new working conditions still lacked adequate rules and frameworks. In the nineteenth century this led to child labour and to the exploitation of workers, as depicted in the novels of Charles Dickens. People had to learn new skills, and employment malpractices had to be addressed.

Even today the process of embedding AI in the world of work is a two-pronged overarching task. First, we need to shift the topic of debate from job loss to job transformation. This requires us to dispense with the idea that man has to compete against the machine – a point nicely illustrated by Dutch chess grandmaster Jan Hein Donner. When asked how he would prepare for a match against a computer, he replied, "I would bring a hammer."[27]

Rather than 'man versus machine', the focus should be 'man with machine'. Seen in this way, AI is primarily about boosting human intelligence rather than replacing it – a process known as 'intelligence augmentation' (IA). Frank Pasquale feels that contrary to all kinds of alarmist stories ("Software is eating the world"), AI actually supports and empowers people in the performance of their work.[28] The renowned AI researcher Geoffrey Hinton once stated that we should stop training radiologists right away. However, the authors of *Prediction Machines* show that AI can be a useful aid to these specialists in their work. Furthermore, they as human beings play at least five roles that cannot yet be replaced by AI systems.[29]

To create effective man-machine combinations, people need experience with – and knowledge of – AI. Practical knowledge, in particular, is relevant here. As with electricity (see above), during this societal integration phase we need to consider how the new technology might enrich all kinds of domains, devices and practices, and how this can be achieved responsibly. We discuss the specific implications of this approach, in terms of human work, below.

People also need to explore AI's impact on working conditions in greater depth. Today, as during the Industrial Revolution, the jobs created by new technology are subject to all kinds of employment malpractices. The conditions and rights of workers on platforms like Uber and Deliveroo are a case in point. Then there is the plight of those employed at Amazon distribution centres. Their toilet breaks are meticulously monitored, and their working conditions are determined by the 'rate', which formulates objectives dynamically. At the same time employers are using AI to expand employee surveillance. This is a rapidly growing trend throughout the

[26] Ibid.

[27] Brynjolfsson & McAfee, 2014: 189.

[28] Pasquale, 2020: 13–14.

[29] Agrawal et al., 2018: 145–148.

economy. The AI Now Institute has documented a variety of cases in which technology requires people to work under appalling conditions. These range from migrant labour in agriculture to sensors that tell workers how to walk and what to do.[30] Other organizations have also warned about the growing trend of digital monitoring in the workplace, especially in the light of people working from home during the COVID-19 pandemic.[31] So although AI will not replace massive numbers of people in the short term, it will partly automate and transform some jobs while also impacting working conditions.

The second major macroeconomic issue associated with system technologies is the productivity paradox. This is where people often have wildly overblown expectations of such new technologies when in fact their actual impact on economic productivity is often disappointing, at least in the short term. In this context Robert Solow famously remarked in 1987 that "you can see the computer age everywhere but in the productivity statistics".

This is also an issue that has arisen in the context of AI. There is an article on this topic in the National Bureau of Economic Research (NBER) publication *The Economics of Artificial Intelligence*, in which the technology is treated as a GPT.[32] The authors point out that, despite our lofty expectations of AI, we are experiencing a period of weak productivity growth. Between 2005 and 2016, US productivity grew by just 1.3% per year, compared with 2.8% in the period 1995–2004. Various OECD studies show that this is a widespread global phenomenon. The authors also conclude that the slowdown cannot be attributed to the impact of the 2008–2009 global recession. They explore three explanations that could account for this phenomenon to a limited extent, if at all. These are false hopes about the impact of AI, inaccurate measurements of productivity growth and the limited dissemination of AI's benefits. The latter is the only explanation for which there is any significant evidence.

The explanation concerning false hopes for AI merits further examination. In his book *The Rise and Fall of American Growth*, Robert Gordon develops a detailed argument in support of this view.[33] He also draws comparisons with previous far-reaching technological revolutions – the railways, the steamship and the telegraph. Those brought immense improvements to everyday life. Mechanization and household appliances made work easier, better sanitation meant less disease, electric lighting and canned food made our lives more pleasant and there were huge gains in life expectancy. According to Gordon, that kind of progress was a one-off development; current digital technologies will not be able to repeat it. He uses productivity figures to illustrate this point. From 1920 to 1970 productivity grew at an average annual rate of 2.8%. It subsequently declined (with a brief exception between 1995 and 2005) to 1.7–1.8%, its previous level. Gordon accounts for this numerical

[30] Crawford et al., 2019: 14–16.

[31] See, for example, Das et al., 2020; EPRS, 2020; TUC, 2020; Scassa, 2021.

[32] Brynjolfsson et al., 2019.

[33] Gordon, 2016.

discrepancy by noting that digital technology has primarily impacted communications. In other areas of life, it has had less overall effect than older technologies. Moreover, current developments in the areas of inequality and education will also contribute to lower productivity growth in the future.

Although Gordon makes a sound argument, some feel that he has not taken sufficient account of recent breakthroughs in AI and tends to underestimate their potential impact. Today many people's expectations concern sectors outside the field of communication, such as mobility, healthcare and education. Carlota Perez argues that there will be productivity increases in a future phase, as the effects of a technological revolution spread throughout the economy.[34] The phenomena spotlighted by Gordon, like economic inequality, can certainly have an adverse impact on efforts to spread the benefits of AI far and wide, but such phenomena are not necessarily connected to AI.

Accordingly, the authors of the NBER's book on AI and the economy argue that the productivity paradox is more likely due to the time taken to implement and restructure as a result of the new technology. The other three explanations are based on the assumption that one side of the paradox is incorrect. They argue either that there will be no productivity growth in the case of false hopes (first explanation) or unequal dissemination (third explanation), or that such growth is already taking place but has not yet been measured (second explanation). In the fourth explanation, based on delay, both observations are correct. People quite rightly have lofty expectations, but these have yet to be realized. In fact, the impacts involve such a big change that it is naturally going to take time to make that transition.[35] That is in keeping with our ideas concerning contextualization. We have already pointed out the technical factors that need to be in place in order for AI to work. From a macroeconomic perspective this involves the development of new business models, the design of various other types of processes in organizations, efficiency gains and price reductions.

The roboticist Rodney Brooks, who we encountered earlier in the context of the overarching task of demystification, sees AI in the same way. He goes so far as to state that it takes 30 years to progress from the laboratory to a practical product. In the case of AI, technical breakthroughs in the backpropagation algorithm, for example, date back to the 1980s.[36] The same applies to autonomous vehicles. Even if these are technically feasible, people still doubt that they could be integrated into the processes and rules of road traffic. At what points along the road would autonomous vehicles be able to stop and pick up passengers? How might other road users respond to them? Will we still need traffic lights and other road features designed for human use rather than for autonomous vehicles?[37] To paraphrase Robert Solow, we could say that we currently see autonomous vehicles everywhere, except on the

[34] Perez, 2016.

[35] Brynjolfsson et al., 2019: 42.

[36] Ford, 2018: 428–429.

[37] Pasquale, 2020: 21.

roads. In addition to its technological prerequisites, embedding AI requires a process of societal change – and that will take time.[38]

Key Points – The Social Ecosystem: Macroeconomic Context

- People are afraid that AI technology will lead to mass unemployment. Nobody can predict the future, but there are reasons to suspect that such fears are groundless. There are, however, more pressing questions about the impact of AI on work.
- On balance, AI may not eliminate jobs. It mainly seems to require different skills on the part of employers and employees.
- However, AI could adversely impact working conditions – through the use of employee surveillance, for example.
- Economic and political choices underpin the way in which AI is used in practice and how it impacts employment.
- Besides its impact on work, questions have also been raised in the macroeconomic context, with regard to the productivity paradox. AI has the potential to trigger a great deal of change, so there is all the more reason to assume that a lag effect will be involved.

6.2.2 The Behavioural Context

At the micro level, too, we need to focus on how AI is embedded in the social ecosystem. More specifically we must examine the behavioural context in which it will be used. We can start by pointing out that developers frequently fail to give due consideration to a new technology's intended role within existing procedures and working methods.[39] This is a particular issue in the field of healthcare. Developers produce new items of software or apps without considering how medical professionals will use them in everyday practice. Can they rely on the app? Who has access to its data? How should doctors respond to patients who use apps to make a self-diagnosis at home? Developers should avoid devising solutions that address only specific individual aspects of the care process. The best approach is to embed that technology within broader behavioural patterns, in this case those of medical professionals and their patients.

Another point in this context is that those involved need to take receptiveness into account. Even if something works well, people can still find reasons to reject it. One key aspect to reckon with is that the technology could pose a threat to the work of the person concerned. Many hospitals or healthcare professionals are assessed by the number of treatments they administer. In these situations, any technology that renders such treatments redundant is a potential threat. If we want the new

[38] In this respect a Social and Economic Council of the Netherlands advisory report on robotization explored the need for 'social innovation' (SER, 2016).

[39] Dealing with people's reticence to use new technologies stems from what Jane Bennett refers to as the "material recalcitrance of cultural products" (Greenfield, 2017: 307).

technology to function properly, we may need to redesign an entire process in order to change people's motivations.[40] In the educational system, too, teachers may see AI as a threat. In this regard a Dutch study has highlighted the importance of acceptance, by encouraging teaching staff to acquire digital skills, and of conducting experiments.[41]

Another important behavioural issue concerns the specific nature of AI. In many contexts it may autonomously take decisions that would normally be a human responsibility. In this respect it is entirely unlike previous technologies. The key issue here, then, is achieving the optimum degree of interaction between man and machine when taking particular decisions.

This can be tackled by a model that distinguishes three forms of human-machine interaction: 'human in the loop', 'human on the loop' and 'human out of the loop'. In the first of these, while an AI system may be involved in the process, ultimate responsibility for any decisions rests with a human being. This is a standard aspect of the 'loop'. It means that if no people are involved, no decisions can be taken. In the second type, 'human on the loop', people play a smaller part. In theory, an AI system of this kind can take independent decisions without any human intervention. Nevertheless, the process is monitored by a human being who is able to intervene and make changes. In the final type, 'human out of the loop', the AI system acts completely autonomously. People are no longer involved in the process.

The latter form of interaction is used in many situations involving activities not of vital importance to people, such as recommending certain films or products. In some uncomplicated situations, we rely on algorithms to make the right decisions. When a roadside camera records a speeding violation, for instance, the driver is fined automatically. No human operators are involved.

In situations of great importance to people's lives, it is essential to include a human in the process. This right is enshrined in European privacy legislation. According to Article 15 of the EU Data Protection Directive:

> Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.[42]

The directive gives no precise definition of decisions that 'significantly affect' people, however, so there is ongoing debate concerning their delineation. Using this as a starting point, the EU has since published recommendations concerning the use of AI.[43]

---

[40] A similar thing happened several centuries ago when coffee was first introduced to Europe. Brewers in many countries saw the new drink as a threat, so innkeepers objected the increasingly popular coffee houses. The authorities solved this problem by allowing each type of establishment to serve both beer and coffee. This meant that they were no longer competitors – and that the government could tax them all equally (Juma, 2016).

[41] Van der Vorst et al., 2019.

[42] General Data Protection Directive, Article 15.1.

[43] The draft AI Act (European Commission, 2021) is based on a risk approach to AI systems.

In some domains it may be sufficient to have a human being 'on the loop' to check that no mistakes are being made. In some domains, however, decisions have such a major impact that the authorities consider it essential for them to be monitored by a person 'in the loop'. This applies to autonomous vehicles, for example, see the detailed explanation in Box 6.7. Life-and-death situations play an even greater role in military applications. Autonomous weapon systems that independently identify and destroy their targets are a case in point. The armies of various countries are already conducting extensive trials with systems of this kind, but their potential application has attracted widespread opposition calling for 'meaningful human control'.[44]

In other contexts, too, such as combating fraud or allocating benefits, people can be very severely impacted by decisions. This was illustrated by the childcare allowances scandal in the Netherlands (see Box 6.6). In the UK, too, government uses automated systems for a variety of purposes, including the allocation of social security benefits.[45] Applications like this have direct impacts on people, which is a strong argument for permanent human monitoring. Those involved in integrating AI into the social ecosystem thus face challenges concerning the form that monitoring should take.

It seems that the three types of human-machine interaction offer a clear means of selecting the right design in various contexts. At the same time, though, this approach does also suffer from a number of problems.

Box 6.6: The Dutch Childcare Allowances Scandal

Between 2013 and 2019 the Dutch tax authorities used a self-learning algorithm to identify tax fraudsters. It picked out individuals who, supposedly, were wrongly receiving childcare allowances and demanded repayment. But in many cases these accusations turned out to be totally unfounded. This mistake went unrecognized for years, leaving thousands of parents and families with enormous debts.

In the Netherlands people can apply for government benefits if they need financial support with their fixed costs. For example, working parents can apply for an allowance to meet the costs of childcare. In 2013 however, the authorities discovered that Bulgarian criminals were abusing the system by applying for this allowance in the Netherlands and then returning to Bulgaria. The national tax administration responded by designing an algorithm to detect fraudulent claims. This created a risk model based on several indicators that

(continued)

[44] References include: COMEST, 2017; Horowitz & Scharre, 2015; AIV & Advisory Committee on Public International Law, 2015.

[45] Gov.uk, undated.

Box 6.6 (continued)
supposedly could identify those receiving payments they were not entitled to. The algorithm assigned a high-risk score to childcare allowances in particular. If an administrative error led to a discrepancy in a claim, for example, the recipient was placed on a blacklist. Their payments were then suspended, and they were required to refund any money they had already received.

In 2018 this approach became a political scandal when a group of journalists published details of the affected parents' stories. Further investigation revealed that the algorithm had assigned a higher risk score to holders of dual nationality and to low-income households.[46] The victims were promised €30,000 each in compensation, but many of those payments were delayed. On 15 January 2021 the affair led to the fall of the government when prime minister Mark Rutte and his cabinet submitted their collective resignation.

Firstly, some people may exhibit behaviour that does not align with the selected model. For example, those who are officially 'on the loop' or even 'in the loop' may suffer lapses of concentration. Alternatively, they may act recklessly due to their unwarranted trust in the technology in question. 'Automation bias' is a psychological mechanism that causes people to blindly follow a computer's suggestions, even if these are incorrect. The phenomenon of 'alert fatigue' has the opposite effect. When systems generate too many reports, people become overloaded with information and take these signals less seriously.[47]

A second problem concerns an insidious process that gradually erodes the significance of the human decision-making role. A prime example would be an algorithm that helps healthcare professionals to reach a diagnosis. Doctors still make the decisions and check the algorithm's suggestions. Over time however, the staff involved become habituated to this procedure and so their checks may become less rigorous. This is especially true of algorithms with a good track record. Today's doctors have all the skills needed to reach proper diagnoses without the aid of a computer. But over time successive generations of doctors may be less well trained in that particular skill set. Calculators have had a similar impact on the skills of mathematics students. Long-term familiarity and an increased work rate can also make it more difficult for human decision-makers to question the results produced by an algorithm. The people involved must be increasingly sure of themselves before they cast doubt on a commonly used and efficient process.

This mix of dynamics makes the human decision less meaningful, yet those implicated in these situations still bear responsibility for the outcome of that decision. Which can present the risk of a problematic intermediate phase emerging, where the algorithms are not yet good enough to make decisions autonomously, but

[46] Dutch Data Protection Authority, 2020.

[47] Pasquale, 2020: 37.

people are no longer able to intervene effectively. A situation that can lead to mistakes and, as a result, human suffering.

By extension, this involves a third challenge for the interaction model. After all, human control only makes sense in situations where algorithms are performing tasks that are normally undertaken manually. In many contexts, though, the algorithm's activities could quite conceivably become much faster and more complex over time.

When this happens, human control often becomes impossible or even hazardous. For example, the law only permits autonomous vehicles to use the roads if a person is behind the wheel to intervene if necessary. Cars could drive much closer together in the future, thanks to C2C communication (see Box 6.2). Human reaction times are too slow to be of use in this case, so human control would actually pose a hazard to other road users. Moreover, vehicles could use I2C communication to communicate directly with their surroundings. This technology may ultimately render road signs and even traffic regulations obsolete. But if they were to be discarded, it would then be very difficult for human drivers to navigate the road network. The use of autonomous weapons poses similar difficulties. People are capable of successfully attacking individual enemies, but what happens if the battlefield becomes much more complex? How would they cope with combat that involves large formations of drones, for example? Humans would be of no use here, as they cannot see the bigger picture and their reaction times are far too slow.

John Danaher presents a topical example from a very different domain. The products stored in traditional warehouses are organized by category. Anyone familiar with the category index can easily find their way around a facility of this kind. But Amazon warehouses use a dynamic storage algorithm to shelve products in the most efficient manner. This is based on complex calculations about future demand, involving a logic beyond human comprehension. To the casual observer, everything just appears to be jumbled up. People need algorithms to find their way around facilities like this. Which, says Danaher, poses the risk of creating an 'algocracy': a system governed by complex algorithms, which is beyond human comprehension (Box 6.7).[48]

In many contexts people place an overly simplistic emphasis on human control. The three challenges mentioned above raise doubts about the wisdom of this approach. They show that we need to focus on the complexity of issues associated with human-machine interactions. That complexity includes efforts to identify the strengths and weaknesses of humans and machines, which are very different. Machines are much better at detecting patterns in large quantities of data, for example. Humans on the other hand are generally more competent at using reason to resolve anomalies. Man and machine can interact effectively if their characteristics are co-ordinated properly. They can compensate for each other's weaknesses and

[48] Danaher, 2016.

Box 6.7: The Behavioural Context of Autonomous Transport

AI's behavioural context involves a range of issues that feature prominently in autonomous transport. Autonomous vehicles are still prohibited by law. This has nothing to do with their technical capabilities. It is simply that cars using the public highway must all be under the responsibility and control of a human driver. At the same time people often fail to behave appropriately, with serious consequences.

At a basic level this is already the case with navigation software. Drivers sometimes follow the instructions given by their satnav systems even when common sense and current road signs dictate otherwise. From time to time there are reports of people driving into the sea or along impassable nature trails, and some have even died in incidents known as 'death by GPS'.[49] These are classic examples of 'automation bias'.

So, although autonomous vehicles are very much in the spotlight, full autonomy is not yet a reality. In the meantime, all kinds of decision support software are now available, such as ADAS. As long as users remain responsible for making decisions, accidents will still happen if they fail to act appropriately.[50] People should therefore make greater allowance for the human factor when estimating the risks involved in automating transport.

Tesla's 'autopilot' function is a very specific case in point. The name suggests that the motorist can just sit back and leave the driving to the car. However, the owner's manual points out that this is not the case. Nevertheless, the company still refuses to change this misleading name even though many people are critical of the suggestion it creates. So effective communication and instruction are key factors in terms of human behaviour.

The behavioural effect of updates is a related issue. They are designed to improve the vehicles, causing them to respond to a specific situation in a certain way. Yet a later update may cause the same vehicle to respond to exactly the same situation in an entirely different manner. That can be difficult and confusing for the driver. Ergonomic features are important, too, as they can mitigate the risks posed by human behaviour. For instance, they can clearly show drivers which vehicle functions are currently active, and which are not.

Here again, the risk of a problematic intermediate phase may arise, as described above. Vehicles are not yet capable of handling all road traffic-related decisions autonomously. Yet people cannot be expected to keep their attention focused on the road during a long journey when the car is doing the driving. Accordingly, some people contend that this human factor is sufficient reason to ban experiments with semi-autonomous vehicles. They are unwilling to compromise, insisting that cars should either drive themselves or be driven by people.

[49] Bridle, 2018.

[50] OvV, 2019.

gain the maximum benefit from each other's strengths. Catholijn Jonker describes this as 'hybrid intelligence'.[51]

Different AI systems may have different properties, depending on how they are set up. Human editors can use AI redaction tools to automatically obscure certain passages of text. These tools can be set up in various ways. If the main aim is to prevent the disclosure of sensitive information, the algorithm can be set to 'heavy'. Conversely, if people feel that too little information is being disclosed, a much 'lighter' algorithm setting can be used.[52] The system's settings should therefore align with its operational context.

Key Points – The Social Ecosystem: Behavioural Context

- In the behavioural context, we need to take various factors into account when embedding AI systems. These are existing organizational structures, working methods and motives for human behaviour.
- The 'human in the loop'/'human on the loop'/'human out of the loop' model is a way to design interactions between man and machine. It can also distinguish between different degrees of human control.
- However, these highly distinct categories can be undermined by many kinds of behavioural factors. For this reason, we need a detailed examination of the design and use of technology.

6.3 In Conclusion

In this chapter we have explored the status quo regarding the overarching task of contextualization – integrating AI into the sociotechnical ecosystem. To a large extent this process cannot be centrally controlled. All sorts of organizations will go through it, both internally and externally. They will use it to innovate, to experiment with their processes and to achieve efficiency gains through improved production methods.

Nevertheless, governments can still play a key part. They could start by investing in good digital infrastructure, for example, or in further training. They could also capitalize on their own use of AI to influence contextualization. Public-sector organizations, especially executive agencies, can help others develop good contextualization practices and even to set standards. Governments could provide further assistance through their procurement policies. As major players and 'launching customers' they can foster emerging markets or nudge existing ones in a certain direction.

[51] TU Delft, undated.

[52] Agrawal et al., 2018: 68.

In Chap. 8 (Positioning) we review the issues associated with a country's competitiveness. It is important to remember that governments possess a broad palette of tools. This enables them to prioritize domains for AI applications and to encourage contextualization in those areas. Some of these could be associated with competitiveness and with a country's economic engine. Others could be of enormous importance to its society. This category includes healthcare and sustainability, domains in which government is specifically responsible for pioneering new developments. Countries can use this approach to focus more intensively on establishing an 'AI identity' of their own.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

AIV, CAVV. (2015). *Autonome Wapensystemen: de noodzaak van betekenisvolle menselijke controle*. AIV.

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the modern productivity Paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans en A. Goldfarb (2019) *The Economics of Artificial Intelligence: An Agenda* (pp. 23–57). University of Chicago Press.

Bridle, J. (2018). *New Dark Age: Technology and the end of the future*. Verso.

COMEST. (2017). Report of COMEST on Robotics Ethics. [Report] UNESCO. Geraadpleegd van: https://unesdoc.unesco.org/ark:/48223/pf0000253952

Crawford, K., Dobbe, T., Dryer, G., Fried, B., Green, E., Kaziunas, A., Kak, V., Mathur, E., McElroy, A., Sánchez, D., Raji, J., Rankin, R., Richardson, J., Schultz, S. W., & Whittaker, M. (2019). *AI Now 2019 Report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accomodation. *Philosophy & Technology, 29*(3), 245–268.

Das, D., de Jong, R., Kool, L., & Gerritsen, M. M. V. J. (2020). *Werken Op Waarde Geschat – Grenzen Aan Digitale Monitoring Op De Werkvloer Door Middel Van Data, Algoritmen En AI*. Rathenau Instituut.

Ding, J. (2019). The interests behind China's Artificial Intelligence dream. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 43–47). Air University Press.

EPRS. (2020). *Data subjects, digital surveillance, AI and the future of work*. European Parliament. Retrieved from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf

EuroHPC JU. (2021, April 20). Vega Online: The EU First Eurohpc Supercomputer Is Operational. Eurohpc-ju.europa. Retrieved from: https://eurohpc-ju.europa.eu/press-releases/vega-online-eu-first-eurohpc-supercomputer-operational

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.
France. (2018). *AI for Humanity: French strategy for Artificial Intelligence.* President of the French Republic. Available at: https://www.aiforhumanity.fr/en/
Gordon, R. (2016). *The rise and fall of American growth: The U.S. standard of living since the Civil War*. Princeton University Press.
Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.
Horowitz, M., & Scharre, P. (2015). *Meaningful human control in weapon systems: A Primer* (Working paper). Center for a New American Security. Retrieved from: https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf
Juma, C. (2016). *Innovation and its Enemies: Why people resist new technologies*. Oxford University Press.
Keynes, J. M. (1930). Economic possibilities for our grandchildren. In J. M. Keynes. (1932). *Essays in Persuasion* (pp. 358–373). Harcourt Brace.
Lee, K. F. (2018). *AI Superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.
Marx, K., & Engels, F. (2010 [1932]). *Die Deutsche Ideologie* (Vol. 17). Akademie Verlag.
OvV. (2019). *Wie Stuurt? Verkeersveiligheid En Automatisering In Het Wegverkeer*. Onderzoeksraad voor Veiligheid. Available at: https://www.onderzoeksraad.nl/nl/media/attachment/2019/11/28/wie_stuurt_verkeersveiligheid_en_automatisering_in_het_wegverkeer.pdf
Pasquale, F. (2020). *New Laws of Robotics: Defending human expertise in the age of AI*. Harvard University Press.
Perez, C. (2016). Capitalism, technology and a green global golden age: The role of history in helping to shape the future. In M. Jacobs and M. Mazzucato (reds.), *Rethinking Capitalism: Economics and policy for sustainable and inclusive growth* (pp. 191–217). Wiley.
Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.
RVS. (2019). *Waarde(N)Volle Zorgtechnologie. Een Verkennend Advies Over De Kansen En Risico's Van Kunstmatige Intelligentie In De Zorg*. Raad voor Volksgezondheid en Samenleving.
Scassa, T. (2021, June 8). *Privacy in the Precision Economy: The Rise of AI-Enabled Workplace Surveillance during the Pandemic*. CIGI. Retrieved from: https://www.cigionline.org/articles/privacy-in-the-precision-economy-the-rise-of-ai-enabled-workplace-surveillance-during-the-pandemic/
SER. (2016). *Mens En Technologie, Samen Aan Het Werk*. Sociaal-Economische Raad. Available at: https://www.ser.nl/-/media/ser/downloads/adviezen/2016/mens-technologie-publieksversie.pdf
Tielbeke, J. (2018, May 16). Lessen van de Luddieten. *De Groene Amsterdammer*. Available at: https://www.groene.nl/artikel/lessen-van-de-luddieten
TNO. (2021a). *Het Technologische Ecosysteem Van AI In Nederland* (WRR Working Paper nr. 47). Wetenschappelijke Raad voor het Regeringsbeleid.
TNO. (2021b, October 4). *Veiligere Europese Wegen Dankzij Doorbraak In Truck Platooning*. Retrieved from: https://www.tno.nl/nl/over-tno/nieuws/2021/10/veiligere-europese-wegen-dankzij-doorbraak-in-truck-platooning/
TOP500. (2021, November 16). *TOP500 list*. Retrieved from: https://www.top500.org/lists/top500/2021/11/
TUC. (2020). *Technology managing people. The worker experience. [Report]*. Trade Union Congress. Retrieved from: https://www.tuc.org.uk/sites/default/files/2020-11/Technology_Managing_People_Report_2020_AW_Optimised.pdf
Vorst, T. van der, Jelicic, N., de Vries, M. en Albers, J. (2019). *De (On)Mogelijkheden Van Kunstmatige Intelligentie In Het Onderwijs, Nr. 2018.068.1828*. Dialogic. Available at: https://www.dialogic.nl/wp-content/uploads/2019/04/Dialogic-De-onmogelijkheden-van-kunstmatige-intelligentie-in-het-onderwijs-v1.0.116.pdf
Vermaas, P., Nas, D., Vandersypen, L., & Elkouss Coronas, D. (2019). *Quantum Internet: The Internet's next big step*. TU Delft.

WRR. (2013). *Naar Een Lerende Economie*. Wetenschappelijke Raad voor het Regeringsbeleid.
WRR. (2015). *De Publieke Kern Van Het Internet. Naar Een Buitenlands Internetbeleid*. Amsterdam University Press.
WRR. (2020). *Het Betere Werk. De Nieuwe Maatschappelijke Opdracht*. Wetenschappelijke Raad voor het Regeringsbeleid.

Chapter 7
Engagement

The next overarching task we have identified for AI's integration into society concerns the engagement of stakeholders. This raises the following question: 'who should be involved?' When any new technology is introduced, after all, various parties are involved right from the start. The previous chapter on contextualization made this apparent; it showed that both companies and government started working with AI at an early stage. In discussing their involvement then, our focus was the question 'how do we make AI work'? Companies and government organizations have the resources and impetus needed to become key drivers of AI's use in society. As a result, they also have a lot of influence over how it is implemented in practice. In this chapter, by contrast, we home in on parties that do not initially use AI themselves but – given its ubiquitous use – are likely to encounter this technology in their activities. Our particular focus is parties in civil society.

In Chap. 4 we saw that the introduction of any new system technology is accompanied by social tensions and growing inequality. This is because certain groups are better able to use such new technologies than others. During the Industrial Revolution workers found themselves in precarious situations. The electrification of society was patchy, and for many years rural regions lagged behind other areas. Cars became associated with wealthy sections of the population, marginalizing less affluent road users. These developments also caused more indirect suffering. Companies used electric lighting to supervise workers more effectively. Cars polluted the air and made the roads hazardous for cyclists. The process of integration or embedding was thus almost always accompanied by malpractices and by irresponsible use of the new technology. We also saw that, over time, civil society groups began to actively oppose these wrongs and to correct imbalances in the use of the new technology. So, when introducing a new system technology, it is important to involve various groups in this process. This helps shape it and its application.

All societies are made up of numerous different parties. This is why civil society needs to be engaged in the embedding of AI. Only the most authoritarian regimes designate a single player – a leader or political party – to chart the course to be taken by society. Democratic constitutional states have a range of institutions designed to

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_7

counterbalance the power of the state. However, these are also protected by that very same state through such mechanisms as constitutional rights. A strong and well-developed civil society is thus an important precondition for the proper operation of the state and of the market.[1] Civil society can involve itself in the embedding of a new system technology in a variety of ways. Many of these options were not yet available during the Industrial Revolution. This was because workers were not united and universal suffrage had not yet been introduced. Today's stakeholders have many ways of expressing their views and making their presence felt. These include filing lawsuits, establishing new interest groups and participating in decision-making by both public and private bodies. Organizations may not always have a choice when it comes to involving stakeholders. In some countries works councils are legally entitled to participate in companies' decision-making processes. This gives them a say in the use of employee monitoring systems, such as cameras on the shop floor or smart tracking systems in vehicles.[2] These are not simply private initiatives; companies are formally obliged to engage with stakeholders.

In democratic societies stakeholders are free to engage in matters that have an impact on their own lives. This is valuable in itself. In addition, a society that has a broad-based engagement with technology can help to improve that technology.[3] Here people's responses to AI are not limited to the impact of its use, they also contribute their own knowledge and experience. Indeed, they can even start using the technology to promote their own interests and values. Various initiatives to involve stakeholders in technology development have been launched since the 1960s. Their goal is to raise awareness about its impact on society and to make use of technology more socially accountable. It is important to include values and moral considerations in the development of a new technology. This can also help avoid any risk of societal resistance to its use at a later stage.[4]

In some ways AI is no different from previous system technologies. As we shall see in this chapter, it is associated with all kinds of malpractices and worsening imbalances of power. A case from the UK exemplifies that dynamic in which AI perpetuates existing inequalities. This concerns an algorithm that was supposed to help predict final school exam grades. In fact, it put pupils from certain schools at a considerable disadvantage compared with others (see Box 7.1). By raising issues of this kind and making people aware of them, civil society is making a significant contribution towards the further societal integration of AI. In other words, engagement is a key overarching task when it comes to embedding AI in society. Various civil society parties are particularly important to this overarching task. These include interest groups, the media, scientists and other experts.

---

[1] Schuyt, 2006.

[2] Take the Works Council's right of consent, which is enshrined in Dutch legislation (the Works Councils Act).

[3] Sykes & Macnaghtan, 2013: 85–107.

[4] One example from the Netherlands is the electronic health record (EHR), another is the smart energy meter. For details see van den Hoven, 2013: 75–83.

Box 7.1: Unequal Opportunities for Success
In 2020 lockdowns prevented school students in the UK, like those in many other countries, from taking their final exams. Instead, their final grades were determined by an algorithm. The input was the expected grade per pupil and their individual rankings relative to other pupils. The authorities also included the school's performance in recent years in the calculation. They expected these estimated final grades for individual students to be more accurate than the teacher's estimate alone. This is because teachers often tend to overestimate their own pupils' performance.

In more than 35% of cases the algorithm did indeed predict a lower final grade than the teachers had. However, it downgraded pupils at state schools to a far greater extent than those who attended private institutions. The algorithm placed great emphasis on results from previous years. As a result, both state schools and individual pupils were unduly penalized due to these schools' relatively poor past performance. This focus on the past thus placed state schools that were on an upward trend, or individual pupils whose performance was improving, at a disadvantage. Private schools on the other hand have traditionally achieved better results. The current cohort has benefited from this. This example shows how algorithms intended to produce fairer predictions can confirm and prolong existing differences.[5]

The expertise of interest groups is a key issue when it comes to the impact of a new technology on disadvantage and equality. Disadvantaged groups in society are served by numerous national and international organizations. Many of these, however, are ill-equipped to pursue that work when confronted with a new technology like AI. This is because it opens up new dimensions to the unfair disadvantage suffered by certain groups. Accordingly, this changes the nature of those organizations' fields of work and the problems they have to address. We could say that AI versions of different forms of inequality are emerging, which require a change of perspective and additional expertise.

This applies to discrimination against people of colour, for example. Ruha Benjamin coined the term 'New Jim Code' to describe this phenomenon. That refers to the so-called Jim Crow laws that codified racial segregation in the US South. Today's code also disadvantages racial minorities, but in a different way. Benjamin defines the New Jim Code as "the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective and progressive than the discriminatory systems of a previous era".[6] In other words AI here provides a channel for discrimination. Moreover, this is far more insidious in nature.[7] Unlike discrimination by police officers, which has attracted so much

[5] BBC, 20 August 2020.

[6] Benjamin, 2019: 5.

[7] See also Wallace, 2021.

attention recently, this form is presented as objective and is less visible. That also makes it harder to identify and oppose. There are no racist bosses, bankers or shop owners to report here.[8] Indeed, many people present the principle of discrimination in a positive light. For example, services like Netflix tailor different trailers to different target groups. So, someone who feels that actors of colour are underrepresented in the movie industry will be shown a trailer that mainly features members of this group. But this can give the impression that a series is more diverse than is actually the case. The diversity or otherwise of Oscar winners is there for everyone to see. But there is no such clarity in the world of Netflix because everyone is presented with a different representation of actors and producers. Safiya Umoja Noble refers here to "algorithms of oppression".[9]

Also consider the exclusion of people with low incomes. Virginia Eubanks explains how, in times gone by, poor people in the US were oppressed and stigmatized in the poorhouse. Those sent these institutions were said to be lazy. They often had to work without pay for their upkeep. According to Eubanks, today's equivalent is the 'digital poorhouse'.[10] People's data points stigmatize them, which can make it more difficult for them to obtain insurance, mortgages and benefits.[11] Eubanks documents the insidious impacts of this digital poorhouse in all kinds of places. Here too, discrimination is openly presented in a positive light. A case in point is the growing range of insurance products that offer discounts in exchange for personal data. The companies involved use this to better predict their policyholders' behaviour and to customize their offers accordingly. But those applying this kind of price profiling are rarely transparent about the criteria, margins of error, parameters and analytical insights involved.[12] The creeping acceptance of this practice is insidiously fostering inequality.

Another issue is the violation of human rights. This has traditionally involved the incarceration of activists and dissidents. Nowadays though, technology such as AI can be used to facilitate digital exclusion and incarceration. For instance, China's highly developed social credit system excludes people with a low rating from trains and planes. Human rights organizations also refer to the 'open-air prison' in the Chinese province of Xinjiang. We discuss this in greater detail in Chap. 9.

Yet another example of an AI variant of inequality pertains to gender and sexual orientation. Caroline Criado Perez shows how many domains revolve around male views and interests, from work to care and politics. Which means that when they are digitalized, data relating to women is underrepresented. The authorities in these domains tend to view women simply as 'smaller men' rather than giving due consideration to all the ways in which the sexes can differ. As a result, many AI

[8] Benjamin, 2019: 33.

[9] Noble, 2018.

[10] Eubanks, 2018.

[11] For this reason Dutch regulators are assessing the risk of discrimination associated with the use of AI in the insurance sector. This survey broadly covers the models used for pricing, acceptance and fraud detection in particular. See AFM & DNB, 2019.

[12] Moerel & Prins, 2016a, b.

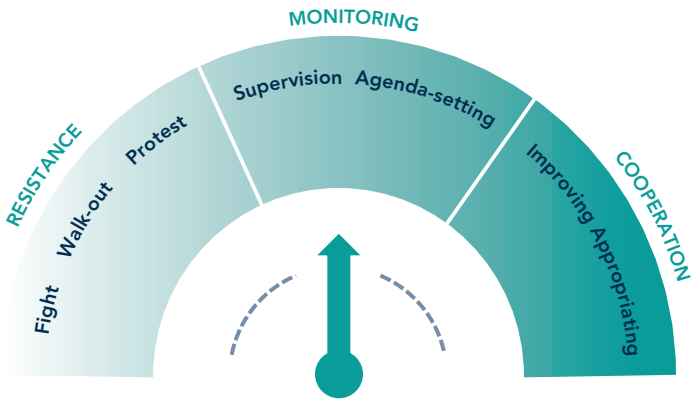


Fig. 7.1 A spectrum of different forms of engagement

applications do not work well for women.[13] New ways of excluding those with a different sexual orientation are also appearing. In the world of credit ratings, for example, Frank Pasquale shows that the presence of gay men is seen as a positive indicator for house prices.[14] In the past people were hostile to homosexuality, seeing it as a bad thing. That does not apply in this case. Yet this is still a type of different treatment, which can insidiously foster inequality.

All of the above areas share a common pattern. Throughout history various groups labelled as 'deviant' were pressurized to conform to societal norms. An AI world would not oppose difference in that way. Instead, this technology serves indirectly as a source of unequal treatment. People often present that difference as something positive, partly under the guise of providing opportunities for a more individual-centred approach.

In this chapter we discuss engagement as an overarching task that can take a variety of forms: fight, walkout, protest, supervision, agenda-setting, improving and appropriating. We have arranged those forms a continuum that reflects their relationship to AI (Fig. 7.1). At one extreme are those with an antagonistic attitude: individuals or groups opposed to AI or in favour of a ban on the technology, for example. We refer to this cluster as 'resistance'. Those at the other extreme have a symbiotic attitude. Such individuals or groups engage positively with AI by incorporating the technology into their everyday lives. We summarize this attitude as 'co-operation'. Intermediate types adopt a critical but not necessarily negative approach, which we refer to as 'monitoring'.

These different forms of engagement are archetypal and in practice often overlap. As we shall see, parties can utilize various forms simultaneously. The spectrum we present is simply a means to gain an overview of the extensive field of activities undertaken by diverse players in civil society with a view to tightening, bending,

[13] Perez, 2019.

[14] Pasquale, 2015: 25.

breaking or shifting existing practices and, on occasion, applicable laws and regulations as well. We discuss the various forms of engagement below, including their status regarding AI, so as to identify those currently prevalent and those requiring more work.

7.1 Resistance

New technologies often trigger resistance. This certainly applies in situations where there is rapid technological change and where people become convinced that the technology will only benefit a very limited section of society, while its risks are widespread.[15] We place resistance at the left end of the spectrum of engagement and subdivide into three forms. At the far left is the most antagonistic relationship with AI: fight. Stakeholders reject the new technology out of hand and resort to violence to oppose it. Next comes 'walkout', also referred to by Albert Hirschman as 'exit', characterized by stakeholders leaving the negotiating table. Hirschman contrasts 'exit' with 'voice', in which people articulate their dissatisfaction without terminating the relationship. The third form of resistance we have identified, protest, is an example of this. Here again, people oppose the technology. However, they do so in a peaceful manner while putting forward a clearly articulated counterproposal. That could involve a ban on a technology or further regulation of its use.

7.1.1 Fight: Violent Resistance

Historically, violent resistance has often been an iconic form of negative engagement with a new technology. In Chap. 4 we have mentioned the infamous Luddites who, worried that they might lose their jobs and incomes, proceeded to smash the newly introduced machines. In terms of resistance, they had very few other options. The owners of the machines were not prepared to listen to them, nor did the workers have any political representation.

In modern democracies groups at risk from technology have a range of non-violent options to make their voices heard. Also in Chap. 4, we argued that democratization distinguishes the embedding of later system technologies from that of their predecessors. Yet even in democratic societies some forms of resistance deliberately transgress legal boundaries and do not shy away from violence. The anti-nuclear campaign is one example: its members have invaded power plants and destroyed equipment.[16]

Is anyone fighting AI at the moment? Not yet, it seems. Perhaps this is because both the technology itself and its impact are not immediately visible. That makes it more difficult for people to locate and destroy. The intangible nature of AI means

[15] Juma, 2016.

[16] van der Vleuten et al., 2017: 135.

that resistance to this technology could become interwoven with opposition to more physical things, such as computers, robots or the companies that develop AI.

In 2014 an anarchist movement called The Counterforce took up arms against the influence of Silicon Valley in general and that of companies like Google in particular. The group accused these firms of driving up house prices in San Francisco, and as a result undermining ordinary people's quality of life. It also campaigned against the impact of digital technology on attention spans and against the construction of an infrastructure that could be used to facilitate totalitarianism. The Counterforce's resistance was not limited to demonstrations; it also encouraged people to obstruct staff buses in Silicon Valley, to steal software engineers' belongings and to tear down surveillance cameras. One of the people targeted was Anthony Levandowski, at the time responsible for the technology behind Google's autonomous vehicle.[17] In Hong Kong protesters used power saws to damage lampposts equipped with facial recognition equipment.[18]

Closer to home, AI has been targeted by groups with an agenda violent resistance – most recently as part of the social media narrative surrounding 5G and COVID-19 vaccines. American video-clips on YouTube link 5G, Huawei and AI with alleged Chinese plans to surreptitiously gather data on a global scale. There are also people who thought that 'COVID' stood for 'certificate of vaccination identification by artificial intelligence'. They saw a connection between A (the first letter of the alphabet) and I (the ninth) and the number 19 in COVID-19.[19] This theory, the brainchild of osteopath Carrie Madej, circulated on the internet in the spring of 2020. She claimed that the coronavirus vaccine was designed to rewrite our DNA to assimilate everyone into an interface (API or application programming interface) between man and machine that would enable our behaviour to be completely controlled by external agents. The prime suspect was Bill Gates. In conspiracy theories of this kind, AI is part of a narrative that has led people to vandalize telecommunications masts. This questionable form of engagement does not need to be reinforced. In fact, we should guard against potential escalation. The overarching task of demystification, which has already been discussed, can also help people to engage with AI more peacefully and democratically.

7.1.2 Walkout: Refuse to Co-operate

People can resist in a non-violent way by refusing to co-operate with something. This is the second form of engagement we have identified. Known as a 'walkout', participants include people working in the technology sector. They have a special weapon in their armoury, after all: the ability to 'down tools'. Those with specialist expertise can exert pressure by refusing to co-operate. Without their input some projects will be unable to take off. People with more widely available expertise can exert collective pressure, especially if they are able to publicize their campaign

[17] Jeffries, 15 april 2014.

[18] Fussel, 30 augustus 2019.

[19] Reuters, 24 april 2020.

successfully. This form of engagement has grown in recent years. Several so-called 'walkouts' have taken place in Silicon Valley. In the Netherlands a recent legal battle between the University of Amsterdam and its students falls into the same category. The students refused to allow themselves to be observed by AI-based online surveillance software (proctoring) during exams held in lockdown. In that sense, they opted for a 'walkout'.[20]

Many walkouts are in fact work stoppages that, as you might expect, are associated with working conditions. Campaigns to improve these occur in all types of companies, of course, but in some cases are linked specifically to the use of technologies such as AI. This is because new system technologies can create new working conditions, which can sometimes be very detrimental to employees. When electric lighting was first introduced, it enabled employers to monitor their workers more effectively – just as algorithms are doing now, with tools ranging from trackers of office workers' internet surfing behaviour (even using biometric information such as eye movements) to the micromanagement of staff in warehouses and delivery services (see Box 7.2).[21]

Box 7.2: Worker Surveillance

Employers have been gathering data about their workers and using it to manage them for at least a hundred years. But now they are deploying today's improved technology to conduct surveillance and monitoring that are more in-depth, more variable, more fine-grained, larger in scale and more rapid than ever before.[22] AI is often used to analyse employee data in the context of personnel policy.[23] Based on the information gathered and the purposes to which it is put, four sub-trends can be identified.

1. Systems that use various types of data to predict worker behaviour, including potentially unacceptable conduct.
2. Systems that make inferences about working conditions based on biometrics and health data, giving employees insights into their own health but also serving as tracking systems.
3. Systems that remotely monitor worker behaviour to measure their performance and determine their pay.
4. Systems designed to facilitate the 'algorithmic control' or 'gamification' of work through the continuous collection of performance data.[24]

[20] In the court case that followed, both the district court and the court of appeal ruled that the UvA's use of the software was lawful (Rechtbank Amsterdam, 11 juni 2020; Gerechtshof Amsterdam, 1 juni 2021).

[21] For example, the AI Now Institute's 2019 report describes the effect of using 'the rate' to manage workers in Amazon's warehouses.

[22] Mateescu & Ngyun, 2019.

[23] See, for example, Das et al., 2020.

[24] Mateescu & Ngyun, 2019.

One example from the Netherlands concerns an app used by PostNL, a postal company. This calculates the routes delivery workers should follow and how long their rounds should take. Any employee exceeding the allotted time can expect that to have disciplinary consequences, but the app takes little account of factors like the weather or leftover mail from the previous day.[25] Amazon uses a similar app, Mentor, to track and assess its delivery staff. It plans to expand this with AI-compatible cameras.[26]

To develop a more fact-based personnel policy, many other companies and government bodies collect data about their employees' state of mind, their health and even their job motivation. These organizations frequently use language analysis to scan e-mails sent between employees for characteristics such as 'enthusiasm'. In many cases, however, their staff are entirely unaware of this. But analyses of this kind are problematic in that their validity has not been proven.

Platform work is a new occupational category that has actually been created by technology.[27] Taxi drivers or riders who deliver meals or parcels are not official employees with the right to a minimum wage or secondary benefits. As a result, many have precarious livelihoods. This development is being countered by the rise of trade unions and by lawsuits to compel employers to recognize these people as employees.[28]

Employees in the US have launched several major legal actions that specifically target AI. In 2018 a group of engineers at Google stated that they did not want to participate in Project Maven. This was a programme for the US military. The aim was to create drones with advanced image recognition capabilities that would be able to automatically recognize people and objects. So many of its employees objected that Google was compelled to terminate this collaboration with the US Department of Defense. Later that year Google engineers signed a petition against Dragonfly, a censored search engine. The company had developed this for use in China, in the hope of gaining a foothold in that country. The workers refused to be party to oppression and this project, too, was subsequently discontinued. In 2019 Microsoft employees sent an open letter that publicly opposed bidding for tenders for the Jedi project – a cloud computing venture for the US military that involved augmented reality equipment – because they did not want to profit from war.

Companies that supply AI technology to Immigration and Customs Enforcement (ICE, the US border security agency) have also encountered resistance from their staff. In 2018 employees at Palantir, Salesforce, Microsoft, Accenture, Google, GitHub and Tableau signed petitions and open letters against working for that organization. In a letter published in *The New York Times*, Microsoft employees called on their CEO, Satya Nadella, to "take an ethical stance and to place children and

[25] Kuijper et al., 31 October 2018.

[26] Palmer, 12 February 2021.

[27] Frenken & Fuenfenschilling, 2020.

[28] In February 2021, the Amsterdam Court of Appeal confirmed an earlier court ruling that Deliveroo delivery riders have an employment contract (Gerechtshof Amsterdam, 16 February 2021).

families above profit".[29] Nadella responded by speaking out against President Trump's immigration policy. Chef Robotics was another company that worked for ICE. When he discovered that this company was using his code, programmer Seth Vargo removed it from online libraries, forcing the company to suspend its service for several days. Chef Robotics eventually terminated its co-operation with ICE.

When Google fired Timnit Gebru it became apparent that even single individuals can refuse to co-operate. Gebru had been a member of the firm's Ethical Artificial Intelligence team. Her scientific work focused on bias and data mining. She wanted to publish a paper on bias in language models, but her employer objected. The company asked her either not to publish the paper or to remove the names of any Google employees. When she refused, she was fired. That sparked outrage. Thousands of company employees, scientists and civil society parties signed a letter denouncing her dismissal. Members of the US Congress asked Google to explain its actions.

Technology businesses depend on talented personnel, so these individuals have sufficient leverage to influence company policy. That applies even to potential future employees. Stanford University is a prestigious institute in the field of AI. In 2018 its students voted to have no dealings with Google until the company shut down Project Maven. They also held on-campus protests against recruitment by companies that support border controls and police activities. More than 1200 students from seventeen campuses signed a pledge never to work for Palantir because of that company's ties to ICE. In addition, students at Central Michigan University opposed the creation of a university Army AI Task Force.

One of the key ways in which parties can make their voices heard is engagement in the form of a walkout, at least in the initial phase of a technology's societal embedding. The number of AI applications is growing rapidly, and employees, students and platform workers are on the front line, as it were. If these individuals take action based on their knowledge of developments, they can play a key part in spotlighting any questionable uses of AI. Walkouts are deemed successful if they receive legal endorsement, for example. In cases such as these, the workers' engagement has a corrective effect.

A more institutionalized form of protest is when unions call for a strike. The right to strike is enshrined in the European Social Charter (Art. 6, paragraph 4). It also derives from the freedoms of association and assembly. A few years ago, a number of Deliveroo delivery riders in the Netherlands went on strike for better working conditions. They were supported by the Riders' Union, part of the FNV (the largest Dutch trade union federation). The pressure exerted by this and other campaigns helped place the working conditions of platform workers on the national political agenda. Those involved are now making every effort to create better legal protection for these workers.[30]

[29] Frenkel, 19 June 2018.

[30] Houwerzijl, 2018.

7.1.3 Protest: Campaigning for a Ban

Protest is the third form of resistance against a given technology or a particular application it can be put to. This approach is particularly common in democratic societies. In such cases people mobilize peacefully to call upon the authorities to, say, impose some kind of ban. Unlike the walkout, protest is not organized from within, nor is it aimed at a particular company and its policy. Instead, it focuses on government and often has a broader base in civil society. A case in point is the anti-nuclear energy movement, which used peaceful protests to call on government to stop building atomic power plants. Similarly, people have staged all kinds of protests against military technologies such as chemical weapons and cluster bombs. In many instances broad-based civil society movements like this have ultimately led to international treaties banning certain weapons.

With regard to AI, protest is one of the most conspicuous forms of engagement. This is especially true of three specific applications: its use by the police for surveillance and prediction, facial recognition and autonomous weapons. One particularly prominent movement against the use of AI in law enforcement was launched in Los Angeles a few years ago. A community group filed a lawsuit to ban the city's police department from using its LASER predictive policing system. This group called itself the Stop LAPD Spying Coalition. It argued that the police were using unjust means – involving proxy data – to predict crime. In doing so they discriminated against people from the Latino and African American communities. University of California Los Angeles (UCLA) students also joined this movement. In its support they highlighted the results of a UCLA study into PredPol (a predictive policing program developed by the university) showing that tools of this kind trigger excessive levels of police activity in communities of colour.

Residents of St Louis, Missouri, also demonstrated against police technology, and in particular a collaborative venture between their city's police and a company called Predictive Surveillance Systems. This uses surveillance aircraft or drones to gather images of members of the public. The residents took to the streets claiming that such "suspicionless tracking" would constitute a massive invasion of their privacy. In the Netherlands Amnesty called for the police's Sensing project at a shopping centre in the town of Roermond to be halted. This used smart cameras to combat 'mobile banditry' – defined by the EU as "...an association of criminals who systematically enrich themselves by perpetrating crimes against property or fraud, (in particular shop and cargo theft, break-ins of homes and companies, fraud, skimming and pickpocketing), within a widespread area in which they carry out activities and are internationally active" – but according to Amnesty this involved the use of mass surveillance and discrimination against certain groups based on their nationality.[31]

[31] Amnesty International, 2020.

A second AI application, facial recognition, has also triggered a great deal of protest. Cameras are increasingly being equipped with this form of computer vision, enabling specific individuals to be monitored with great precision. Concerned citizens see this as a tool for totalitarian surveillance. That has led some people to push for a ban on all use of facial recognition. Others emphasize that government bodies in particular should avoid its adoption. Others still want to impose very strict requirements on its use, such as prohibiting the storage of data or restricting its deployment to, say, searches for missing children. Besides more general concerns about surveillance, some people worry that this technology will be ineffective in the case of individuals from minority groups. They believe that its main effect could be to aggravate the oppression of those communities.

Civil society organizations throughout the world have demonstrated against facial recognition (see also Box 7.3). San Francisco has Stop Secret Spy Tech and Face Surveillance. Successful protests along the same lines have been held in several other American cities, too. Police in San Francisco and Boston are now banned from using this technology. In Portland, Oregon, any use whatsoever is prohibited. Other American movements protesting against facial recognition include Why ID, the Electronic Frontier Alliance and Public Voice. Their European counterparts include Privacy International in the UK and Techno Police in France.

Box 7.3: A Ban on Facial Recognition?

During the preparatory phase of the draft European AI Act, numerous parties called for it to include a ban on facial recognition technology. Amongst them were dozens of civil society organizations.[32] More than 60 MEPs and 50,000 EU citizens also backed the campaign.[33] They had two main demands.

1. A ban on the indiscriminate or arbitrary use of biometric identification in public or in publicly accessible areas, which could lead to mass surveillance.
2. Legal restrictions or hard limits on uses that endanger fundamental rights, such as AI applications for border control, predictive policing, access to social security systems and risk assessments in the context of criminal law.

The call appears to have had some effect. The final version of the draft act prohibits such uses as ‘social scoring’ and the deployment of biometric identification systems in public spaces. This is because, in the European Commission’s view, they pose an ‘unacceptable risk’ to European values.

[32] A campaign entitled ‘Reclaim Your Face’.

[33] Reclaim Your Face, 16 April 2021.

A third AI-related topic to attract protest is autonomous weapons. Movements throughout the world are calling for these to be prohibited. The Campaign to Stop Killer Robots was founded in 2012. This coalition of non-governmental organizations is committed to a ban on fully autonomous weapons and to upholding 'meaningful human control' over the use of force. In 2015 more than a thousand AI experts, including Stephen Hawkins, Elon Musk, Steve Wozniak and Noam Chomsky, signed an open letter warning of an AI arms race and calling for autonomous weapons to be outlawed. In 2017 a similar letter calling for a ban on lethal autonomous weapons was submitted to the United Nations. This was signed by 166 robotics pioneers and by the directors of several technology companies. The Dutch government received a similar exhortation at the end of 2020; more than 150 scientists active in the fields of robotics and AI asked it to support a ban on lethal autonomous weapons. Protest – in the sense of being able to speak out against something – is thus an important form of engagement in a democratic society. With regard to AI, this approach is already highly developed and will continue to play a prominent role.

Key Points – Resistance: Fight, Walkout, Protest

- New technologies often provoke resistance, especially in cases of rapid technological change and where people become convinced that a technology will benefit only a very limited sections of society, while its risks are widespread. Resistance expresses an antagonistic attitude There are three different forms: fight, walkout, and protest.
- In the past groups opposed to new technology regularly resorted to violence in their *fight* against it. AI has not yet been associated with this form of resistance, a questionable aspect of engagement that does not need reinforcement. Democratic engagement with AI is preferable.
- People who engage in *walkouts* are refusing to co-operate with AI in various ways. One option is work stoppages, literal 'walkouts', where pressure from within compels companies to change course. This form of engagement has grown in recent years. It is typical of the initial phase of AI.
- In democratic societies *protest* is a highly developed and significant form of resistance. Here people mobilize peacefully to call upon the authorities to, say, impose a ban on something. This is currently one of the most prominent forms of engagement regarding AI, targeting three of its applications in particular: its use by the police for surveillance and prediction, facial recognition and autonomous weapons.

7.2 Monitoring

The next two forms of engagement we have identified fall under the collective heading 'monitoring'. This cluster occupies the central part of the spectrum between the antagonistic forms discussed above and the symbiotic ones we examine in the next section. The two forms that count as monitoring are supervision and agenda-setting.

Both subject actions by other parties –public and private alike – to critical scrutiny. Where necessary these actions are corrected and adjusted in line with alternative proposals. This approach is in line with the historical trend signalled by John Keane in his book *The Life and Death of Democracy*. Basically, he argues that many hundreds of new types of institutions came into being after 1945 to track the actions of influential parties and subject them to intense scrutiny.[34] He characterizes this development as 'monitory democracy' and refers to options such as the use of surveys, online petitions and focus groups, but also to self-appointed watchdogs and NGOs committed to weaker or underrepresented groups in society.[35] Building on this basis, we take 'supervision' to mean co-ordinating stakeholders to address malpractices in the use of a new technology. We refer to the fifth and slightly more neutral form of engagement as 'agenda-setting'. This involves civil society parties who identify both positive and negative aspects of the technology, but are dedicated primarily to turning a public spotlight on the theme.

7.2.1 Supervision: Reporting Malpractices

Supervision has a different goal from the forms of engagement discussed above. It is not so much about preventing specific uses of AI by banning them as about correcting the applications themselves or the conditions under which they operate. Specific parties could be informed, say, or public campaigns conducted. Alternatively, people could bring lawsuits or submit notifications to regulators to address malpractices. In practice a critical benchmark here is the matter of rights – human rights first and foremost. Civil society parties assess the nature of AI applications to determine whether these are legally permitted. This is a key feature of supervision.

There is some uncertainty about the impact of AI at this early stage of its integration into society. Lawsuits play an important part in dealing with these 'grey areas'; they enable any malpractices to be identified and case law to be developed. In this way, directly disadvantaged groups can be protected by restoring their rights. The development of the law also benefits. Accordingly, jurisprudence can also spotlight issues and provide guidance. It helps people understand what is happening in the field and gives them the clarity needed to respond appropriately. It also helps create a framework for further applications, and possibly, future legislation as well. History shows that lawsuits challenging abuses by railway companies and telegraph services have specifically served this purpose.[36]

In addition, there are situations in which AI applications must be subjected to mandatory 'assessments'. This is because stakeholders' views concerning the implementation of new technological capabilities need to be heard. In the Netherlands this applies to the statutory remit of works councils: their right to

[34] Keane, 2009.

[35] Keane, 2011.

[36] Pasquale, 2015: 90–91.

consultation (concerning investments and so on), right of consent and right to be informed.[37]

Supervisory activities can be undertaken by interest groups, experts or the media. This growing form of engagement is also important, even if it does not usually attract quite so much public attention as protests. Its expansion is due mainly to the relatively recent emergence of organizations linked to the use of AI in society and, more broadly, to the rise of digitalization. It is also due to the fact that the Netherlands has quite recently broadened the legal scope for collective action in judicial proceedings.[38] That change has facilitated this form of engagement. Nevertheless, people often express concern that the use of AI is at odds with the current legal protections available to victims. This is because those safeguards are organized at the individual level whereas AI applications categorize people into group profiles.[39]

In this section we first discuss a number of national and international organizations that are helping to monitor the use of AI, mainly through knowledge sharing. Then we review a number of prominent lawsuits.

Various international organizations have taken on a supervisory role by disseminating knowledge about the use of AI. In 2017 Kate Crawford and Meredith Whittaker founded the AI Now Institute in New York. It issues reports and analyses that focus on AI's impact in four areas: rights and liberties, labour and automation, bias and inclusion and safety and critical infrastructure. For instance, it has raised the issue of malpractices associated with poor working conditions in the technology sector (including at Amazon's warehouses). It has also spotlighted AI systems' ecological footprint, something that generally receives little attention. In its annual report the institute describes the current state of affairs regarding the use of AI. It then goes on to make recommendations concerning the further development of AI in society.

Another such organization was the Google Transparency Project, launched in 2016. Its goal was to conduct research and analysis that would shed light on the ways in which Google influences government and policy. Under the new name Tech Transparency Project, the organization is now focusing more broadly on the technology sector. It acts a non-profit watchdog pursuing corporate accountability through investigations, litigation and the disclosure of misconduct.

In Europe, Germany's AlgorithmWatch is one of the key players. This non-profit organization focuses on algorithmic decision-making processes that have a social impact. These include algorithms that predict or direct people's behaviour or are designed to make automatic decisions. Its approach is to analyse the ways in which algorithmic decision-making influence human behaviour. In this context it explains to the general public how decision-making works, brings experts together and develops ideas and strategies for the beneficial use of AI in society. AlgorithmWatch's

[37] Moreover, the importance of this role is reflected by the Social and Economic Council of the Netherlands' (SER) guideline informing works councils about their role regarding technological developments (SER, 2016).

[38] van Boom & Weber, 2017. This concerns the Dutch Class Action (Financial Settlement) Act.

[39] Kosta, 2020; van der Sloot & van Schendel, 2020; Taylor et al., 2016.

annual *Automating Society Report* tracks the use of automated decision-making in Europe. It also publishes sub-studies on topics such as the use of algorithmic decision-making in response to the COVID-19 crisis. The organization identifies ethical dilemmas and puts forward proposals for more responsible use of algorithms.

Such bodies contribute to the supervision of new technologies by providing knowledge. In the Netherlands the digital rights organization Bits of Freedom showed that it was possible for people in one country to place advertisements on Facebook in another country – during elections in the latter, for example. The organization revealed just how easily Dutch users could upload German memes (and hence ideas) conveying party political messages.[40] This contradicted testimony to the Dutch House of Representatives from senior Facebook staff.

Another way to deal with malpractices is to take such matters to court. In a further example from the Netherlands, a coalition of civil society organizations and individuals sought an injunction against the Dutch state to ban the use of System Risk Indication (SyRI) (Box 7.4). Various local authorities, working in co-operation with the Ministry of Social Affairs and Employment (SZW), had been using this tool for purposes such as detecting cases of benefit fraud.

The coalition stated that SyRI involved unlawful automated decision-making. The plaintiffs also argued that SyRI was used mainly in areas already labelled 'problem neighbourhoods'. As a result, the system had a discriminatory and stigmatizing effect. In reaching its verdict, the court first considered the nature of SyRI itself. It took the view that this tool was in line with forms of AI such as deep learning and self-learning systems. Given that SyRI uses risk profiles, the court felt that this could lead (unintentionally) to biased connections being made. They could be based on lower socio-economic status or migrant background, for example. This would mean that SyRI has a disproportionately large impact on poor people. According to the court, the infringement of privacy that results from the state's use of this system is out of all proportion to the importance of detecting benefit fraud. Following this verdict, SyRI was discontinued.

Box 7.4: System Risk Indication

System Risk Indication (SyRI) is a technical application that calculates the probability of a particular individual fraudulently claiming social benefits. To do this it links seventeen different types of data. The government states that SyRI compares files of existing factual data from sources such as the Employee Insurance Agency (UWV), the Social Insurance Bank (SVB), local authorities, the Dutch tax authorities and the Netherlands Labour Authority. It then checks for discrepancies between the information garnered from these sources. If this comparison and an assessment against the risk model reveal any irregularities, these must first be investigated by one or more of the above organisations. Only then can a decision be taken that might have legal consequences for the individual concerned.

[40] Austin, 20 May 2019.

This is not just about government, though. Lawsuits concerning the use of AI have been instigated against other parties as well. Private companies have been sued for all kinds of malpractice. In the UK Uber was sued because its facial recognition system failed to effectively identify drivers and couriers of colour.[41] In a case in the Netherlands the court ruled in favour of Uber by finding that the company had lawfully used algorithms to determine which drivers had or had not been fired.[42]

Incidentally, it can be quite challenging for civil society parties to bring cases to court. Those concerned must have the resources and knowledge needed to raise the issue of malpractices and take action. Moreover, more traditional interest groups (many of which originated in the analogue era) are usually unaware of how AI is changing their field of work.[43] Consequently, they do not yet have a sufficient grasp of how AI can marginalize the groups they represent or jeopardize their interests. Take consumers. In economic transactions with companies, they are viewed as the weak party. Accordingly, they are afforded legal protection in Europe. The use of AI systems can have an impact on their autonomy because it is algorithms, not the buyers themselves, that search for the ideal purchase based on an identified need.[44] Many consumers are ignorant about the underlying workings of AI systems. As a result, companies can persuade them to make purchases that are not in their own interest, perhaps because they are more expensive. This could blur the distinction between personalized offers and manipulation.[45] So both legislators and consumer organizations need to understand developments of this kind and, if necessary, take a stand against them.

The lawsuit against SyRI in the Netherlands was unusual in that it was driven by a broad coalition. These included some traditional interest groups as well as experts in the field of law and digital technology. Alliances like this are very helpful inasmuch as they fulfil civil society's supervisory role quite effectively. Organizations less familiar with AI and the problems associated with digitalization can access the expertise of others that do possess the requisite knowledge. This expertise is not restricted to digital rights groups. Human rights organizations are also increasingly acquiring knowledge and expertise in this domain, and developing it further.[46] Moreover, both human rights and digital rights organizations are part of larger international networks where AI has long been on the agenda.

Such organizations are incentivized to initiate joint lawsuits by a type of procedure known as public interest litigation.[47] In these cases the organizations involved must be able to demonstrate that rules or policies directly impact the public interest in general or the particular collective interests they represent. This form of litigation

[41] The Guardian, 2021.

[42] See Gerechtshof Amsterdam, 11 March 2021.

[43] Steijns, 2021.

[44] Fierens et al., 2021: 974–975.

[45] Fierens et al., 2021: 969.

[46] Steijns, 2021.

[47] Braun & Stolk, 2020.

is not yet routinely used everywhere because not all legal systems are receptive to it. At the same time lawyers are pushing for innovation in this area, especially with a view to advances in digitalization. They point to Germany, for example, where competitors can hold each other accountable for compliance through the courts.[48]

7.2.2 Agenda-Setting: Information About the Importance of AI

The next form of engagement with AI we identify is positioned slightly more towards the right-hand end of the spectrum from antagonism to symbiosis. Parties here are committed to generating more attention for AI because they believe that that is important in itself, whether it focuses on the positive or negative aspects. Various civil society organizations are helping to place AI on the agenda. Some specifically focus on drawing attention to new technology, but thought leaders and artists also play their part in this respect. Moreover, they use a wide range of platforms for this purpose. In addition to artistic events and reports from think tanks, we discuss the ways in which civil society parties are involved in the development of policy and legislation for AI. We also spotlight the interests they represent.

In addition to supervising AI, many of the abovementioned international organizations like AI Now and AlgorithmWatch also publish reports and stage events on this topic. The artist Trevor Paglen has created ImageNet Roulette, an app people can use to upload photographs of their faces to see how they are ranked by the influential ImageNet database.[49] The Dutch organization Waag has made an especially outstanding contribution in this area; with its origins in the hacker movement and the early rollout of the internet in the Netherlands, its goal is to achieve the open, fair and inclusive use of digital technology. In particular, it defends public values and interests against the influence of commercial logic. People can also use art projects to raise awareness about AI. We discuss two Dutch examples of this approach in Box 7.5.

Box 7.5: Agenda-Setting Through Art

We Are Data

This artists' collective is helping to create a more profound awareness of the types of personal information that can be recorded in databases. The idea is that, by experiencing this phenomenon at first hand, you gain a better idea of the impact of various technologies – old and new. To this end We Are Data has developed a 'mirror room'. Visitors enter an enclosed space one at a time.

(continued)

[48] Moerel & Prins, 2016a, b. 116; recently also Barkhuysen, 2021.

[49] Crawford, 2021: 141.

Box 7.5 continued
There they are subjected to an impressive and very personal experience. It is also a smart space in which the visitor is surreptitiously observed and measured. They thus find out what it is like to be processed into data and can decide which of their personal information remains their private property. In this way they are literally held up to a mirror.

Wouter Moraal

Moraal aims to inform people about how deep-learning algorithms work. He also wants to warn them about the potential repercussions of misusing algorithms. To do this he has developed Artificial Impact, a board game modelled on a deep-learning algorithm.

The players first have to train the algorithm. At the end of the game their performance is rated by their own creation – a self-taught risk-prediction algorithm. The assessments made during the game are based on situations in which AI was used without due care and attention, leading to malpractices. The project therefore has much in common with Monopoly, the board game developed by Lizzie Magie in 1903. She wanted to make people aware of the harmful consequences of people owning huge estates and of capitalist exploitation.

Various aspects of AI need to be placed on the political agenda. This is a particularly important aspect of political decision-making processes. In a representative democracy, elected representatives have the final say in political decisions. However, they are still accountable to voters and to society at large. In practice various more or less optional processes have been set up for this purpose. Civil society parties can independently present their views to the legislature or to ministries working on specific policy and/or legislative proposals. These, after all, increasingly concern AI and its use in various sectors. Here we discuss the engagement of civil society parties with the European Commission's draft AI Act.

The European Commission published this document on 21 April 2021.[50] Civil society parties were also involved at various stages throughout its development. First, they participated in the High-Level Expert Group on Artificial Intelligence (AI HLEG), founded in 2018. Through this forum 52 experts advised the European Commission on the implementation of its AI strategy, details of which were published on 7 December 2018.[51] The expert group consisted of 18 academics, 37 business representatives and four representatives of civil society. The AI HLEG presented its final Assessment List for Trustworthy Artificial Intelligence (ALTAI)

[50] European Commission, 2021.

[51] European Commission, 7 December 2018.

on 17 July 2020.[52] Even before it had completed its meetings, however, civil society organizations were accusing the industry of unduly influencing that list. In particular, they claimed that the sector had blocked a number of proposals to ban some forms of AI.[53] Another relevant point of criticism was that those parties with practical knowledge and ties with groups in society who had to deal with AI systems were not being properly heard. Michael Veale, a British researcher in the field of digital rights, says that these 'low-level experts' are the very people who will have to deal with the ethical considerations when AI applications are implemented.[54] He also states that there is a much greater need for such practice-based experts than for the professors of applied ethics advocated by the AI HLEG.

Civil society was also involved through the Alliance for Artificial Intelligence, which provides a platform for approximately 4000 stakeholders. Its initial purpose was to provide feedback to the AI HLEG. Over time however, the alliance has become a benchmark for stakeholder-driven discussions about AI policy.

Finally, several civil society parties participated in the public consultation that preceded the publication on 19 February 2020 of the White Paper entitled *On Artificial Intelligence – A European approach to excellence and trust*. EU Member States contributed 84% of the content, with the remainder coming from other parts of the world. Civil society actors were responsible for 13% of contributions.[55] Many of them felt that the Commission should have done much more to safeguard human rights, especially regarding the use of facial recognition. A case in point was when dozens of civil society organizations jointly appealed to the European Commission to ban certain forms and uses of AI (see Box 7.3).[56]

The European Commission consulted parties across the board, including civil society actors, to give them an opportunity to present their views on the white paper. As we have contended in the introduction to this chapter, while issues important to specific groups in society need to be put on the agenda, this is not the sole responsibility of civil society parties. It is primarily government's duty to represent numerous different interests as far as possible. It therefore needs to develop a vision that encompasses the full range of views concerning the integration of AI into society. Government also needs to understand the technology in terms of its potential implications for different groups of stakeholders. In the case of political decisions, one aspect of this task concerns formal stipulations to consult stakeholders or to allow participation in political decision-making.

A broader and more structured process of this kind should have taken place during the preparatory phase of the draft AI Act. For example, the existing internet consultation mechanism could have been used to reach groups that operate below

[52] High-Level Expert Group on Artificial Intelligence, 2020.

[53] In particular, see Access Now, 2020: 16–18.

[54] Veale, 2020.

[55] European Commission, undated (b).

[56] Reinhold, 22 April 2021.

Key Points – Monitoring: Supervision, Agenda-Setting

- Monitoring subjects the actions of public and private parties to critical checks. Where necessary these actions are corrected and adjusted in line with alternative proposals. We have identified two forms of this type of engagement: supervision and agenda-setting. These occupy the middle ground between antagonistic forms of engagement on the one hand and more symbiotic forms on the other.
- Supervision involves correcting AI applications themselves or the conditions under which they operate. For instance, specific parties could be informed, or public campaigns could be conducted. Alternatively, people could bring lawsuits or submit reports to regulators to address malpractices. In practice one critical benchmark here is the matter of rights (human rights first and foremost). This provides insight into the impact of AI at an early stage.
- In agenda-setting, civil society parties, opinion formers and artists commit themselves to spotlighting certain aspects of AI and its use. Despite its undoubted importance, this form of engagement is often underdeveloped. In addition, it is rather like preaching to the converted.
- Agenda-setting is another important aspect of political decision-making processes, at both national and international levels. Here it is essential for government bodies to approach a broad spectrum of civil society groups. Weaker or vulnerable groups often experience the adverse impacts of AI systems.

the radar of government bodies. On the one hand AI is an early-stage systems technology. Understandably therefore, it is not immediately clear to government which groups should be involved in plans to manage its impact. On the other hand, the individuals with experience in everyday practice are the very people who, by definition, are in a position to identify the challenges that will arise as AI integrates into society. This is borne out by recent evidence that weaker or vulnerable groups tend to be affected by the adverse impacts of AI systems. We therefore recommend that government bodies actively and formally involve a broad spectrum of civil society groups in the process of formulating AI policy.[57]

[57] European Center for Not-for-Profit Law, 2020: 7. In the Netherlands the government is not obliged to submit parliamentary bills, draft orders in council or draft ministerial regulations to interested parties for their approval. But it is required to consider doing so. In 2007 the Dutch government expressed an aspiration to strengthen its dialogue and consultations with civil society, using channels such as the internet.

7.3 Co-operation

Our third and final cluster of forms of engagement comes under the heading ‘co-operation’. First and foremost, this entails a commitment to improving the technology. That could include civil society parties that draw up principles of good practice or are involved in standardization processes. Co-operation also includes appropriating the new technology, whereby parties incorporate it into their existing activities and use it to achieve their own goals and values. People co-operate for a variety of reasons, some related to the particular nature of AI.

7.3.1 Improving: Knowledge of Good Practice

Improving AI is positioned towards the ‘symbiosis’ end of the engagement spectrum. Involved here are those who work in the field itself or possess related know-how or other relevant expertise. They work with AI because they are convinced that the technology will enrich society. They are prompted to mobilize by the desire to put their expertise on the subject to good use, with the aim of improving AI and its application. Some might draft principles, while others write open letters and others still develop instruments for good AI practices (toolkits) or other types of publication. Many of these initiatives take place at the international level. Institutions with regulatory powers are actively involved in drawing up principles or standards; they include the EU, the UN and various standardization bodies. Here however, we confine ourselves to the bottom-up initiatives launched by various civil society parties. These include professional organizations, academic institutions and non-profit organizations. In Box 7.6, we describe one example, the Dutch ALLAI initiative, which focuses on developing responsible AI through research and collaborative projects.

An AI security conference was held in Puerto Rico in 2015. The participants issued an open letter stressing the importance of broadening AI research. This was based on the notion that AI was conceived ‘in a lab’. The participants stated that ethicists, philosophers, economists, legal scholars and cybersecurity researchers should be more engaged with the interdisciplinary research agenda.[58]

In 2017 the Future of Life Institute hosted the Asilomar Conference on Beneficial AI. A hundred people, including AI scientists, economists, philosophers and lawyers as well as politicians, joined forces to develop 23 principles for ‘beneficial AI’. These are divided into questions for research into AI, ethics and values and long-term issues.[59] Several prominent researchers attended the conference. The list of principles was signed by such eminent figures as Elon Musk, Nick Bostrom, Demis Hassabis, Yann LeCun, Yoshua Bengi, and Stuart Russell.

[58] Future of Life Institute, undated (b).

[59] Future of Life Institute, undated (c).

Box 7.6: ALLAI
The Alliance for Artificial Intelligence Netherlands (ALLAI) was launched at the World Summit AI in 2018.[60] This made the Netherlands the first European country to have an independent organization dedicated entirely to the responsible use of AI. Amongst other things, ALLAI focuses on developing ethical preconditions for AI through projects, research, policy advice and education. Basing its approach on 'responsible AI', it aspires to create national and international environments that will deliver the benefits of artificial intelligence while at the same time safeguarding civic values such as security, autonomy and inclusion. To this end alliance founders Catelijne Muller, Virginia Dignum and Aimee van Wynsberghe (all former members of the AI HLEG) encourage stakeholders across the field to co-operate. They also make every effort to involve policymakers, scientists, entrepreneurs, lawyers and consumers in their projects. Since the outbreak of the COVID-19 crisis the organization has been exploring options for the responsible use of AI in tackling the pandemic. In this domain it is working with policymakers and researchers.

Meanwhile, a team at the University of Montreal has developed a set of ethical principles for responsible AI. This group of ethicists, legal scholars, public administrators and AI experts prepared a draft proposal listing seven principles. Five hundred academics, members of the public and stakeholders were mobilized to respond to that in writing and at meetings. The goals were to establish frameworks for the development and application of AI, to create principles that enable everyone to benefit from it and to facilitate the debate on equity-oriented, inclusive and sustainable AI. This process culminated in the Montreal Declaration for the Responsible Development of Artificial Intelligence. In another effort to improve use of the technology, the AI Now Institute has been developing an Algorithmic Accountability Policy Toolkit and Algorithmic Impact Assessments.

The Partnership on AI is yet another prominent body dedicated to improving AI. Its members include large companies like Amazon, Facebook, Google, DeepMind, Microsoft and IBM, as well as China's Baidu. The partnership itself is a non-profit organization committed to the responsible use of AI, its approach being to identify good practices and share knowledge.[61] For example, it has developed a database of AI incidents involving autonomous vehicles or so-called 'flash crashes' on stock exchanges.

Finally, there is the organization OpenAI. This partly for-profit venture has originated products such as GPT-3, the AI program that wrote the article in *The Guardian* mentioned at the very start of this report. Its activities also include non-profit research aimed at developing 'friendly AI'. OpenAI has received significant funding from Elon Musk and Microsoft.

[60] Alliance for Artificial Intelligence, undated.

[61] Russell, 2019: 250.

Public courses are another form of engagement based on co-operation. 'Elements of AI' was the first of these, a series of lessons intended to give people a basic understanding of the topic. It was developed by the University of Helsinki in co-operation with Reaktor, a technology company, and originally funded by the Finnish government. 'Elements of AI' is now backed by the European Commission and available in dozens of languages. More than 750,000 people have taken these courses.

AI is playing an increasingly important part in everyday life. Yet people still have a lot of mistaken ideas about what the technology is and what it can do. Courses like Elements of AI are designed to provide information in an accessible way to anyone wanting to find out more about the subject. Also included in this category of tools are impact assessments that can be used to identify the effects of using AI. These forms of engagement are based on improving technology and the ways in which we make use of it. The momentum behind them is growing, and they will become increasingly important as AI becomes more deeply embedded in our society.

7.3.2 Appropriating: Diversity in Goals and Interests

Our final – and most symbiotic – form of engagement is appropriating AI. Whereas improving AI is about working on good practices and its lawful future use, its appropriation means civil society parties actually adopt it. The business community and government bodies have the resources to put new system technologies into practice, so they are usually the first to do so. Civil society parties usually take much longer to follow their example. These are mainly groups of individuals and professional organizations. Here we discuss a number of initiatives that would enable these latter two groups to appropriate AI.

Several projects have been launched to assist social groups disadvantaged by AI (see paragraph 6.2). These mainly involve the critical monitoring and assessment of its use by companies and government bodies. As a more extreme option, AI itself can be used to represent the interests of those groups. Ruha Benjamin stresses the importance of community-wide technology use to counteract any exclusive effects. She explores the democratization of data, citing initiatives such as DiscoTech ('Discovering Technology'). These make technology accessible in ways that allow particular groups to appropriate it in practice.[62]

The Mijente group describes itself as a 'political home base' for Latino Americans and Mexicans. Its projects include identifying the relationship between AI and immigration. MediaJustice is a US organization that champions people of colour and those on lower incomes. It is working to achieve a fair economy, connected communities and a political landscape in which these groups are not only visible but have a voice and power. Its founders say that to achieve this we need a media and technology environment able to sustain real justice. Numerous organizations are

[62] Benjamin, 2019: 188–189.

currently spotlighting the interests of minorities. These include Women in AI, Black in AI (co-founded by Timnit Gebru, who was fired by Google) and Queer in AI.

In the Netherlands appropriation takes place in AI labs and numerous other places. Many of these are working on the application of AI by companies or by government. But civil society parties are also becoming involved. The Civic AI Lab is one example. This collaborative venture between the University of Amsterdam, VU Amsterdam university and the City of Amsterdam was established in 2021. Scientists at Tilburg University are co-operating with partners such as Greenpeace, the World Food Programme and the Jeroen Bosch Hospital to use AI for public-interest tasks in the fields of climate, food shortages and healthcare.[63] A final example is The Hague Institute for Innovation of Law (HiiL), co-founded by the legal scholar Maurits Barendrecht. This has been using a so-called 'justice accelerator' to launch various projects involving the use of AI, mainly in Africa.[64]

So, there is already a great deal of activity in the field of appropriation, although civil society parties still seem rather slow in grasping the opportunities presented by AI. As yet, organizations representing more traditional interest groups like tenants, patients, consumers and teachers do not seem to be very active in this domain. This is partly because we are still just starting the process of embedding AI in our society. Accordingly, groups that defend public values by appropriating AI (which helps to shape that process) are only now emerging. Many of these still have a limited understanding of the technology, let alone ideas about how to use it for their own purposes. At the same time, it is important that they not be left behind, as their grassroots have much to gain from AI (Box 7.7).

Box 7.7: PublicSpaces

While it does not focus specifically on AI, the Dutch PublicSpaces coalition is a great example of civil society appropriating digital technology. This is a collaborative venture by more than twenty parties from the public media, cultural heritage, festivals, museums, education and healthcare sectors.

The coalition was created to reimagine the internet as a public space and revive its founding principles. PublicSpaces is campaigning against our reliance on big tech companies for communications, information and media circulation. Its goal is an alternative software ecosystem that revolves around public values rather than commercial interests.

In that context the organization is developing tools such as 'public badges'. These are quality labels for the coding and tooling of websites and software applications based on the values espoused by PublicSpaces. It is also working to implement open-source initiatives.

[63] See Tilburg University, undated; Data Science Center Tilburg, undated

[64] HiiL, undated.

Besides communities and specific sections of the general population, appropriation is also important for professional groups. This is an enormous field and AI is triggering workplace changes in all kinds of occupational arenas (see Chap. 6). Here we focus on the interests and values embodied by certain professions in particular, such as doctors, teachers and lawyers. They possess specific expertise derived from their educational background and work experience, which we need to safeguard when AI systems are introduced. In other words, these groups need to appropriate AI in a way that gives their students a good education, enhances their patients' health or safeguards the rights of their clients.

Many people claim that AI applications can replace this kind of expertise. Some assert that robot judges or medical algorithms can render traditional professions superfluous. As we have seen, misconceptions like this are typical of antagonistic relationships between AI and society. Whereas AI's increasing integration into society in fact requires a symbiotic relationship. That means combining it with human professional expertise. Frank Pasquale has shown that rather than undermining expertise in general, AI in its current form actually tends to place more emphasis on some types than others. The skills possessed by computer scientists and economists is central to many AI applications. As things stand these are taking precedence over other forms of know-how.[65] Consequently, applications of this kind are based solely on a single, simple criterion. This is in stark contrast with the real-world situation, which involves a complex web of standards, goals, interests and knowledge from all kinds of professional groups. Algorithms that write articles may perhaps be able to simulate part of a journalist's work, but they do not come close to replacing every aspect of their day-to-day responsibilities. These include considering different perspectives, treating people equitably and conducting in-depth research.

During the initial phase of AI's entry into society, people tended to focus mainly on its revolutionary nature. However, they have since become increasingly concerned about the jobs that would be rendered obsolete by this technology. In the next phase it is vital for all kinds of professional groups to appropriate AI based on their responsibilities as professionals. Such groups are subject to various forms of self-regulation. They also have regulatory bodies that issue licences and monitor practices. These provisions need to include the use of AI in their field of work.[66] Before that can happen, professionals need to master the technology and understand exactly how it can contribute towards their everyday work.

[65] Pasquale, 2020: 23.

[66] Pasquale, 2020: 88.

Key Points – Co-operation: Improving, Appropriating

- Co-operation involves a symbiotic attitude towards AI. It encompasses commitment to improving the technology and to appropriating it to achieve your own goals and values.
- Improving AI involves people who work in the field or possess related know-how or other relevant expertise. Their efforts in this domain are motivated by a belief that the technology will enrich society. They use their expertise in the subject to improve AI and its use. More specifically, some might draft principles while others write open letters or develop instruments for good AI practices (toolkits) or other types of publication.
- As yet, few individuals or groups are involved in appropriating in AI. It is mainly the business community and government bodies that are putting the technology into practice. Civil society parties and professional groups seem to be rather slow in grasping this opportunity. Appropriation is important for a variety of reasons. For example, these parties can use AI to counteract its own exclusionary effects or to safeguard the values they embody in their own professional practice.

7.4 In Conclusion

The overarching task of engagement is all about who should be involved in AI. Companies and government bodies are often the first to use new system technologies. This gives them a huge amount of influence over these technologies' developmental paths. Civil society parties, too, are gradually becoming involved in this process. They can include interest groups, academic institutions, the media and specific professions.

Engagement is important for any society, especially democracies. The process of embedding a new technology responsibly within a society hinges on the interests, values and knowledge of a wide range of actors. This means their voices need to be heard not only during the design process, but also if they are impacted by the use to which that technology is put. When all is said and done, they should be able to use it themselves to achieve their own goals. Or to put it another way, civil society parties provide valuable feedback (based on their own experience and knowledge) for AI. We need to take this into account to ensure that the technology becomes properly integrated into our society.

So far though, there are few formal channels for feedback of this kind. As a result, companies and government bodies are developing all kinds of AI applications without fully understanding how they will impact the lives of individuals and

specific social groups. They are also failing to exploit the knowledge and expertise that such groups could contribute. Teachers and students have a part to play in the development of AI in education, doctors and patients have a part to play in healthcare AI and so on and so forth.

This chapter has focused on engagement with AI. In this respect we have identified a spectrum of different forms, ranging from an antagonistic relationship with AI to a symbiotic one. Some of the antagonistic forms are already highly developed, such as protest and supervision. These efforts are key to preventing the malicious use of AI, and they must be continued. Supervision also plays an important part in spotlighting issues, thereby helping to create frameworks, standards and regulations. The same goes for walkouts. The employees of technology companies are on the front line, so they can identify any problems at an early stage. The most antagonistic form, fighting, is not yet widely used in connection with AI but it can send a clear signal to society.

Some parties adopt a neutral stance when placing AI on the public agenda. At an international level too, people have launched initiatives to improve the technology. Their approach is to develop principles and to share knowledge and experience. Engagement in the form of appropriation is enormously important as well. It enables civil society actors, communities and professional groups in particular to use AI in ways that suit them, helping them to achieve their own goals and safeguard their own values. As yet though, traditional interest groups and professions only have limited capabilities when it comes to appropriating AI.

Progress is being made with neutral monitoring and the symbiotic forms of engagement. But unlike the more antagonistic forms, these are still quite poorly developed. There is also a great deal of activity at the international level. Government's task is to encourage national forms of engagement as a way of more effectively involving civil society in embedding AI. First and foremost, government bodies can do this by augmenting stakeholder expertise. That is, by equipping particular groups of stakeholders with the means to participate in constructively critical forms of engagement. Which is all the more important given the civic values at stake here, or potentially so.

References

Access Now. (2020). *Europe's approach to Artificial Intelligence: How AI strategy is evolving*. Access Now. Available at: https://www.accessnow.org/cms/assets/uploads/2020/12/Europes-approach-to-ai-strategy-is-evolving.pdf

AFM en DNB. (2019). *Artificiële Intelligentie In De Verzekeringssector Een Verkenning*. Autoriteit Financiële Markten, De Nederlandsche Bank. Available at: https://www.afm.nl/~/profmedia/files/rapporten/2019/afm-dnb-verkenning-ai-verzekeringssector.pdf?la=nl-nl

Amnesty International. (2020). *Out of control: Failing Eu Laws for digital surveillance export*. Amnesty International.

Barkhuysen, T. (2021). Handhaving van de AVG: de AP kan het niet alleen. *Nederlands Juristenblad, 572*(8), 585.

Benjamin, R. (2019). *The race after technology: Abolitionist tools for the New Jim code*. Polity Press.

Braun, C., & Stolk, R. (2020, March 4). Procederen Uit Naam Van Het Algemeen Belang. *Montesquieu Instituut*. Available at: https://www.montesquieu-instituut.nl/id/vl6ck1v35y97/nieuws/procederen_uit_naam_van_het_algemeen

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Das, D., de Jong, R., Kool, L., & Gerritsen, M. M. V. J. (2020). *Werken Op Waarde Geschat – Grenzen Aan Digitale Monitoring Op De Werkvloer Door Middel Van Data, Algoritmen En AI*. Rathenau Instituut.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Center for Not-for-Profit Law. (2020). Being aware: Incorporating Civil Society into national strategies on Artificial Intelligence. Country Papers On Participatory Processes In Drafting National Ai Policies In The Czech Republic, The Netherlands, Australia And Canada. ECNPL. Available at: https://ecnl.org/publications/being-ai-ware-incorporating-civil-society-national-strategies-artificial-intelligence

European Commission. (2021 [2018]). *EU coordinated action plan on AI 2021 review*, COM(2021) 205 final. Available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-reviewhttps://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

Fierens, M., van Gool, E., & De Bruyne, J. (2021). De Regulering Van Artificiële Intelligentie (Deel 1) – Een Algemene Stand Van Zaken En Een Analyse Van Enkele Vraagstukken Inzake Consumentenbescherming. *Rechtskundig Weekblad, 84*(25), 962–980.

Frenken, K., & Fuenfenschilling, L. (2020). The rise of online platforms and the Triumph of the Corporation. *Sociologica, 14*(3), 101–113.

High-Level Expert Group on Artificial Intelligence. (2020). *Assessment List For Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Houwerzijl, M. (2018). Juridische Vraagstukken Rond Arbeid In De Klusseneconomie. *Beleid En Maatschappij, 45*(2), 208–216.

Juma, C. (2016). *Innovation and its Enemies: Why people resist new technologies*. Oxford University Press.

Keane, J. (2009). *Life and death of democracy*. Simon & Schuster.

Keane, J. (2011). Monitory democracy. In S. Alonso (red.), *The future of representative democracy* (pp. 212–235). Cambridge University Press.

Kosta, E. (2020). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*. https://doi.org/10.1111/rego.12331

Mateescu, A., & Ngyun, A. (2019). *Explainer: Workplace monitoring En surveillance*. Data en Society Research Institute. Available at: https://datasociety.net/wp-content/uploads/2019/02/DS_Workplace_Monitoring_Surveillance_Explainer.pdf

Moerel, L., & Prins, C. (2016a). Privacy Voor De Homo Digitalis: Proeve Van Een Nieuw Toetsingskader Voor Gegevensbescherming In Het Licht Van Big Data En Internet Of Things. In *Homo Digitalis, Preadviezen 2016 Nederlandse Juristen-Vereniging 2016* (pp. 9–124). Kluwer.

Moerel, L., & Prins, C. (2016b). *Privacy for the Homo digitalis: Proposal for a new regulatory framework for data protection in the light of big data and the Internet of Things*. Wolters Kluwer. Available at: https://doi.org/10.2139/ssrn.2784123

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce Racism*. New York University Press.

Pasquale, F. (2015). *Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

Pasquale, F. (2020). *New Laws of Robotics: Defending human expertise in the age of AI*. Harvard University Press.

Perez, C. C. (2019). *Invisible Women: Exposing data Bias in a world designed for Men*. Vintage.

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.
Schuyt, K. (2006). *Steunberen van de samenleving*. Amsterdam University Press.
SER. (2016). *Mens En Technologie, Samen Aan Het Werk*. Sociaal-Economische Raad. Available at: https://www.ser.nl/-/media/ser/downloads/adviezen/2016/mens-technologie-publieksversie.pdf
Steijns, M. (2021). *"Van Repliek Gediend?" Een Verkenning Van Tegenmacht Vanuit Maatschappelijke Organisaties* (WRR Working Paper nr. 50). Wetenschappelijke Raad voor het Regeringsbeleid.
Sykes, K., & Macnaghtan, P. (2013). Responsible innovation: Opening up dialog and debate. In R. Owen, J. Bessant, and M. Heintz (reds.), *Responsible Innovation: Managing the responsible emergence of science and innovation in society* (pp. 85–107). Wiley.
Taylor, L., Floridi, L., & van der Sloot, B. (reds.). (2016). *Group Privacy: New challenges of data technologies*. Springer.
The Guardian (2021). 'Ex-Uber driver takes legal action over 'racist' face-recognition software', 5 October 2021. https://www.theguardian.com/technology/2021/oct/05/ex-uber-driver-takes-legal-action-over-racist-facerecognition-software
van Boom, W., & Weber, F. (2017). Collectief Procederen – Ontwikkelingen In Nederland En Duitsland. *Weekblad voor Privaatrecht, Notariaat en Registratie*, wpnr 2017/7145: 291–299.
Hoven, J. van den (2013). Value sensitive design and responsible innovation. In R. Owen, J. Bessant, and M. Heintz (reds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 85–107). Wiley.
van der Sloot, B., & van Schendel, S. (2020). Tien Voorstellen Voor Aanpassingen Aan Het Nederlands Procesrecht In Het Licht Van Big Data. *Computerrecht, 1*, 4–13.
van der Vleuten, E., Oldenziel, R., & Davids, M. (2017). *Engineering the future, understanding the past: A social history of technology*. Amsterdam University Press.
Veale, M. (2020). A critical take on the policy recommendations of the EU high-level expert group on Artificial Intelligence. *European Journal of Risk Regulation, 11*, 1–10.
Wallace, R. (2021). 'The names have changed, but the game's the same': Artificial Intelligence and racial policy in the USA. *AI and Ethics, 389*, 1–6.

Chapter 8
Regulation

Embedding or integrating AI into society depends on the existence of frameworks, and therefore regulation. Now that the technology is making the transition from the lab to society, its effects on the economy and the society are subject to widespread scrutiny. This has led to debate about the nature of the regulatory measures needed to ensure that AI is properly embedded in society and government processes.[1]

Attention has focused not only on the opportunities, but also particularly on AI's potential negative consequences. Hundreds of guidelines, codes of conduct, private standards, public-private partnership models and certification schemes have been developed with a view to both promoting opportunities and addressing adverse repercussions.[2] One of the more important initiatives is the European Commission's AI Act21.[3] Moreover, many existing legal provisions and frameworks are potentially applicable to AI, ranging from fundamental rights to liability law, intellectual property rights and the rules on archiving and evidence. In other words, the effects of AI are now controlled by means of a wide range of frameworks and specific rules, many more of which are likely to be laid down in the years ahead.

Formulating desirable and necessary regulations involves not only deciding on the actual content of the norms to be applicable but also a need to determine by means of what specific regulatory instrument these norms will apply (legislation or private arrangements, such as codes of conduct) and the level at which the rules are laid down (international, national, local). In short, the overarching task of regulation relates to question 'what frameworks are required?' Because AI is a system technology, that general question has a number of more specific aspects pertaining not only to such matters as the applicability of existing rules and the need for new ones, but also matters of scope and regulatory level. In this chapter we are concerned specifically with regulation arising from the role and position of government (national and

[1] Meijer et al., 2021.

[2] For an overview, see Jobin et al., 2019.

[3] European Commission, 2021b.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_8

international), and particularly the legislature. We also explore the extent to which, in regulating AI, government can and should rely on the engagement of other actors such as the technology companies that introduce AI applications into society.

Specific questions pertaining to the regulation of AI can be divided into two groups, which are considered separately in this chapter. The first concerns the relationship between regulation and the scope for innovation. Because AI is a system technology, government will need to establish what is required on many fronts. After all, many of the implications of AI's introduction into society remain uncertain and unclear. Government decision-making regarding regulation and its effects must reflect that. This group thus includes the following questions. Do the existing legal rules provide sufficient legal certainty and legal protection? Do those legal rules sufficiently facilitate innovation? And what should be done about the fact that legislation almost always lags behind technological development?

If new rules are deemed necessary, the question of what should be done at the national level and what at the international level then comes into play. As does the question of what can be left to the market and what government should deal with. Although a decision has now clearly been made at the European level to set up a legal framework specifically for the regulation of AI, there remain countless questions – some of them quite fundamental – still not addressed by the proposed European AI regime. For example, it fails to address the potential of algorithmic decision-making as it relates to citizens' legal position, in that it may restrict or even transform constitutional principles such as the principle of legality.[4] The issue of what the amended EU Copyright Directive implies for access to the data used to train AI systems is also left unanswered,[5] along with countless other questions concerning the data used by AI. Other pertinent matters that fall outside the scope of the AI Act include competition and market failure issues in the field of digital services, implications for administrative law, the need to archive algorithms in order to comply with the Dutch Archive Act[6] and even, in the light of the European Charter of Fundamental Rights, whether the Dutch constitution is up to the challenge of AI.[7] In short, the European AI regime represents an important step forward but still leaves countless matters to the national legislature. In the first part of this chapter, we therefore consider various generic issues relevant to the available means of regulating AI in particular legislation. We do not consider the substantive legal issues that exist in various domains, but instead concentrate on possible ways the legislature can address them in general terms.

From the history of earlier system technologies, it is clear that the process of their embedding is consistently accompanied by increasing government

[4] Goossens et al., 2021.

[5] EU Directive 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market (L130/92).

[6] Helwig, 2020.

[7] Passchier, 2020.

involvement. That serves as the starting point for the second part of this chapter, where we deal with such aspects as the influence of time on the regulatory frameworks and practical rules governing AI. As a system technology is embedded in a society, courts, regulators, NGOs and parliament all send out signals about the ability of that society or its public sector to manage the process without intervention by the legislature. One point of particular relevance here is society's and government bodies' ability to ensure that the applications of a system technology take proper account of public values. In practice, many such signals highlight the need for intervention by the legislature, judiciary or regulators.

Almost all system technologies require increasing government intervention over time, and therefore a more explicit role for legislation. For example, the use of steam engines led to high levels of hazardous air pollution in cities that did not decrease until industry was required to build taller chimneys.[8] Similarly, government's role in the regulation of AI is likely to increase over time – and that makes it pertinent to ask how it can prepare. We argue that, at the very least, a broader perspective is needed: the current relatively narrow focus on the technology itself should make way for an outlook that takes in the process of societal embedding and its effects. As well as regulation concerned mainly with the development, characteristics and use of AI, there is a need for regulation that addresses the effects of its integration into society. We also show that as AI increasingly becomes part of our lives, there is a growing need to make fundamental decisions about the design of what we refer to as 'the digital living environment'. The practical implication is that the regulation debate in the years ahead cannot be confined to matters of reliability, transparency and privacy but must also address broader issues concerning the organization of a society in which AI has a prominent place (see Fig. 8.1). Moreover, that debate must at least involve those actors with the ability to shape the digital living environment and the means they use to do so – particularly data.

8.1 Government Standardization of AI

By proposing its AI Act, Europe is clearly signalling the need for specific rules to govern this technology. But there are many issues not covered by the act or that remain to be clarified before the proposal passes into law. So, it remains necessary to consider whether existing frameworks are applicable to AI. Will the Netherlands have to amend its General Administrative Law Act, for example, or the many regulations that apply to specific sectors such as care and mobility?

In the Netherlands and beyond, such issues have been the subject of considerable commentary and debate in recent years, resulting in numerous changes to the relevant frameworks.[9] These matters are outside the scope of this report, however. What

[8] Bakker & Korsten, 2021: 58.

[9] For a recent overview, see Fierens et al., 2021; Van Gool et al., 2021; Chavannes et al., 2021.

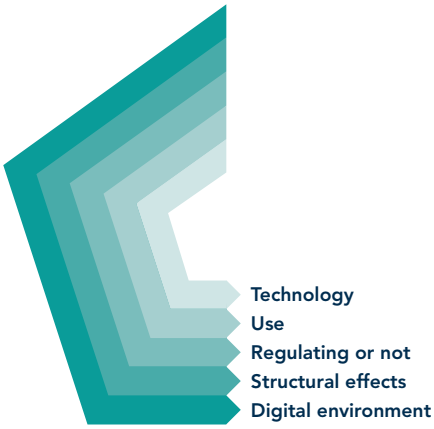


Fig. 8.1 Various levels of regulation

we are concerned with are specific questions regarding the means to be used (type, level), particularly ones that stem from the systemic nature of AI.

First there are questions regarding the breadth of AI's impact. As a system technology, it has the potential to become ubiquitous and to trigger complementary innovations in many areas. In other words, the technology can be utilized in numerous domains and for very wide-ranging purposes. The growing awareness of this fact is illustrated by the research under way into AI's use in sectors such as healthcare, education and defence, as discussed in Chap. 3. Questions therefore arise regarding the legal and other requirements that the development and deployment of AI in those sectors must meet. Given the technology's systemic nature, the debate on the required regulation should focus on whether rules should be tailored to the individual sector, the type of AI used or its practical application. In some cases, generic rules not necessarily specific to AI may be sufficient.

In that context, the distinction made by Lyria Bennett Moses in her research on regulation and technological change is useful. She distinguishes four categories of issue that a new technology may raise.[10] Firstly, it may be necessary to regulate new practices. For example, the General Data Protection Regulation (GDPR) requires that a human being must be involved when an autonomous system processes personal data in a way that has implications for the legal position of a data subject. Secondly, Bennett Moses distinguishes issues that make it necessary to clarify existing rules – because it is unclear whether they apply to the new technology, for instance. After all, many existing rules were not drafted with AI in mind and so may unfairly block the development of new, societally significant applications of the

[10] Bennett Moses, 2007a.

technology. At the EU level there has recently been debate concerning who is legally responsible for the actions of AI.[11] Can the AI system itself be held responsible?[12] That question forms part of a broader debate about the attribution of rights and obligations to AI systems.[13] Historically, the introduction of a system technology has often involved a phase when the applicability of existing rules needed to be clarified. In 1921, for example, the Dutch supreme court had to decide whether electricity could be stolen. After all, its immaterial nature meant that taking it could not be deemed 'the removal of goods'.[14]

Bennett Moses' third category relates to issues necessitating regulation to prevent the uncontrolled introduction of high-risk applications. Historical examples include the risk of 'death by wire' (electrocution) associated with the increasing density and chaotic installation of power networks in urban centres towards the end of the nineteenth century. That issue was resolved by regulations making private utilities responsible for the safety of the electricity network. Similarly, in its proposed AI Act the EU is seeking to ban certain applications of AI, such as those that exploit vulnerable people and those that involve indiscriminate mass surveillance for law enforcement, social scoring (as with the Chinese government's social credit system) or harmful manipulation. In the Netherlands, meanwhile, parliament has called for an end to the use of 'discriminatory algorithms'.[15]

The fourth and final category identified by Bennett Moses comprises issues arising where existing rules are based on assumptions that are invalidated by a new technology, making enforcement of the rules in line with those assumptions inappropriate. Earlier this century, for example, it became necessary to widen the protective scope of the law criminalizing the production of child pornography. Before modern digital editing techniques were available, the relevant legislation (Dutch Criminal Code, Article 240b) was designed to prevent the exploitation of children through child pornography. However, advances in digital image manipulation technology created a situation where pornography could be produced without subjecting the depicted children to actual abuse. The law was therefore amended in 2002 to prohibit the production of images that are harmful to children, even if their production does not involve actual abuse of the subject.

In the years ahead government will have to assess whether the rules that apply within many domains of society and the associated legal domains are appropriate for the new entities, activities and relationships created by AI. The four categories outlined above can be helpful in this regard. If the conclusion is that new or amended regulations are needed, several further questions arise. First, should the regulations

[11] See the study on civil liability conducted by Bertolini (2020) on behalf of the Committee on Legal Affairs of the European Parliament, and the resolution of the European Parliament on the same subject (European Parliament, 20 October 2020).

[12] Hage, 2017.

[13] Brown, 2021.

[14] The so-called "electricity judgment" (Hoge Raad, 23 May 1921).

[15] Kamerstukken II 2020/21, 28362, no. 44.

be specific or generic? Second, which is most appropriate: a technology-neutral approach or a focused one? The third and fourth questions relate to the appropriate regulatory level and to the actors involved and the means available to them. Below we briefly consider each of these matters in turn, particularly in light of our characterization of AI as a system technology.

8.1.1 Specific or Generic Policy?

The pervasiveness of AI can lead to a sense that it is best regulated using generic frameworks. This line of thinking is encountered in the debate around transparency and explainability and in that regarding the formation of new regulatory bodies. However, for the reasons outlined below we regard a generic approach as impractical in the long run.

In the debate regarding the regulation of AI, there is particular emphasis on transparency and explainability.[16] Not only do the workings of the technology, such as its decision rules, need to be explained in a way that people can understand, it also has to be possible to clarify the choices underpinning the use of AI technologies and the actual decisions made by AI systems.[17] After all, clarity is a prerequisite when ascertaining whether or not citizens' fundamental and legal rights are being compromised.[18] Transparency and explainability are also important in determining liability and responsibility for decisions taken by AI – especially if there is a need to understand how an AI system reasons and how particular decisions are reached. But disclosing how an algorithm works, and therefore its operator's business model, may have competitive disadvantages. Transparency has the potential to distort competition and undermine intellectual property rights.[19]

Moreover, what exactly do transparency and explainability entail? Both concepts are open to interpretation, and both may be pursued for a variety of reasons. The interpretation and objectives adopted have a major bearing on the type of information made available, and to whom. For example, a study undertaken in partnership with Statistics Netherlands has found that scientists tend to interpret explainability as meaning explainable to their peers, not the general public.[20] Sometimes the context in which AI is used will imply that transparency and explainability are subject to limitations. For instance, when it is used for medical diagnosis and associated treatment. A strict interpretation of the informed consent requirement has

[16] See, for example, Kamerstukken II 2018/19, 26643, no. 570: 3–4; Kamerstukken 2019/20, 26643 & 32761, no. 641.

[17] In this context, consider public administrative decision-making: Coglianese & Lehr, 2019.

[18] Van Eck et al., 2018. See also the 2020 annual report of the Dutch Council of State, which calls for particular attention for the risks of stigmatization, stereotyping and discrimination (Raad van State, 2020a: 42).

[19] Gerbrandy & Custers, 2018: 108.

[20] De Ree, 29 April 2021.

implications for the level of detail required of the explanation given by a doctor regarding the basis of an AI system's recommendations.[21] Finally in relation to the question of whether generic or specific rules are preferable, we need to consider whether retrospective transparency is sufficient or if prior transparency is also required. The importance of prior transparency varies from one application to another and depends partly on the seriousness of any potential consequences. Demanding it may constrain some applications of AI, such as neural networks. Christopher Reed therefore argues that prior transparency should be required only where AI poses a risk to fundamental rights or where society needs reassurance regarding the safety of its use.[22]

So, although generic transparency and explainability rules may seem sufficient at first sight, specific regulations are often necessary in practice. The judgments of the Dutch Council of State, the nation's supreme court in administrative law issues, in the so-called 'Aerius case' (2017 and 2018) are illustrative in this respect.[23]

Other practical examples demonstrate that numerous factors influence both the requirements for transparency and explainability and their scope. The opportunities presented by AI, and its risks, depend very much on the domains and the organizational context in which algorithms are used.[24] A fine-collection officer being erroneously prompted by an AI system to call a person who does not have payments outstanding bears no comparison to a traffic accident caused by an autonomous vehicle as a consequence of a system misinterpreting sensor data. Similarly, the moderation process of an online platform, where an algorithm gives advice but does not make decisions, cannot be compared with the algorithmic anonymization of court judgments. According to Stefan Kulk and Stijn van Deursen, such differences between domains, organizational contexts and the associated stakeholder interrelationships mean that it is preferable to tackle problems on a domain-specific basis wherever possible.

As indicated above, the choice between specific and generic regulation is also relevant to the debate regarding regulatory oversight of AI (as has been proposed in the Netherlands, the EU and the US), and the creation of a new overall AI regulator or authority.[25] In this context too, a generic approach – a general regulator with access to specific expertise – looks attractive. Nevertheless, there are various arguments against it. A regulatory authority needs a defined field of activity and a set of overarching principles as a basis for its oversight. As with other technologies in the

[21] Klincewicz & Lily, 2020.

[22] Reed, 2018.

[23] Afdeling Bestuursrechtspraak van de Raad van State, 18 July 2018; Hoge Raad, 17 August 2018; Rechtbank Amsterdam, 4 July 2019; Centrale Raad van Beroep, 15 May 2019.

[24] Kulk & Van Deursen, 2020.

[25] See the various contributions to the special issue of the Tijdschrift voor Toezicht (no. 1, 2020) on predictive models, algorithms and AI. The coalition agreement underpinning Dutch prime minister Mark Rutte's fourth administration, which took office early in 2022, states that the government intends to appoint an algorithm regulator tasked with monitoring the transparency, discriminatory potential and randomness of algorithms. He or she will be attached to the Data Protection Authority.

early stages of their application, it is not currently possible to define such principles for AI because not enough is yet known about its risks.[26] Moreover, AI differs from non-system technologies when it comes to the feasibility of designing a supervisory regime suitable for monitoring all possible applications: that would need to have an extraordinarily wide scope yet still be capable of addressing an enormous variety of issues in detail. AI's applications are highly diverse, and their implications are not always comparable. A regime suitable for autonomous vehicles would not be appropriate for smart refrigerators that order food based on consumption patterns. The challenge, therefore, lies not in oversight policy but in the risk that it remains overly generic and consequently requires the definition of countless exceptions for particular applications. More generally, the potentially enormous mandate of a general AI authority or regulator is also problematic given that, in the Netherlands, legal protections related to supervisory activities are already in need of improvement due to the far-reaching enforcement powers currently at the disposal of the authorities.[27] Furthermore, delegating multiple tasks to independent agencies (including regulatory bodies) unduly limits scope for democratic control.[28]

Whether generic or specific frameworks should be used for the regulation of AI is a question also picked up by the European Commission's AI Act, in relation to both the management of risk and the associated supervisory regime (see Box 8.1).

Box 8.1: General and Specific Frameworks Provided for in the Proposed AI Act

The European Commission distinguishes four categories of risk, each associated with certain AI technologies, purposes and sectors. The implication is that AI technologies and applications cannot all be treated in the same way and do not all have the same impact on society. The Commission has therefore chosen to adopt a specific approach to AI. The next question is which applications should fall under which risk management regime. The ban on biometric identification does not go far enough for some commentators,[29] while the decision to restrict the ban on social scoring to public organizations has been questioned given that the private sector is heavily involved in the datafied welfare state.[30] The dual-use nature of certain AI applications is also relevant in this context, as is the fact that AI vendors can design their systems to be modifiable by their purchasers, opening the way for manipulative use. The proposed prohibition on the sale of manipulative AI systems to repressive

(continued)

[26] Nemitz (2018) nevertheless identifies a few areas in which such principles might be sought.

[27] See Verhey & Verheij, 2005.

[28] Raad van State, 2020b.

[29] For example, European Data Protection Board and European Data Protection Supervisor, 2021.

[30] Chiusi et al., 2020.

Box 8.1 (continued)
regimes would therefore be relatively easy to circumvent. Finally, the Commission's proposals are vulnerable to the fundamental criticism that they pay insufficient attention to the injustice and the damage, both tangible and intangible, that AI systems can do to fundamental rights – making the proposed controls inadequate in that regard.[31]

As for whether policy should be general or specific, we find that question addressed in the Commission's proposed governance system. For the most part this builds on member states' existing structures. For example, it envisages each state designating one or more national authorities or regulators to share responsibility for implementing and enforcing the Act. The Commission also proposes that, depending on the sector in which an AI system is to be implemented, regulators should be appointed for that particular sector.

The Act is intended to address, as far as possible, situations where AI may pose risks in practice, now or in the near future. However, it also provides for flexible mechanisms that can be adapted as AI develops and new risks emerge.

The pervasiveness of AI means that the related legal requirements must consider a wide range of factors and so cannot be generic. Generic frameworks are nevertheless relevant, particularly for the regulation of government use of AI. In that context the Dutch Council of State highlights the importance of cohesive legislation for the protection of citizens' rights.[32] With regard to concrete generic frameworks, moreover, both the Council of State and administrative lawyers emphasize the role played by general principles in ensuring good government.[33] For example, Johan Wolswinkel regards the 'guidelines on government use of algorithms' drawn up by the former Minister for Legal Protection as a 'direct consequence' of such principles.[34] Another example of generic regulation, albeit designed for ICT rather than AI, is Franken's 'general principles for good ICT use', formulated in the 1990s.[35] Almost 30 years on, these principles – availability, confidentiality, integrity, authenticity, flexibility and transparency – are still valid in guiding the search for an appropriate balance between effectively safeguarding civic values and allowing scope for the further development of AI.

Our conclusion, therefore, is that weighing up the respective merits of and choosing between generic and specific approaches is always complex. Thorough exploration of the relevant issues is always important, albeit based on a recognition that regulation primarily requires an appropriate balance between effective safeguarding

[31] Smuha et al., 2021.

[32] Raad van State, 2021: 115.

[33] Raad van State, 2021: 105–108.

[34] Wolswinkel, 2020.

[35] Franken, 1993.

of civic values and allowing scope for innovation. Moreover, undue emphasis on certain sectors can lead to the neglect of general matters given that AI, as a system technology, cannot be constrained within a particular policy area or legislative domain.[36] In its fulfilment of this task, government must constantly engage in dialogue as to how specific or general the frameworks regulating AI should be.

8.1.2 Technology-Specific and Technology-Neutral Rules

The second key issue for the regulation of AI is the extent to which the statutory rules should be neutral or tailored to particular technologies. Technology-neutral regulation has several advantages.[37] First, it means that rules are generic and can be efficiently applied in different technological contexts. Second, a technology-neutral law or provision is less likely to become obsolete when technology changes. The underlying rationale is that it is easier to determine how such legislation should be applied by referring back to more general principles. So, technology-neutral legislation may be seen as a more futureproof form and therefore suitable for the regulation of AI.

Nevertheless, the systemic nature of AI means that such legislation is not necessarily the best option. First because technology-neutral legislation depends on a good understanding of the working of technologies that are functionally more or less equivalent. Such understanding enables the legislature to define requirements regarding vehicle braking distances without specifying the nature of the braking system to be used, for example. With a new technology, however, such an approach is difficult because its characteristics remain unknown. Also, a new technology may have qualities that require a different balance to be struck between legislative objectives, such as between accuracy and explainability. If explainability is prioritized over accuracy, rule-based AI systems gain an advantage over those that use deep learning. Finally, with a new technology very different solutions may be required to meet the generic objectives of a law, such as the protection of other road users. If cars are one day able to fly, for instance, the whole idea of braking distances might become obsolete. It was probably with such considerations in mind that the European Commission opted for a functional definition of AI in its proposed AI Act, supported by a dynamic list of actual technologies (see Box 8.2).

Another relevant point is that although AI is a system technology, it is unlike earlier system technologies in certain respects. In Part I we described AI applications as ‘semi-finished products’, which by their nature are constantly changing. Moreover, AI usually exerts an influence over other technologies (computers, communication systems and so on), many of which already operate without human intervention. It is also a technology that is subsumed by, and therefore ‘disappears’ within, society’s everyday processes. These characteristics raise particular

[36] Black & Murray, 2019.

[37] For a critical discussion, see Bennett Moses, 2007b and Koops, 2006.

Box 8.2: Technology-Specific and Technology-Neutral Legislation, and the Proposed AI Act

With its proposed AI Act, the European Commission has sought to create a futureproof regime. It there defines AI as "software that is developed with one or more of [certain] approaches and techniques … and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions influencing the environments they interact with." By focusing on the function of AI systems rather than defining the technology itself, the Commission is aiming to avoid the need to modify the legislative framework as new developments occur.

Nevertheless, the Act does include an annex defining the technologies and approaches that fall within its scope.[38] One criticism of the Commission's approach is that, as a consequence of this, the Act's overall scope is much broader than the fields of application of its more targeted requirements.

questions about autonomy, liability and responsibility, and consequently about the need for forms of regulation that recognize AI's characteristics and are therefore technology-specific.

Inevitably, technology-specific regulation lags behind innovation: the legislation becomes outdated and requires amendment, which takes time during which innovation continues. It is important to note, however, that a great deal is now known about how to address that challenge.[39] We should also avoid falling into the trap of thinking that technological innovation and legislation must always move in step. Or as former chief justice of the US Supreme Court Warren Berger put it half a century ago, "It should be understood that it is not the role and function of the law to keep fully in pace with science."[40]

The choice between technology-specific and technology-neutral legislation must therefore be made on a case-by-case basis.[41] Again, history teaches us that that is the more or less natural course of events. For example, we have legislation requiring third-party insurance cover that applies specifically to motorists, reflecting the seriousness of the potential consequences of accidents involving motor vehicles, but not cyclists. On the other hand, the Dutch Road Traffic Act applies to all road users, not just drivers. In other words, those of its provisions applicable specifically to motorists operate within a generic framework.

---

[38] The list includes: (a) machine-learning approaches including supervised, unsupervised and reinforcement learning using a wide variety of methods including deep learning; (b) logic and knowledge-based approaches including knowledge representation, inductive logic programming, knowledge bases, inference and deductive engines, symbolic reasoning and expert systems; and (c) statistical approaches, Bayesian estimation, search and optimization methods.

[39] On the importance of further characterization of the mismatch between legislation and technological innovation, see, for example, Brownsword & Goodwin, 2012; Bennett Moses & Gollan, 2015.

[40] Cited in Marchant, 2011: 27.

[41] Also Raad van State, 2021: 124.

8.1.3 Framework Levels

The third issue of AI regulation facing government is the level at which frameworks should be established. Because system technologies are by definition universal, they require both national and international policies. Consider, for example, the international arrangements and obligations regarding electricity network voltages and quality, as laid down in the UCTE agreements. More recently numerous global agreements have been made to facilitate the working of the internet, including basic protocols such as TCP/IP, DNS and routing protocols. The recognition that many of the challenges in this field are global has also led to the development of various international consultative mechanisms. Indeed, the proliferation of national AI directives has been accompanied by regulatory convergence at the international level.[42]

In addition to bilateral initiatives on AI, such as those between the EU and Japan, France and Canada and Germany and India, there have been several multilateral ones aimed at the development of common rules. Examples include the OECD's common ethical principles for AI, based on the concept of 'trustworthy AI' developed by the European Commission's AI HLEG.[43] In June 2019 the G20 also formulated a set of ethical principles, based largely on the OECD's.[44]

Other forums are considering rules on AI, too, such as UNESCO[45] and the Council of Europe.[46] Meanwhile, various countries are intensifying their efforts in the field of international standardization. Although the organizations active in this area are concerned mainly with the technical aspects of AI, they are increasingly looking to address ethical aspects as well. There are other reasons for seeking international co-operation, especially where standardization is concerned. For example, China has explicitly stated its ambition to be actively involved in the process of global standardization, particularly in the field of facial recognition. In fact, Huawei's director chairs the ISO's IEC Joint Technical Committee for IT, one of the world's most important standardization bodies. The USA and EU have similar ambitions and aim to promote their vision of AI within relevant organizations. Consequently, standardization bodies such as the ISO, the IEEE and the ITU have become battlegrounds where countries strive to have their own standards adopted globally to give their companies a competitive advantage.[47] For more on this, see Chap. 9.

The choice between applying existing frameworks and developing new ones therefore depends not only on the technology but also on the level at which AI-related issues emerge, not to mention the related strategic ambitions of the country

[42] Smuha, 2019.

[43] See High-Level Expert Group on Artificial Intelligence, 2019.

[44] OECD, undated.

[45] UNESCO, undated.

[46] Council of Europe, undated.

[47] Smuha, 2019: 21.

Box 8.3: Regulatory Levels and the Proposed EU AI Regulation
The European Commission's decision to define European regulations implies a choice in favour of a directly applicable horizontal regulatory framework for AI, and for high-risk AI applications in particular. The legal basis of the AI Act is provided by the Treaty on the Functioning of the European Union, a pact whose intended purpose is to reinforce the single market. The Commission's decision was informed to a significant extent by its desire to promote a healthy internal market for AI systems and thus prevent fragmentation.[48] By assuring the values and fundamental rights recognized by the EU, the AI Act additionally gives the public the confidence to embrace AI applications as well as signalling clearly to companies that only applications that respect those values and rights are welcome in the EU market.[49] However, the use of European single-market instruments as the basis for the regulation of AI represents a limitation, particularly on AI applications that serve broader societal interests. The reason being that the AI Act merely defines certain minimum requirements designed to manage AI-related risks and problems; no positive ethical framework is provided. The portion of AI's societal potential that the market cannot realize – without assistance, at least – is not addressed by the act.

concerned. Here again, the implications of AI's systemic nature are apparent. Given the association between autonomous weapon systems and warfare, such systems cannot be regulated at the same level as AI-based medical devices. The European Union has long had its own licensing framework for medical devices, in which the Dutch Health and Youth Care Inspectorate and other bodies are involved. Although, as the next chapter makes clear, international positioning is crucial as AI makes the transition from lab to society, the ultimate choice of regulatory level depends on many other – largely national – issues. With its AI Act, the European Commission has clearly signalled that the preferred regulatory level for some issues is the European arena (see Box 8.3).

Despite the harmonization process that the European legislature has begun, companies and individual citizens still have to contend with regulatory inconsistency amongst member states. This gives rise to uncertainty. While the AI Act does not require implementation in national law, which would inevitably lead to differences between countries, it leaves many issues unaddressed. Given the systemic nature of AI, there remains a need to ensure that, in the international context, legislation is not (or does not become) unduly inconsistent and does not fail to provide adequate legal certainty. Especially in a cross-border context, legal uncertainty as to what rules apply to the digital world constitutes an increasing problem. Not only because rules

[48] European Commission, 2021a: 1–4.
[49] Floridi, 2021.

differ from country to country but also because it is often unclear which rules – that is, which country's rules – apply. This is the case, for example, when AI systems make use of datasets stored in cloud applications and there is uncertainty as to which country's law applies because their actual location varies in line with available capacity. On occasions, one country's rules require something to be done that another country explicitly prohibits. In a preliminary advisory report for the Royal Dutch International Law Association, Australian professor Dan Svantesson warns about such a problematic scenario, which he refers to as 'hyperregulation'.[50] Not surprisingly, many lawyers and other commentators have called for a far more uniform global legislative agenda. Europe could take the lead in this regard – by means of a European Digital Rule of Law, for example.[51]

8.1.4 Actors and How They Exert Control

Finally, the way that power relationships develop is very important when it comes to the question of how AI should be regulated. Where can and should the market play a role, and is self-regulation appropriate? Where is it possible to rely on citizens' personal responsibility, and where and when is state regulation necessary? In the Netherlands the state's recognition that there is scope for self-regulation is anchored in formal regulatory guidelines. In other words, self-regulation is an explicit policy option for government.[52]

Technology companies are now under pressure and must accept responsibility for defining rules on the development and use of AI. In this regard the landscape has changed considerably in recent years. Whereas they were previously averse to regulation, nearly all the large tech firms are now responding to mounting criticism by working on codes and guidelines clarifying the rules that AI should meet. In some cases, substantive proposals have also been made, such as the creation of ethical review bodies. Moreover, an increasing number of companies is calling for government regulation – in part because they apparently fear losing market share if they do nothing. Various CEOs publicly expressed their views on this matter around the time of the 2020 global summit in Davos. Google's Sundar Pichai said that AI required regulation because of its "potentially negative consequences". Microsoft president Brad Smith (and the company's chief legal officer) warned that governments should not wait until the technology is mature before acting to regulate its use. Microsoft accordingly set up its own committee to make policy recommendations. Meanwhile, IBM CEO Ginni Rometty announced the launch of an internal research lab to devise policy initiatives. Google did set up an Advanced Technology External Advisory Council (ATEAC), but soon scrapped it following a controversy

[50] Svantesson, 2020: 121.

[51] Hagedoorn, 2021: 140.

[52] Staatscourant, 2017, 69426.

about its membership. Facebook too announced the formation of an internal Oversight Board, which the media dubbed the company's 'supreme court'. In May 2021, for example, this board reviewed the legality of banning former US president Donald Trump from the platform.

Self-regulation is a form of regulation practised within an organization or its operational setting. In its most developed form, self-regulation implies private actors themselves defining, implementing, policing and enforcing appropriate norms and rules.[53] The tools used in this context may include private standards, voluntary programmes, professional guidelines, codes of conduct, statements of best practice, public-private partnerships and certification programmes. Some people also favour the use of process-based approaches, where ethical principles are programmed into machines and internal supervision is provided.

Self-regulation can have the advantage of increasing the engagement and support of relevant actors, since its principles are defined from the bottom up by the actors themselves. Theoretically, it also facilitates the use of much more precise standards because these are developed by people who know what works in practice. Furthermore, self-regulation need not be the final regulatory mode adopted; it can also serve as a steppingstone to legislation. Partly for this reason, various public authorities apply 'light-touch' regulation. In the Netherlands, for example, the Ministry of Justice and Security has defined a set of 'guidelines on the use of algorithms by government'. For its part, the UK has proposed an ethical code for AI as a means of avoiding the harmful effects of premature legislation.[54] Generally speaking, the adoption of bottom-up initiatives is a faster process than the implementation of legislation, making them useful for addressing urgent issues. The resulting rules are also easier than legislation to amend or rescind in line with changing circumstances.

But, of course, self-regulation entails a democratic deficit that is potentially problematic in that it may undermine the legitimacy of the rules concerned. Another significant shortcoming is that rules defined privately are more difficult to enforce.[55] Often, therefore, not all actors subscribe to them. Where AI is concerned, it is mainly benevolent actors that participate in initiatives to ensure conformity with ethical principles and societal values. Meanwhile, more problematic and controversial applications remain unregulated. On top of that, the enormous proliferation of charters, guidelines and the like can make co-ordination difficult. Which document should be followed in a particular case, and what happens if they are mutually contradictory? Who acts as referee in such circumstances?

Gary Marchant predicts that 'soft law measures' will become the default in the years ahead because of AI's rapid development and global spread. The most that

[53] See, for example, Giesen, 2007; Smits, 2015.

[54] Select Committee on Artificial Intelligence, 2018.

[55] A court can also interpret and apply the relevant legal rules in the light of the self-regulation regimes agreed by actors in the field. See Giesen, 2007.

government can do is resolve minor problems here and there. According to Marchant, it is therefore necessary to investigate how self-regulatory mechanisms for emerging forms of AI can be indirectly enforced and co-ordinated. Bert-Jaap Koops has made the same point in a slightly older article on ICT regulation. He suggests that pure self-regulation barely exists in practice. More often than not, government also plays a role.[56] In practice, self-regulation and government regulation frequently coexist, supplementing and reinforcing one another in key respects. In many countries, for example, private actors and governments have collaborated on the formulation of AI strategies.[57] It is also common for basic standards to be defined in law, but with the details left to sector-specific self-regulation. According to Koops it is necessary to have a combination of consistent government and public pressure, rewards for prosocial behaviour and monitoring mechanisms to ensure that measures are more than mere window-dressing.

This point is very important in relation to the development of AI, now that it is becoming clear that self-regulation is not sufficient to deal with many issues. As it continues to develop, AI still faces numerous technical challenges that complicate its trustworthy application. Self-regulation assumes that developers are able to bring products to market that meet industry standards, including trustworthiness criteria. According to AI researchers Gary Marcus and Ernest Davis, however, that is not always the case. They see the absence of good development practices as particularly problematic here. AI research tends to focus on short-term solutions, such as code that works immediately, but without the layer of technical safeguards seen in other fields. Stress testing is almost unheard of and machine learning systems with adequate risk margins are never applied.[58] Furthermore, good engineers in other domains always provide their products with fallback options such as duplicate brakes, multiple control systems and fail-safe functions. But these are rare in AI.

The global success of the big technology companies owes much to an approach focused on the large-scale marketing of new but usually unfinished products. Users then take care of further product development and optimization. This development model is diametrically opposed to that used for cars, pharmaceuticals or aeroplanes, which are extensively tested before entering use. Of course, unlike many physical products software can easily be modified and updated remotely; the detection of flaws does not therefore require expensive recalls. But as applications are integrated more and more with real-world processes and – as is the case with AI – come to underpin decisions that have major impacts on people's lives, the practice of releasing unfinished products entails ever greater risk. As the European Commission argues, there is therefore an increasing range of circumstances in which the use of AI is permissible only if certain basic trust requirements are satisfied.

---

[56] Koops et al., 2006.

[57] Mols, 2019.

[58] The authors use the examples of a lift, which is always able to carry a much greater weight than calculations suggest, and servers that can handle more internet traffic than is necessary in everyday practice.

For self-regulation to be an effective and legitimate option, moreover, AI must also conform to a wider set of principles that reflect society's expectations and are codified in national and international treaties. Here existing guidelines could be distilled into a set of principles very similar to those used in medical ethics: respect for human autonomy, harm prevention, honesty and explainability.[59] Those points are already central to the OECD's common ethical principles for AI and the work of the European Commission's High-Level Expert Group on Artificial Intelligence. That said, however, the frequent references to medical ethics cannot hide the fact that many of the conditions required for the implementation of those principles are not currently being met.[60]

Unlike in medical science, AI development activities have no common goal comparable with the promotion of patient health and welfare. As explained in Part I AI development is a much newer field than medical practice, with a very short professional history and consequently barely any clearly articulated norms of good conduct. Third, by contrast with medical practitioners AI developers come from a variety of disciplines and professional backgrounds, with divergent histories, cultures, incentive structures and moral obligations. The most closely related established discipline, software development, is not a legally recognized profession with obligations towards society – in part because it lacks a system of licences and clearly defined professional duties of care. The two biggest professional organizations, the Institute of Electrical and Electronic Engineers (IEEE) and the Association of Computing Machinery (ACM), have published and repeatedly revised various codes, but these are relatively concise and theoretical, and they do not include recommendations or specific behavioural norms.[61]

Perhaps the most important difference, though, is that AI development is not governed by any discipline-specific legal or professional accountability mechanisms. At present there is almost no scope to seek redress or remedy. Data breaches and privacy infringements form exceptions in this regard, but that is because these abuses are covered by formal legislation (GDPR). Protection of other values is left to private self-regulation mechanisms, whereby a long-term commitment to upholding those values is by no means assured.[62] This is particularly problematic now that discussion surrounding AI in the business community and some sections of academia has come to focus primarily on the question of how and under what conditions the technology should be used. The fundamental desirability of such use is barely considered in this debate.[63]

Moreover, it is characteristic of a system technology such as AI that its introduction to society raises issues that transcend the domain of the technology companies

[59] Floridi and Cowls (2019) argue that current AI principles are most similar to those used in bioethics; they also add the principle of explainability.

[60] Mittelstadt, 2019: 503.

[61] Mittelstadt, 2019: 503.

[62] For references see Mittelstadt, 2019: 504.

[63] Greene et al., 2019.

and other private actors involved. Tech firms tend to propose technical solutions to problems.[64] Which is hardly surprising given that that is where their expertise lies. But the scope of such solutions is too limited to tackle these issues effectively. Discrimination, for example, is primarily a societal problem that requires solutions in such domains as institutional access, coupled with a normative debate about what forms of discrimination we find socially acceptable. Secondly, some issues do not operate at the company or application level or cannot be addressed adequately there. Even when companies comply with all relevant legislation and regulations, such second and third-order effects can still arise. One example is changes in employment patterns and the associated need for training. Thirdly, what is actually at issue here is the purposes for which AI may and may not be applied. Should it be used in autonomous weapon systems, for instance? Such matters are not the province of technology companies, at least not exclusively, because of potential conflicts of interest. Fourthly, self-regulation is not an option when human rights and the fundamental standards and values of democracy are at stake,[65] as various academics argue is the case with AI.[66]

Where government standardization of AI is concerned, therefore, debate should not be confined to the characteristics of the technology itself (is it reliable, safe, transparent and explainable?) and the activities of the companies and organizations that develop and utilize it. A system technology requires a much broader discourse, encompassing such matters as the goals we wish to pursue as a society and hence where, for what purpose and under what conditions we want to use AI,[67] as well as whether restrictions or even bans on its use in certain domains (as also proposed in the European Commission's AI Regulation) are needed.

The systemic nature of AI also results in the overlap of societal, political, commercial and research interests. No one actor or group of actors can simultaneously defend all of these. It is therefore impossible, and also undesirable, for a single actor to monopolize the ethics of AI or to dominate the agenda with regard to the regulatory frameworks governing it. In order to prevent the private sector and, to some extent, the academic community defining what constitutes a good AI society, authors such as Corinne Cath believe that a 'bolder' strategy is required. They envisage this as addressing the entire spectrum of unique challenges that AI presents for society with regard to fairness, social equality and accountability.[68] As argued in Chap. 7, the formulation of that strategy should involve all parties affected by AI.[69] Government itself is of course part of this matrix, since it has the task of considering the big picture and the interests of all the various parties concerned. The proposed AI Act also demonstrates that, after thorough consideration by government, further

[64] Häußermann & Lütge, 2021. See also Hagendorff, 2020. Hagendorff observed as well that the more men were involved in defining ethical guidelines, the more often technical solutions came to the fore.

[65] Vetzo et al., 2018.

[66] European Union Agency for Fundamental Rights, 2020; Hirsch Ballin, 2021.

[67] Floridi et al., 2018.

[68] Cath et al., 2018.

[69] Cath et al., 2018: 523.

Box 8.4: Actors and the AI Act

By proposing its AI Act, the European Commission has clearly taken the initiative on regulation in a manner that will influence the course of market developments. Nevertheless, private actors still have a part to play. One aspect of the proposed act that has attracted little comment is that much of the responsibility for regulating AI, in particular high-risk systems, will rest with standardization organizations such as CEN (Comité Européen de Normalisation) and CENELEC (European Committee for Electrotechnical Standardisation).[70] The act requires any party wishing to market an AI system in the EU to consult certain as yet undefined AI standards. These will include mandates to establish a quality system, draw up technical documentation, organize human supervision and undertake logging.

The standardization process is sensitive to commercial lobbying, however, and a lack of resources and expertise often makes it difficult for interest groups to participate. Consequently, some commentators have expressed concern that the new legislative framework underpinning the proposed act will not adequately protect consumers' interests. Another important issue is that high-risk applications have numerous implications for fundamental rights, a field in which standardization bodies have limited expertise and experience.

A further criticism is that the AI Act does not establish procedural rights for individual citizens or interest groups, such as the right to complain, seek redress or dispute a decision.[71] In other fields the existence of such rights has proven an important driver for the development of jurisprudence, particularly when influential companies have appeared to exercise undue influence over policymaking, creating a need for balance. As currently drafted, the act allows only companies that are subject to its requirements to challenge government decisions. Given that AI has implications for fundamental rights, it is pertinent to ask whether and to what extent other parties should also have a say.[72]

In its response to the proposed act, the Dutch government has emphasized the importance of clarity for citizens and consumers as to how they can exercise their rights and has expressed a desire to see appropriate provisions made in specific consumer (and other) regulations.[73]

steps may be required to assure that input is obtained from other parties – during the concretization of legal requirements, for example, or to draw attention to injustice and harm (see Box 8.4).

[70] Veale and Zuiderveen Borgesius (2021) argue that, in practice, very few situations will arise where use is made of the independent 'notified bodies' accredited by national regulators to which the act makes repeated reference. Once the standards are in place, any party seeking to market an AI system will merely have to perform a self-assessment.

[71] European Data Protection Board & European Data Protection Supervisor, 2021.

[72] Cf. Smuha, 2019.

[73] This point is made in Fiche 2: Verordening betreffende Kunstmatige Intelligentie, van het Ministerie van Economische Zaken en Klimaat, Ministerie van Justitie en Veiligheid en Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

It is clear from the first part of this chapter that government has a responsibility to regulate the embedding of AI within society. In this regard it should focus on the tools available for use in that context and should address such matters as appropriate regulatory characteristics and levels, as well as the extent to which private actors are willing and able to protect civic values of their own accord.

The scope and intensity of government's role are separate matters, which are considered in the second part of this chapter. In view of the history of previous system technologies, we argue there that the process of embedding such a technology goes hand in hand with increasing government involvement. In this case that should not be restricted to the regulation of AI and acute AI-related problems but extend to the long-term co-evolution of technology and society, including the associated structural challenges, opportunities and risks.[74] This implies government regulation as a means of shaping the digital living environment, not to mention interaction between that environment and numerous issues in the physical world. The adoption of such a comprehensive, future-oriented view of regulation is a prerequisite if government is to properly discharge its responsibility to protect civic values.

Key Points – Government Standardization of AI

- The systemic nature of AI means that it touches on a variety of societal, political, commercial and research interests. Comprehensive consideration, safeguarding civic values and protecting different parties' interests are possible only if government plays a guiding role.
- As a system technology, AI is going to become ubiquitous. Government must therefore be able to oversee the full spectrum of societal challenges it presents and to intervene promptly with legislation where necessary. Government should not confine itself to the technology itself or to users' activities, but also take a broad view encompassing such matters as the interests we wish to pursue as a society and hence where, for what purpose and under what conditions we want to use AI.
- Government regulation of AI should not take a standard approach. Decisions regarding the regulatory instruments to be used (legislation, self-regulation) and the level at which regulation should take place (international, national, local) will require an appropriate balance to be found between the effective assurance of public values and the provision of scope for innovation.
- In order to address these challenges with prompt, effective and significant interventions while maintaining policy cohesion, government must adopt a broad legislative strategy.

[74] See, for example, Just & Latzer, 2017; Krupiy, 2020.

8.2 AI Regulation and the Digital Living Environment

In the regulation of earlier system technologies, government intervention gradually increased over time. At first a new technology is often given space to develop. As it enters more widespread use and becomes more deeply embedded in society, however, and as its effects become clearer, more formal requirements often become necessary. When motor cars entered use, their growing prevalence gradually led to more dangerous situations and to traffic accidents, prompting large-scale protests in the US and Europe. In the US the car lobby won the day, and it became the dominant mode of transport. But in Europe public transport systems developed and much more explicit allowance was made for pedestrians to facilitate the mobility of the less well-off.

The effects of a system technology and the opportunities and risks associated with it change gradually over time, therefore. Consequently, the focus of regulation widens from the technology itself to its general effects, such as modification of the dynamics of the economy and the context in which it is used. So, intervention to regulate earlier system technologies was extended to address related matters such as road safety, urban pollution and traffic congestion, the safety of consumer electronics and emissions of greenhouse gases. As these examples show, there are always trade-offs to be made and in this respect companies and pressure groups always seek to influence the embedding process in line with their own interests.

Although government intervention is sure to increase gradually with AI as well, it is not possible to say in advance how extensive and intensive the regulation needs to be. Integrating a system technology into society is a process that spans decades and involves considerable uncertainty, particularly regarding the impact it will have and how regulation can manage that. In this section we begin by considering this uncertainty, which can be the cause of both tardy and premature intervention to prevent problems. The timing of interventions is therefore the second theme we explore. In that context we also reflect on the need for government to be alert to outside influences, particularly market forces. The salient point being that, as a technology becomes more embedded, it becomes increasingly difficult for government to counter or redirect the regulatory influence of other actors.

The final major factor affecting the extent and intensity of government regulation is the interaction between a system technology (in this case AI) and more general developments and challenges impacting society. A system technology shapes society and society shapes the technology. The position adopted by government will have a major bearing on the nature of this interaction and whether it can be linked to developments that at first sight seem to have little to do with AI (climate change, say, or the sustainability of the care system). Our discussion of these three issues leads us to the conclusion that there is an urgent need to regulate not only such factors as privacy, liability, transparency, insurability and consumer protection, but also to organize the digital living environment in a way that will enable AI to support public values in the long term.

8.2.1 *Uncertainty*

As the histories of the internal combustion engine and electricity show, regulatory frameworks are not created overnight but often continue evolving for decades. When a system technology has recently made the transition from the lab to society, very specific practical embedding challenges often arise, relating to such matters as liability, insurability and – in the specific case of AI – the performance of legal acts by autonomous systems and copyright on algorithms. In a lot of cases these can be addressed by falling back on and updating existing frameworks, but in many instances the definition of special rules for AI would currently be premature. The same applies to initiatives like the creation of an AI authority or a special AI regulatory body because the sphere of responsibility of such an agency cannot yet be defined. Not enough is known at present about generic patterns in relevant fields such as AI deployment-related risk, legal protection requirements or competition and market regulation.

With a system technology such as AI, government should therefore initially take an incremental approach to regulation. For the management of known risks, existing rules can be clarified or amended fairly soon after introduction of the system technology. Or new rules can be implemented. This is already happening in various AI-related fields, albeit fairly slowly and not on a systematic basis. However, both the complex technological structure of AI and its association with particular usage contexts introduce considerable uncertainty here, with a high degree of complexity and therefore unknown risks.[75] Their details will become apparent only once AI enters more intensive, large-scale use, and so careful monitoring involving early-warning regimes, error registers and the like is required.[76]

Parties close to developments will typically be the first to become aware of issues and problems associated with the embedding of AI. As well as the community organizations discussed in Chap. 7, courts, supervisory bodies and parliament can all perform an early-warning role.[77] As community representatives, members of parliament can hear of incidents that may indicate threats to civic values. Courts are asked to rule on cases where litigants have been affected by the use of algorithms, as in the Council of State cases mentioned earlier in this chapter. Inspectorates and regulators are able to observe the introduction of new applications to the market, such as AI-based medical devices and car driver support systems, and have the task of supervising processes in which AI is increasingly being used, such as risk assessment, cybersecurity, social security and logistics.

Such bodies also perform a societal role in the early detection of developments with implications for the protection of public interests and the balance of power.[78]

---

[75] Burrell, 2016.

[76] For the management of unknown risks and the precautionary principle, see WRR, 2008.

[77] Cf. part IV of Bennett Moses, 2007b.

[78] WRR, 2013.

In the Netherlands, for instance, the Authority for the Financial Markets and DNB (the Dutch Central Bank) have investigated the use of AI in the insurance sector[79] and the Court of Audit has examined how the national government deploys algorithms.[80] The latter concluded that the responsible development of complex automated applications requires better supervision and better quality control, and accordingly developed an assessment framework. A number of inspectorates and market regulators additionally took the initiative to set up an interdepartmental working group to share knowledge and experience of the supervision of AI and algorithms. The more difficult it is to resolve problems within existing frameworks and/or the more generalized those problems become, necessitating the use of generic measures, the more important it is to ensure good feedback of such bodies' observations to the political and public administration communities.

8.2.2 Timing of Government Interventions

It is therefore clear that government, more specifically the legislature, should initially proceed cautiously before assuming a more active role in due course. However, the scope for attaching effective requirements to the use of a technology changes over time. Once a technology is firmly established, influencing its use becomes complicated and sometimes even impossible or impractical. That is due to the so-called 'Collingridge dilemma' and to actors other than the government guiding the process of technological embedding and thus performing a regulatory role.

The Collingridge dilemma is a governmental information and power problem. It was first formulated by David Collingridge in his 1980 book The Social Control of Technology: "When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time-consuming." In the early stages, when it is still possible for government to influence the development of a technology, its effects have yet to become apparent and so there is a significant risk that legislation will prove inappropriate, ineffective or even counterproductive. But by the time its effects are manifest, and it is clear what needs to be done, the technology is so firmly embedded that legislating to bring about change involves considerable cost.[81]

Over the years the Collingridge dilemma has attracted considerable attention. One may interpret it as implying that government should not interfere with new technologies, certainly in their early stages. When a technology is in its infancy, it is vulnerable and therefore generally warrants a careful, nurturing approach. At this stage, moreover, its introduction is surrounded by unknowns and uncertainties. At the same time the Collingridge dilemma implies that the opportunity to intervene

[79] AFM & DNB, 2019.

[80] Algemene Rekenkamer, 2021.

[81] Cf. Bijlsma et al., 2016.

may be lost if nothing is done until the technology has become pervasive. Those dangers are of course two sides of the same coin: if one starts a race late, one has ground to make up before the finish and that may prove impossible. While there is truth in the Collingridge dilemma, that is somewhat simplistic – which prompted Wendel Wallach to describe it as a dogma.[82] Collingridge disregards the many forces that influence how a technology is used in practice, at all stages of its societal embedding.

8.2.3 The Guiding Effect of Technology

Lawrence Lessig's 1999 book Code is a classic treatise on such forces.[83] The author argued that digital technology is strongly influenced not only by legislation, market forces and societal standards but also by its technical design – in other words, by its code.[84] Some years earlier, in his work on the politics of technology, Langdon Winner had demonstrated that the workings of a technology are also a form of regulation.[85] The same is true of AI. Consider the algorithmic moderation that platforms use to proactively police the online content shared by internet users.[86]

The development and application of AI are also subject to various forces, from the existing legal rules and the private actors that develop the technology to societal concepts of autonomy and human dignity, which influence decision-making in the system design process regarding such matters as the prioritization of output accuracy over explainability. Standardization and the extent to which its tone is set by the private sector are further examples. Furthermore, there is another issue we must also consider in relation to AI: the controlling and therefore guiding role played by humans, and hence their influence over the regulatory power of the technical design, are changing. This aspect is particularly problematic because AI systems are generally non-transparent, complex and self-learning.[87]

The fact that regulation becomes more difficult over time is attributable not to the technology's deterministic 'natural' or 'unavoidable' development but to path dependency. This phenomenon is best illustrated by the way the road network operates. When constructing new roads, existing routes are often followed. But many of these are not ideal. They are used nonetheless because the existing urban environment is adapted to them. It would be extremely expensive to move all the homes and businesses along an existing road, for instance. So, when a route is chosen for a new road, the efficiency of the ideal path must be weighed up against a wide range of

[82] Wallach, 2015: 71–72.

[83] Lessig, 2006.

[84] Lessig, 2006.

[85] Winner, 1983: 97–111.

[86] For more on algorithmic supervision and the associated European policy, see Kulk, 2020: 132–140.

[87] Yeung & Lodge, 2019.

other interests associated with decisions made in the past, which often prevail. The same process is evident throughout society. For example, the supposed exponential growth of computing power (Moore's law) is not really a law but a self-fulfilling prophecy – in fact nothing more than an annual goal set for engineers, labs and companies, which then dictates that the digital infrastructure must grow, driving demand for staff, research funding and fast semiconductors.

In such a process there is always a point at which a certain interpretation of the design or use of a technology becomes the norm and 'closure' occurs.[88] It then ceases to be controversial and one of the competing designs is retained while the others are discarded. Around 1900, for example, various types of vehicles were being pioneered, some with electric motors and others with different kinds of internal combustion engine.[89] Then, over the course of the twentieth century, the petrol-fuelled engine became predominant, partly because of the limited range of electric vehicles, a problem that has yet to be fully resolved. Another familiar example is the bicycle. Nowadays this is an efficient mode of transport with two wheels of the same size.[90] But it was initially seen as a masculine mode of transport requiring considerable strength and athleticism. Gradually competition developed between the 'penny farthings' responsible for that perception and machines resembling the modern bicycle, designed for safety and efficiency. After several decades the modern version prevailed, and others disappeared from use. It should be noted, though, that the quality of a technology is not always the deciding factor in closure. In the 1970s and '80s, for instance, there was competition between three mutually incompatible video recording standards: VHS, V2000 and Betamax. Marketing factors such as price and support proved decisive in VHS ultimately dominating, rather than the better technical quality of the competing systems.

Closure has far-reaching consequences: alternatives disappear and cease to be the subject of debate. People, relationships, other technologies and existing practices and procedures – once closure occurs, all their interrelationships are redefined. The lesson is that, between the introduction of a technology and that technology becoming firmly embedded in society, there are always one or more tipping points. At those points problems associated with the technology become apparent while there is still an opportunity to address them.[91] Technological change never comes out of the blue, even in the most dynamic situations. The window of opportunity for action may be very short, or it could remain open for years. The faster technological development proceeds, the less scope for intervention there is. That scope can also be reduced by widespread adoption, as with the mobile phone. Because it is so widely used, this has become an important platform for many other products and services, such as smart domestic appliances and payment services. Various factors can bring about closure, then.

[88] Bernstein, 2006.

[89] Bakker & Korsten, 2021: 37.

[90] Bijker, 1995.

[91] Wallach, 2015: 72; Bennett Moses, 2007a: 600.

In light of the considerations set out above, we may conclude that if government allows itself to be trapped for too long by the uncertainty paradox that every new system technology creates,[92] it runs the risk of missing the opportunity for effective and significant intervention. In this regard it is instructive to consider the regulation of the internet, or rather the lack of it, as a cautionary warning.[93] In such situations failure to act can clearly have major consequences for the proper safeguarding of civic values. When government is inactive, other forces have ample opportunity to control the way the technology is embedded in society. Where the internet is concerned, a handful of American tech companies have been able to drive the development of social media platforms.[94] When it comes to AI, failure to regulate will give free rein to forces specific to the technology itself, potentially reducing the scope for transparency, explainability and human intervention. The crucial point here is that – consistent with the Collingridge dilemma – the more time passes, the more difficult it gradually becomes to counter the influence of non-governmental forces and to guide AI's direction of travel.

The EU's move to regulate various specific AI-related issues is therefore welcome. In relation to the overarching task of regulation as described in this chapter, it is important to note that the Commission has proposed applying rules of varying strictness on the basis of a four-tier risk classification system[95] reflecting the function, intended purpose and modalities of the AI application. In other words, the Commission is not focusing exclusively on the technology or operating based on a generic assessment of its use, but instead bearing in mind the specific context. Biometric identification is in principle prohibited, for instance, since it poses specific risks to fundamental rights – in particular human dignity, respect for private and family life, the protection of personal data and non-discrimination – but it may nevertheless be used in certain strictly defined, limited and regulated circumstances such as targeted searches for missing children or the perpetrators of serious crimes. By adopting a risk classification system, European and national authorities should be able to obtain early insight into the problems and risks associated with AI, thus considerably increasing their ability to influence the course of developments. Such an approach may be compared with the early identification of problems associated with introduction of the internal combustion engine (traffic accidents at the start of the car age) or the construction of the first railways (ownership issues, outdated assumptions about rights of way).

The latter conclusion brings us to the final issue associated with the overarching task dealt with in this chapter: what should be on the regulatory agenda? It will be apparent from the considerations set out above that – no matter how uncertain the process of AI's societal embedding or its outcome may be – government's regulatory activities in the years ahead must not be limited to the maintenance of existing

[92] Van Asselt et al., 2010.

[93] Stikker, 2019; Black & Murray, 2019.

[94] Helberger et al., 2018.

[95] A distinction is drawn between applications according to the level of risk they entail: unacceptable, high, low or minimal.

frameworks and the formulation of principles for the concretization of rules by supervisory authorities or the judiciary. The EU's proposed AI Act illustrates the need for a broader legislative agenda. The European legislature's intervention is consistent with the trend of governments, including ours in the Netherlands, being urged by an increasing number of parties to develop sound technology policies.[96]

With consideration for future developments, the overarching task of regulating AI entails more than addressing countless separate issues, risks and advances. With a system technology, it is not only the technology itself and its practical applications that require legislation but also its wider effects on society. It is very important here that government not become mired in the countless separate individual legal issues at play in the many fields where the technology is used, but instead keep sight of the bigger questions associated with its embedding. This requires it to design its regulatory agenda for AI on the basis that embedding the technology is essentially about responsibility for and the design of our digital living environment.

8.2.4 A Legislative Agenda for the Digital Living Environment

Various historical examples illustrate the need for the legislature to consider AI developments primarily from the perspective of society's general organization in the years ahead. When motor cars, aeroplanes and electricity were introduced, legislators were obliged to make decisions with a scope that extended far beyond individual policy portfolios. In the cases of cars and electricity, that respectively involved developing new visions of land-use planning and the organization of the physical living environment. Like electricity, AI is both a commodifiable good that can be used to gain economic advantage and a good beneficial to large groups within society, or even the entire population. Electricity extended the day, made homes safer and cities cleaner and improved people's lives in many ways. AI can be expected to deliver similar benefits. However, that will require government ensuring that the technology is used in places where it can be most valuable and deployed on a scale and for purposes compatible with the aspirations of Dutch society.

Like the car and the aeroplane, AI is a very energy-hungry technology.[97] It can also be a major cause of pollution in its broadest sense. ''It has a lot to offer, just as industrial innovation, intensive agriculture and the chemicals industry have brought

[96] See, for example, the pre-election joint commitment on digital policy (Digitale Stembusakkord) signed by a broad group of political parties on 12 March 2021. Strict supervision of algorithms is one of the eleven points to which the signatories commit.

[97] It has been recognized for some time that a single online search query consumes enough energy to power a lamp for some seconds; Google itself has said that one search is equivalent to seventeen seconds' use of a 60-watt bulb. There are also many countries in the world whose annual energy consumption is less than that of the Bitcoin network, and numerous wind turbines dedicated to powering the data centres of firms like Google and Microsoft. Cf. Coeckelbergh, 2020 and Van Wynsberghe, 2021.

us a great deal – but at considerable cost to the environment, the climate and our planet. An excessive cost, it now turns out. Similarly, innovative digitalization, economic and social progress and growth all involve risk. Here again, not only are the interests of individual citizens and companies at stake, but also collective goods and values."[98] In short, as AI's role in society increases we will see an growing need to make fundamental decisions about the organization of society and the 'digital living environment'.

Ernst Hirsch Ballin has called for AI policy to be linked particularly to the promotion of resilience over the course of the human lifetime to counter our inherent and inevitable vulnerability as people. In his analysis of the relationship between AI and human rights, Hirsch Ballin formulates three benchmarks for AI policy – two of which are pertinent in relation to the organizational issues being discussed here. The first is that artificial intelligence should be linked primarily to people's aims in life. In other words, we need to develop AI policies "to correct the complex processes that prevent people from feeding themselves, educating themselves and otherwise developing their life projects".[99] The second relevant benchmark is the need for AI's purposes and design to be "linked to humanity, respect for diversity and support for freely accepted life projects" at all times.[100]

When addressing developments in fields quite distinct from AI that affect and shape society in fundamental ways, the Dutch government has previously brought together numerous separate policy portfolios to view them as an integrated organizational issue. Land-use planning is a good example. In recent decades several administrations have published policy documents setting out tasks and objectives in this area, complete with guiding philosophies, toolkits and implementation plans.

A good Dutch example of a policy document tailored to the challenges of AI and the digital living environment is the 234-page 1998 paper from the Ministry of Justice on 'legislation for the information superhighway' (Wetgeving voor elektronische snelweg). Its purpose was "to provide a legitimate basis for government action during the transition to the information society, insofar as legislative instruments lend themselves to that function; to translate that legitimate basis into an assessment framework for the legislature; to differentiate between the physical world and the electronic environment in key areas; [and] to make proposals regarding real-world issues that arise as a result of technological developments." The current Dutch Digitalization Strategy, which is updated annually, is far less explicit than either of these two documents regarding policy tasks, steering principles and the regulatory toolkit. Furthermore, and unlike the information superhighway paper, it does not address the full breadth of its policy domain, digitalization. Although its declared aim is "a successful digital transition in the Netherlands", the strategy in fact consists largely of a list of practical action points, most for the public sector, organized under a number of thematic headings.

[98] Prins, 2018: 1563.

[99] Hirsch Ballin, 2021: 33.

[100] Hirsch Ballin, 2021: 34.

The risk with such a list-based approach is that broader and often far more fundamental developments receive insufficient attention or are not based on a long-term vision. Below we identify three developments that the WRR considers crucial to AI's integration into society. Fundamental to all three is the observation that AI is a largely 'civil technology', as explained in Chap. 4. That is, a technology developed by the business community and not, or only to a lesser extent, by government or independent researchers. It is also a technology based on large volumes of data, the abundant availability of which is related to the design of the digital world, with the internet centre stage. Moreover, the actors that already dominate the collection, processing and dissemination of data over the internet are also AI's largest investors and developers.[101] These factors have a significant bearing on the process of embedding AI in society, and government will have to address them. How it does this in fulfilment of its regulatory task in the years ahead will determine whether AI is embedded in a manner that fully respects fundamental rights and society's core values.

8.2.5 Three Developments That Influence the Embedding of AI

We have identified three key developments that will influence AI's integration and embedding in society. They are increasing surveillance in the public domain, unequal growth in the use of digital resources and the concentration of power within the digital domain, with spill-overs into other areas of society such as its relationship with democracy. A considered government view of these developments is crucial because they have major implications for the ability to regulate AI, be that by government itself or by other actors, in the short and the long term. Moreover, they also shape the relationships associated with the more specific issues raised by the embedding of AI. For example, transparency is important not only so that substantive decisions made by individual AI systems can be understood but also so that the developers and users of such systems can be prevented from exerting undue and unchecked influence over society.

8.2.6 Surveillance

The first development is the large-scale processing of personal and other data for surveillance purposes, and its use to influence how individuals and companies behave. Although such activities are certainly not new in themselves,[102] it does not

[101] Nemitz & Pfeffer, 2020.

[102] Cf. the contributions to Hildebrandt & Gutwirth, 2008.

follow that government should be unconcerned about them. The practical scope for surveillance allowed to companies, other organizations and private citizens in the years ahead will be crucial to the direction digital society takes in the longer term. Observation, even covert, already appears to be viewed as increasingly normal by companies, governments and the general public.[103]

Furthermore, when AI is involved observation data becomes more than the mere input and output of a digital system – it actually helps determine the quality of the system's risk assessments, predictive models and modelling variables and is therefore formative in how the system works. Consider Gijs van Dijck's research into the quality of the OxRec algorithm used by the Dutch probation service to advise courts on the risk of a suspect reoffending.[104] Since this was introduced, he argues, its users seem to have allowed themselves to be guided by predictions that are regularly incorrect even though the new system performs no better than its predecessor and also entails a risk of discrimination on the basis of race, class or other social characteristics. Moreover, data shapes not only systems but also policies. For example, there are now calls for a transition from a policy cycle to a data cycle.[105]

Surveillance of citizens, consumers and others has become standard practice for almost all companies and government agencies, as well as many private individuals. Commercial firms now base their business models on surveillance, with the consequence that any restriction of their capability in this area implies lost income. For government the collection and processing of data, particularly personal data, opens the way to monitoring the activities of people and companies in many different arenas.[106] Another significant point is that, as pointed out previously, personal-data processing now often takes place at the group level as well as the individual level – a practice that existing protection mechanisms are not well designed to address. The point here is not that surveillance is inherently undesirable or dangerous; the real cause of concern is the increasing distortion and imbalance it causes to relationships between citizens, businesses and government. As we elaborate later, that is problematic with regards to control over and access to data, and therefore its wider availability.[107] Another problem is imbalance in the extent to which actors can influence the collection and further processing of data. In recent decades people's insight into and control over what happens to their data has been decreasing,[108] prompting repeated references to a 'black box society'.[109] AI is liable to make the black box still more impenetrable, partly as a consequence of inadequate supervision and judicial control.[110] The reason being that, while AI enables people to monitor 'each

[103] Zuboff, 2019; Couldry & Mejiast, 2019.
[104] Van Dijck, 2020.
[105] Van Ginkel & Strijp, 2020.
[106] WRR, 2011, 2016.
[107] Kop, 2020.
[108] Moerel & Prins, 2016; Solove, 2011.
[109] Pasquale, 2016.
[110] De Poorter & Goossens, 2019.

other', the underlying mechanisms and the business models of the companies supplying applications such as facial recognition-enabled doorbells are hidden from them. Furthermore, AI is making data collection less focused: the definition of selection criteria no longer precedes the collection and processing of data since AI's great strength is its ability to reveal previously unexpected patterns in large volumes of data.

Another significant point is that not only is the volume of data collected by unfocused methods increasing, but its nature is changing as well. Whereas in previous decades fairly harmless personal details were typically harvested, often through direct interaction with data subjects (asking them to provide certain information), nowadays smart devices gather material about our activities without us even realizing. "In more and more spheres of society, information about people's physical and behavioural characteristics, such as their faces, voices and emotions, are digitally collected and processed. Such data is intimate information that may relate to private matters such as health, or that may be used to identify a person remotely."[111] So, for example, insurers can individualize their risk assessments using data about behaviour, emotion and actions.

AI and the scope it offers for facial recognition and many other new and enhanced functionalities are transforming surveillance activities. Not surprisingly, some of the applications set to be banned by Article 5 of the EU's proposed AI Act are surveillance-related, including random mass surveillance for law enforcement and social scoring purposes, as in the Chinese government's social credit system. Such a ban would rightly shift attention from AI itself to a particular field of application. However, it is questionable whether the regulation of AI at the individual application and context level is sufficient. Technology companies go to great lengths to secure people's attention because their advertising revenues depend on user interest. They are also given incentives to collaborate, since that enables the firms to create ever more precise user profiles. With the help of Google Maps, for instance, Spotify can see what music its users listen to when driving while Google itself can refine its user profiles by incorporating information about their musical tastes. Similar network effects are evident wherever organizations collect data. It is therefore more urgent than ever to consider the desirability of a high-surveillance society.

In the surveillance debate, considerable attention is rightly devoted to privacy implications and to associated issues such as banning the collection of certain types of data (biometrics, for instance), transparency and citizens' rights.[112] That debate will need to be broadened and deepened by also considering the revenue models and power of the companies engaged in surveillance.[113] The emphasis they (supposedly) place on ethical AI should therefore be subject to critical examination. A focus on self-regulation and how it deals with ethical issues risks drawing attention away

[111] Gerritsen et al., 2020.

[112] For the citizen-state relationship and AI's significance in that context, see, for example, Van Heukelom-Verhage, 2020.

[113] Häußermann & Lütge, 2021.

from underlying structural problems. Once the way something is done becomes the subject of discussion, the question of whether it should be done at all tends to be forgotten. Kate Crawford makes a similar point in her new book, Atlas of AI. She argues that a narrow definition of AI and an abstract debate about good practice serve the interests of big players by ignoring questions of power, capital and governance.[114] From this she concludes that addressing ethical issues is important, but insufficient. The focus, Crawford asserts, should be less on ethics and more on power.[115]

8.2.7 Imbalance

Data collection, use, control and quality are increasingly pertinent, therefore, to fundamental issues concerning the organization of society, the way people view that society and the behaviour and position of individuals within it.[116] Such activities also touch upon international relations, particularly the dependence of the Netherlands and Europe on other regions. Concentration of the growing volume of available data in the hands of a very small number of companies based outside the EU only serves to amplify concerns in this regard. Which brings us to the second key development influencing the embedding of AI, namely the growth of imbalance between the public and private sectors in terms of their interest in, position relative to and influence over the use of digital resources.

At present private actors are primarily responsible for the development, use and circulation of AI, largely because many recent advances have been made by the business community. But in part also because, at least until recently, the world's governments took a passive approach to regulation of the digital domain. That has led to a growing imbalance between the levels of AI use in the public and private domains, and also increased government's dependence on private actors for digitalization of the public sector. If government does not start using AI sooner rather than later, its failure to do so will result in higher opportunity costs while also further drawing private actors into the fulfilment of public tasks and increasing the public sector's dependence on them. Such developments have the potential to erode democratic accountability and ultimately diminish government's scope to determine its own policies.

For example, the government or organizations in the health or education sectors may find themselves tied to a single vendor, which thus accrues power to dictate what services are provided and on what terms. An ongoing debate concerning the Dutch government's switch from Microsoft 365 to Google Workspace illustrates this. Although Microsoft has repeatedly promised better user privacy safeguards,

[114] Crawford, 2021: 9.

[115] Crawford, 2021: 224.

[116] Hildebrandt, 2018.

the government remains very dubious about its ability to deliver. A switch to Google Workspace nevertheless appears problematic, because Google's services also entail significant privacy risks. A similar situation exists in the education sector, where G Suite for Education (a variant of G Suite Enterprise, featuring Gmail, Docs and Classroom) is used. A data protection impact assessment (DPIA) for two Dutch universities has highlighted the fact that, where metadata is concerned, Google regards itself as the sole data controller. Meaning that it alone and autonomously determines the purposes for which metadata is collected, and the means used. Furthermore, its privacy agreements state that it may unilaterally change terms and conditions regarding metadata without seeking the user's consent. Consequently, educational institutions that use Google G Suite for Education have little or no control over such data, which may relate to staff and students at any level of the educational system. The Dutch Data Protection Authority therefore advises government and educational institutions not to use Google G Suite, or to stop doing so if they already have it.

It is not only privacy or such aspects as exclusive rights and therefore control over AI applications, data and algorithms that are put at jeopardy by the imbalance just described. The public sector is also highly dependent on private companies, including some Dutch firms, in other ways, as reflected in the practical influence they exercise over the translation of policy and rules into digitalized implementation processes. A report by the Netherlands Institute for Human Rights has observed that decisions regarding the design of AI systems used by Dutch local authorities are often made not by those bodies themselves but by the system vendors. "That," the report says, "is leading to standardization and means that the way national rules are interpreted is determined by vendor companies, with implications for the practices of all the local authorities using the software in question."[117]

The Netherlands is certainly not alone in struggling with dependence on large tech firms. The German federal government has commissioned a market analysis with a view to reducing its reliance on individual software providers. Like many national administrations, the Germans use Microsoft Cloud. But they have decided to gradually reduce their dependency on this system because of Microsoft's decision to require users to migrate to its cloud-based 'software as a service' with effect from 2026. From that date Microsoft will effectively be able to dictate what applications customers are able to use. Furthermore, Germany's Federal Data Protection Authority instructed all the country's government departments to close their Facebook pages by the end of 2021. It says that these are not GDPR-compliant and that page administrators cannot therefore fulfil their accountability obligations under Article 5:2 of that regulation.[118] Facebook reportedly has no plans to make changes with a view to achieving GDPR compliance. Yet, dependence levels vary from country to country. A comparative analysis of EU members states' university ICT services, for instance, has found that the Netherlands is far more reliant on US

[117] Choi et al., 2021: 5.

[118] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, 20 May 2019.

cloud service providers than most other European nations, including Germany and France.[119]

A similar imbalance is likely to develop in relation to AI, especially where it is provided as a service. Two forms of dependency are liable to arise: dependency on AI itself and on the supporting technologies needed to make use of it (see Box 8.5). As with earlier technologies, users will probably be offered multiple choices to begin with. Over time, however, the big technology companies may well modify their offerings. The availability of AI applications could be restricted to customers who use the companies' data hosting services, for example, or subscribe to comprehensive service packages.

There are two reasons for the situation described above. First, as previously explained lock-in phenomena have become commonplace as service providers benefit from network effects. The more users they have, the more data they acquire and the more they can optimize their products. It is therefore in their commercial interests to retain users. Offering and later requiring subscriptions to integrated service packages is one way of doing this; another is to make applications incompatible with those from other providers.

Box 8.5: AI and Dependency on Foreign Vendors

High levels of dependency on foreign vendors exist in relation to both AI itself and the supporting technology.[120] The Netherlands is strong in AI research and development, but dependent in terms of access to AI-related services – especially software packages and online library services, which are largely provided by large technology companies. That is the case with both commercial and free or open-source software. For example, Google and others offer access to image-recognition models on a commercial basis. Access to AI management tools and services – including the Machine Learning Engine on Google Cloud, Azure Machine Learning Studio on Microsoft Azure, Einstein on the Salesforce cloud and IBM Watson ML – is also controlled by commercial actors. The more such software is integrated into Dutch AI products and services, the less the Netherlands is able to safeguard its civic values. A further problem is that there is no guarantee that packages like those mentioned will remain open source.

As far as supporting technology is concerned, the Netherlands is strikingly dependent on the services and products of foreign cloud providers – a situation that poses a risk to the entire AI application value chain. The Dutch market is Europe's fourth largest for cloud infrastructure. However, its supply side is dominated by overseas providers. Amazon, Microsoft and Google lead the way, with local firm KPN fourth.

[119] Fiebig et al., 2021.

[120] Based on TNO, 2021.

Second, because of their access to large volumes of data the big technology companies are well placed to develop AI and to invest heavily in it. Over the past decade hundreds of smaller AI companies have been acquired by big tech firms. Apple leads the way here with twenty acquisitions, followed by Google (fourteen), Microsoft (ten), Facebook (eight) and Amazon (seven).[121] This brings us to our third key factor: the concentration of power.

8.2.8 Concentration of Power

A small number of large US technology companies wield disproportionate influence and are dominating the development of AI. They include the vendors mentioned above, whose services government already uses. As AI becomes more deeply embedded in society, these firms are gaining ever-greater influence over many of its aspects, including political processes and democracy.[122] One person who has warned of the implications is Paul Nemitz, an adviser to the European Commission and a member of the German federal government's data ethics committee. He takes the view that because AI's impact on society is so significant, its use should be subject to democratically legitimized decision-making. In other words, Nemitz's call for further regulation is justified not by the nature of the technology itself but by the extent of its use and, as a consequence, the excessive power to shape society that is being concentrated in the hands of a small number of companies. This becomes particularly problematic when AI-based services are integrated within society's infrastructure. It is therefore pertinent to consider whether such services should in fact be considered public goods.[123]

Concentration of power has previously proven problematic in the oil and electricity industries, where initially innovative companies gradually developed into monopolies, leading governments to break them up and either convert them into public utilities or have them continue as smaller commercial entities. Immediately before World War I, GE and Westinghouse in the USA and Siemens and AEG in Germany became the largest companies in the world following mergers designed to increase their scale and access to capital. They even made a pact to divide up the global export market for the electrical technologies and machinery they produced.[124]

In recent years investigative committees in the US, the EU and the UK have turned their attention to the big technology companies, accusing them of abusing their power. The resulting reports warn that there is no longer competition in the

[121] See Nemitz & Pfeffer, 2020: 84.

[122] Poon, 2016; Moore & Tambini, 2018.

[123] Fukuyama, 2021.

[124] Bakker & Korsten, 2021: 34.

market, only competition for the market. Which leads to less innovation, less consumer choice, compromised privacy rights, a weaker press and weaker democracy. The different committees are remarkably consistent in reaching this conclusion.[125]

The Judiciary Committee of the US House of Representatives published a report on Apple, Facebook, Google and Amazon in early October 2020. Its bottom line was that these firms act as gatekeepers for certain distribution channels and abuse that position to deny others access and so maintain their own power. The committee says that a 'kill zone' exists around them, which competitors must stay clear of in order to survive. In addition, they abuse their brokerage role to increase their own dominance. Amazon, for instance, utilizes data on businesses that use its cloud services to offer its own competing products.

This report also considers AI. One of its conclusions is that voice-controlled assistants have a clear network effect. All algorithm-based applications learn and improve with use. The more they are used, the better they become. Consequently, user numbers are decisive when it comes to the success of an AI application. Furthermore, access to a combination of big data and AI is enabling the tech giants to enter new markets where the possession and use of data confers an advantage. They are already exploiting this position in the market for 'smart' devices, for example: Apple and Amazon (Alexa) sell cut-price virtual assistants that only access or recommend the vendor's own services.[126] If this technology eventually becomes the norm for online shopping, the big tech companies will already have a firm grip on the market.

Various proposals to manage such developments are now being debated. But this discussion only highlights the weakness of the instruments currently available to government. Privacy legislation, for example, is the standard vehicle for protecting personal data. Yet serious doubt now exists regarding the sufficiency and effectiveness of that regulatory framework.[127] There is also criticism of existing competition law, which is seen as overly focused on low prices as the primary indicator of consumer welfare.[128] This bias is often inappropriate in relation to technology companies, which typically provide some or all of their services without charge and often serve multiple markets simultaneously. Many commentators have therefore argued for a review of competition law and its underlying objectives.[129] One measure suggested by the US report is the break-up of the big tech companies. Advocates believe that such a move, or the threat of it, would energize the market. Similar strategies have previously been adopted in relation to IBM (whose hardware and software divisions were separated), AT&T (then the world's largest company, split into eight smaller entities) and Microsoft.[130]

---

[125] See Lancieri & Sakowski, 2020. In their report, these authors analyse 22 studies by experts and competition authorities examining competition in digital markets.

[126] Committee on the Judiciary, 2020: 124–125, 307–312, 377.

[127] Purtova, 2018: 40–81.

[128] Gerbrandy & Custers, 2018.

[129] For detail of the various proposals, see Kohlen et al., 2021.

[130] Wu, 2020: Chapter 5.

The European Commission is also active in this area,[131] taking legal action against technology companies that fail to comply with European laws and regulations and imposing increasingly severe fines over a period of years. Two new European legislative instruments have been on the table since the end of 2020: the Digital Services Act (DSA) and the Digital Markets Act (DMA). The first, in full the Proposal for a Regulation on a Single Market for Digital Services, would require large platforms to remove illegal and harmful content without delay and allow users to see how recommendation algorithms work. The DSA is hence intended to improve the liability and security rules applicable to digital platforms, services and products. The largest platforms will be subject to close, systemic scrutiny.

Through the DMA the Commission is also seeking to add new requirements to existing competition law. This measure classifies major technology platforms as 'digital gatekeepers',[132] a status that reflects both their size and their critical role within society. The long-awaited legal assignment of gatekeeper status could prevent big tech companies continuing to evade legislation. If the Commission's proposal passes into law, it will put an end to the argument that these firms fall into a special category of business not subject to various legal provisions. Spring 2022 a political agreement was reached on both proposals. Whatever final form the new legislation takes, it will fundamentally change digital competition law.[133]

In the Netherlands too, competition law has been a subject of debate for several years. In 2019, for example, Kees Verhoeven (then a member of the Dutch parliament) presented a policy proposal for the modernization of competition rules, amendment of the European and national rules to accommodate data and the definition of new criteria to demarcate the digital market and to determine companies' share of it.[134] Various ministers have submitted documents to the House of Representatives concerning the developments outlined above.[135] In 2019, for example, the undersecretary for Economic Affairs and Climate Policy presented a green paper on 'The future resilience of competition policy in relation to online platforms' (Toekomstbestendigheid mededingingsbeleid in relatie tot online platforms). This addressed a number of particular developments relating to the use of algorithms and cartels. "Because consumers' preferences and financial status can increasingly be accurately gauged using data and algorithms," it stated, "individualized price discrimination may develop."[136] Self-teaching algorithms might thus one day even be able to form cartels without human intervention. Building on this perspective, a bill providing for the modernization and better enforcement of consumer protection rules was subsequently tabled.

[131] Crémer et al., 2019; Kohlen et al., 2021.

[132] See, for example, the EU's proposed Digital Markets Act (DMA).

[133] Chavannes et al., 2021: 17–20.

[134] Kamerstukken 2018/19, 35134, no. 2.

[135] Kamerstukken II 2019/20, 26643, no. 672.

[136] Appendix to the memorandum from the undersecretary for Economic Affairs and Climate Change, 17 May 2019: 6.

The Netherlands, Germany and France have since collectively proposed that all mergers and takeovers by large digital platforms performing a gatekeeper role should be subject to review by an EU regulator.[137] The mechanism they suggest would supplement and reinforce the supervisory and other provisions of the DMA. Those provisions include an obligation to share data, interoperability requirements and a ban on digital gatekeepers favouring their own products or services in rankings.

Various other proposals to downsize or reduce the influence of big technology companies are under consideration, such as obliging them to make their services and data available to others.[138] Matters being discussed in this context include interoperability and platform neutrality. The proposal in the latter regard is to follow the existing principle of net neutrality, whereby internet providers are not allowed to price content differently for different users. One fairly recent suggestion is an approach that combines regulation with technological solutions; this would involve the use of 'middleware', an intermediate technological layer inserted above a platform to ensure fair competition.[139]

The companies providing the middleware would have the task of editing news and information, for which they would have their own algorithms and would be able to develop their own profiles. Users would then be able to choose between different information channels, whilst the rapid and extensive dissemination of misinformation and fake news picked up by a single platform's algorithms would be counteracted. The hope is that this approach can address the problems currently caused by platform companies when it comes to spreading fake news and filtering illegal content. It could be difficult to implement such a far-reaching change to the digital infrastructure, though: that would require a new revenue model, co-operation on the part of platform operators and a technical framework that is both compatible with the architectures of the various platforms and enables market diversity.

Exactly how the proposals made by the European Commission and others will eventually be implemented remains uncertain. What is clear, however, is that they will have a major impact on AI's integration into society. So, as well as having work to do when it comes to the questions the technology raises in respect of current regulatory frameworks, government must also invest in structuring the way we deal with AI itself in order that civic values are properly safeguarded in the long term. Embedding AI within society is thus in essence an issue related to the wider 'digital living environment' and as such a fundamental issue that government must address as a matter of urgency. If it fails to do, government may find its scope for action

[137] Proposal of 26 May 2021 (Le Maire, Altmaier & Keijzer). Cf. Considerations of France and the Netherlands regarding intervention on platforms with a gatekeeper position, 15 October 2020.

[138] Graef & Prüfer, 2021.

[139] Fukuyama et al., 2021.

restricted by others taking the lead in determining how and for what purposes AI is used, or by the public losing trust in or even rejecting AI.

Key Points – AI Regulation and the Digital Living Environment
- Government's role in the regulation of AI will inevitably increase as the technology enters more widespread use and situations arise that require intervention.
- The timing of government intervention is crucial. If it waits too long, AI may become embedded in ways that are inconsistent with or fail to serve civic values.
- A system technology like AI requires a legislative agenda that addresses not only issues associated with the technology and its use, but also its broader societal effects.
- Mass surveillance, extreme dependency on private actors and power concentration represent threats to civic values in the context of AI's societal integration, and therefore require urgent government intervention.

8.3 In Conclusion

In this chapter we have considered the overarching task of government regulation. We perceive that task as having two dimensions associated with the systemic nature of AI. The first is its pervasiveness, which is such that it will require new or adapted regulatory frameworks in many areas.

Generally speaking, we are at a very early point in the process of AI's societal integration or embedding. It might easily be supposed, therefore, that government should refrain from intervening at this stage and instead monitor developments until 'the time is right'. Such a policy is undesirable, however, because of the major impact AI is likely to have. If government wishes to retain its capacity for significant and effective intervention now and in the longer term, particularly with a view to safeguarding civic values, it must be vigilant and start preparing now for the more forceful role it will inevitably have to play in due course. Fortunately, this process of preparation is already under way in certain spheres, both in the Netherlands and in the European Union. Against that background it is important that government be aware of the various issues surrounding the regulation of AI. It must also commit to structural investment in the collection and collation of signals regarding the opportunities and risks associated with the societal embedding of AI, otherwise its ability to make appropriate changes or define new rules – or to do so in good time – will be seriously curtailed.

The second conclusion of this chapter is that as AI becomes more deeply embedded in society, government's regulatory task will inevitably increase. At the same time, the nature of the issues it faces will change as increasing use of AI gives rise to second and third-order problems. It will then be necessary to address problems

posed not only by the technology per se, but also by the extent of its use and the scale of its effects. Which in turn will require active management of the wider digital living environment into which AI is ultimately going to be embedded. Precedents for this kind of approach have previously been set in other fields where developments have fundamentally affected and reorganized society. In the period ahead, government must therefore focus on converging currently separate policy portfolios and lines of development so that they are viewed as elements of a more comprehensive design challenge.

It should also be recognized that AI is entering a society where data is already collected on a large scale, where digital products, services and infrastructures are made available largely by private actors and where the leading AI developers occupy a dominant position in the global internet economy. That context is bound to have a major bearing on the way AI is eventually used in society, by whom and for what purposes. If government wishes to retain its ability to influence developments, it must act now. Delay is not only undesirable, it is unnecessary. Numerous research reports, other documents and plans are already available, which government can use to support and guide its interventions. The important thing now is to be energetic in converting those plans into regulatory action.

References

AFM en DNB. (2019). *Artificiële Intelligentie In De Verzekeringssector Een Verkenning*. Autoriteit Financiële Markten, De Nederlandsche Bank. Available at: https://www.afm.nl/~/profmedia/files/rapporten/2019/afm-dnb-verkenning-ai-verzekeringssector.pdf?la=nl-nl

Algemene Rekenkamer. (2021). *Aandacht voor Algoritmes*. Algemene Rekenkamer.

Bakker, S., & Korsten, P. (2021). *Artificiële Intelligentie Als Een general purpose technology: Strategische Belangen Van Verantwoorde Inzet In Historisch Perspectief* (WRR Working Paper nr. 41). Wetenschappelijke Raad voor het Regeringsbeleid. Available at: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Bennett, M., & Gollan, L. N. (2015). *The illusion of Newness: The importance of history in understanding the law-technology interface* (UNSW Law Research Paper, 2015-71). University of New South Wales.

Bennett Moses, L. (2007a). Why have theory of Law and Technological Change? *Minnesota Journal of Law, Science & Technology, 8*, 589–606.

Bennett Moses, L. (2007b). Recuring Dilemmas: The Law's race to keep up with technological change. *University of Illinois Journal of Law, Technology and Policy, Fall*, 239–285.

Bernstein, G. (2006). The Paradoxes of technological diffusion: Genetic discrimination and internet privacy. *Connecticut Law Review, 39*(1), 243–297.

Bertolini, A. (2020). *Artificial Intelligence and Civil Liability. Onderzoek in Opdracht van de Commissie Juridische Zaken van het Europese Parlement*. European Parliament. Available at: http://www.europarl.europa.eu/supporting-analyses

Bijker, W. (1995). *Of Bicycles, Bakelites, and Bulbs. Toward a theory of sociotechnical change*. MIT Press.

Bijlsma, M., Overvest, B., & Straathof, B. (2016). *Marktordening Bij Nieuwe ICT-Toepassingen. Vroegtijdig Ingrijpen Nodig* (CPB Policy Brief 2016/09). Centraal Planbureau. Available at: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-2016-09-Marktordening-bij-nieuwe-ICT-toepassingen.pdf

Black, J., & Murray, A. (2019). Regulating AI and machine learning: Setting the regulatory Agenda. *European Journal of Law and Technology, 10*(3). Available at: https://ejlt.org/index.php/ejlt/article/download/722/980

Brown, R. (2021). Property ownership and the legal personhood of artificial intelligence. *Information and Communications Technology Law, 2*, 208–234.

Brownsword, R., en Goodwin, M. (2012). *Law and the technologies of the twenty-first century*. Cambridge University Press.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.

Cath, C., Wachter, S., Mittelstadt, B., Toledo, M., & Floridi, L. (2018). Artificial intelligence and the 'Good society'. The US, EU, and UK approach. *Science and Engineering Ethics, 24*, 505–528.

Centrale Raad van Beroep. (2019). ECLI:NL:CRVB:2019:1737, ruling 15 May 2019. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:CRVB:2019:1737

Chavannes, R., Strijbos, A., & Verhulst, D. (2021). Kroniek Recht en Technologie. *Nederlands Juristenblad, 2021*(16), 1350–1370. Available at: https://blog.chavannes.net/2021/04/kroniek-technologie-en-recht-2021/

Chiusi, F., Fischer, S., Kayser-Bril, N., & Spielkamp, M. (2020). *Automating Society Report 2020*. AlgorithmWatch. Available at: https://www.ivir.nl/publicaties/download/Automating-Society-Report

Choi, W., van Eck, M., & Hukshorn, H. (2021). *Hoe Gemeenten Besluiten Over Algoritmen En Mensenrechten*, onderzoek in opdracht van het College voor de Rechten van de Mens. Hooghiemstra en partners.

Coeckelbergh, M. (2020). AI for climate: Freedom, Justice, and other ethical and political challenges. *AI and Ethics, 1*, 67–72.

Coglianese, C., & Lehr, D. (2019). Transparancy and algorithmic governance. *Administrative Law Review, 71*(1), 12–57.

Committee on the Judicary. (2020). *Investigation of competition in digital markets*. *Majority Staff Report and Recommendations*, Subcommitte on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary. Available at: https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf

Couldry, N., & Mejiast, U. A. (2019). *The Costs of Connection*, *How Data Is Colonizing Human Life and Appropriating It for Capitalism.* Stanford University Press.

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Crémer, J., De Montjoye, Y., & Schweitzer, H. (2019). *Competition policy for the digital era*. Europese Commissie. Available at: http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

De Poorter, J., & Goossens, J. (2019). Effectieve Rechtsbescherming Bij Algoritmische Besluitvorming In Het Bestuursrecht. *Nederlands Juristenblad, 44*, 3303–3312.

De Ree, M. (2021, April 29). Onderzoek Naar Eerlijke En Uitlegbare Algoritmen. *CBS.nl*. Available at: https://www.cbs.nl/nl-nl/corporate/2021/17/onderzoek-naar-eerlijke-en-uitlegbare-algoritmen

Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. (2019). Facebook-Auftritte von öffentlichen Stellen des Bundes, brief d.d. 20 May 2019, 61924/2021. Available at: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Rundschreiben/Allgemein/2021/Facebook-Auftritte-Bund.pdf?blob=publicationFile&v=2

European Commission. (2021a). *2030 Digital Compass: the European way for the Digital Decade*. Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118

European Commission. (2021b). *Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Data Protection Board and European Data Protection Supervisor. (2021). *Joint Opinion 5/2021 on the proposal for a proposal for the regulation of the European Parliament and of the council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act), 28 June 2021*. EDPS. Available at: https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

European Parliament. (2020). European Parliament resolution of 20 October 2020 with recommendations to the commission on a civil liability regime for Artificial Intelligence (2020/2014(inl)). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html

European Union Agency for Fundamental Rights. (2020). *Getting the future right. Artificial Intelligence and fundamental rights*. Publications Office of the European Union. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

Fiebig, T., Gürses, S., Gañán, C., Kotkamp, E., Kuipers, F., Lindorfer, M., Prisse, M., & Sari, T. (2021). *Heads in the Clouds: Measuring the implications of Universities migrating to Public Clouds*. Available at: https://arxiv.org/abs/2104.09462

Fierens, M., van Gool, E., & De Bruyne, J. (2021). De Regulering Van Artificiële Intelligentie (Deel 1) – Een Algemene Stand Van Zaken En Een Analyse Van Enkele Vraagstukken Inzake Consumentenbescherming. *Rechtskundig Weekblad, 84*(25), 962–980.

Floridi, L. (2021). The European Legislation on AI: A brief analysis of its philosophical approach. *Philosophy and Technology, 34*, 215–222. Available at: https://link.springer.com/article/10.1007/s13347-021-00460-9

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). Available at: https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707.

Franken, H. (1993). Kanttekeningen Bij Het Automatiseren Van Beschikkingen. In H. Franken, I. TH. M. Snellen, J. Smit, and A. W. Venstra *Beschikken en automatiseren* (pre-advies Vereniging voor Administratief Recht, pp. 7–50). Samsom H.D. Tjeenk Willink.

Fukuyama, F. (2021). Making the Internet safe for democracy. *Journal of Democracy, 32*(2), 37–44.

Fukuyama, F., Richman, B., Goel, A., Schaake, M., Katz, R., & Melamed, D. (2021). *Report of the working group on platform scale*. Stanford University. Available at: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/platform_scale_whitepaper_-cpc-pacs.pdf

Gerbrandy, A., & Custers, B. (2018). Algoritmische Besluitvorming En Het Kartelverbod. *Markt en Mededinging, 3*, 101–109. Available at: https://www.bjutijdschriften.nl/tijdschrift/marktenmededinging/2018/3/MenM_1387-6236_2018_021_003_002

Gerritsen, J., Hamer, J., Kool, L., & Verhoef, P. (2020). Beter beschermd tegen biometrie. *Beleid en Maatschappij, 47*(4), 451–466.

Giesen, I. (2007). *Alternatieve Regelgeving En Privaatrecht*. Kluwer.

Goossens, J., Hirsch Ballin, E., & van Vugt, E. (2021). Algoritmische Beslisregels Vanuit Constitutioneel Oogpunt. Tweedeling Tussen Algemene Regels En Concrete Toepassing Onder Druk. *Tijdschrift voor constitutioneel recht, 12*(1), 4–19.

Graef, I., & Prüfer, J. (2021). Governance and data sharing: A law and economics proposal. *Research Policy, 50*(9), 104330.

Greene, D., Hoffman, A., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A critical assessment of the movement for ethical Artificial Intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 2122–2131).

Hage, J. (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law, 25*(3), 255–271.

Hagedoorn, P. (2021). *The digital challenge for Europe*. Peter Hagedoorn.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*, 99–120.

Häußermann, J., & Lütge, C. (2021). Community-In-The-Loop: Towards pluralistic value creation in AI, or—Why AI needs business ethics. *AI and Ethics, 2*, 1–22.
Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society, 34*(1), 1–14.
Helwig, P. (2020). Rekenen en Rekenschap. Algoritmes en de Archiefwet. *Tijdschrift voor Toezicht, 1*, 54–59.
High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines For Trustworthy AI*. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics. *The University of Toronto Law Journal, 68*(5), 12–35.
Hildebrandt, M., & Gutwirth, S. (reds.). (2008). *Profiling the European Citizens*. Springer.
Hirsch Ballin, E. (2021). *Mensenrechten Als Ijkpunten Van Artificiële Intelligentie* (WRR Working Paper nr. 46). Wetenschappelijke Raad voor het Regeringsbeleid.
Hoge Raad. (1921). NJ 1921, 564 (Elektriciteitsarrest) ECLI:NL:HR:1921:186, ruling 23 May 1921. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:1921:186
Hoge Raad. (2018). ECLI:NL:HR:2018:1316, ruling 17 August 2018. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:2018:1316
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.
Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture and Society, 39*(2), 238–258.
Kamerstukken II 2018/2019, 26643, nr. 570. (2018, October 9). *Brief Van De Minister Voor Rechtsbescherming*, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-570.pdf
Kamerstukken II 2019/2020, 26643 nr. 641. (2019, October 8). *Brief Van De Minister Voor Rechtsbescherming*, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-641.pdf
Klincewicz, M., & Lily, F. (2020). *Consequences of unexplainable machine learning for the notions of a trusted doctor and patient autonomy*. Paper presented at the 32nd International Conference on Legal Knowledge and Information Systems, Madrid, Spain, 2020.
Kohlen, J., van de Sande, M., & Cox, M. (2021). 'Rebooting' Het Mededingingsrecht – Ook Het Mededingingsrecht Ontsnapt Niet Aan De Digitale Transitie. *Markt en Mededinging, 1*, 6–14.
Koops, B. (2006). Should ICT regulation be technology-neutral. In B. J. Koops, C. Prins, M. Schellekens, and M. Lips (reds.), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners* (pp. 77–108). T.M.C. Asser Press.
Koops, B. J., Lips, M., Nouwt, J., Prins, C., & Schellekens, M. (2006). Should selfregulation be the starting point?. In B. J. Koops, C. Prins, M. Schellekens, en M. Lips (reds.), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners* (pp. 109–149). T.M.C. Asser Press.
Kop, M. (2020). AI and intellectual property: Towards an articulated public domain. *Texas Intellectual Property Law Journal, 28*(1), 297–341.
Krupiy, T. (2020). A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective. *Computer Law & Security Review, 38*, 1–25.
Kulk, S. (2020). 'Platformaansprakelijkheid – Van 'Notice And Takedown' Naar Algoritmisch Toezicht', *Nederlands tijdschrift voor Europees recht*, nr. 5/6*:* 132–140.
Kulk, S., & van Deursen, S. (2020). *Juridische Aspecten Van Algoritmen Die Besluiten Nemen. Een Verkennend Onderzoek*. Wetenschappelijk Onderzoek- en Documentatiecentrum.
Lancieri, F., & Sakowski, P. (2020). Competition in digital markets: A review of expert reports. *Stanford Journal of Law, Business & Finance, 26*, 65.

Le Maire, B., Altmaier, P., & Keijzer, M. (2021). *Strengthening the Digital Markets Act and its Enforcement*. Non-paper DMA. Available at: https://www.rijksoverheid.nl/documenten/publicaties/2021/05/26/non-paper-dma

Lessig, L. (2006). *Code and other Laws of Cyberspace, Version 2.0*. Basic Books.

Marchant, G. (2011). The Growing Gap between emerging technologies and the Law. In G. E. Marchant, B. Allenby, and J. Herkert (reds.), *The growing gap between emerging technologies and legal-ethical oversight the pacing problem*. (pp. 19–33). Springer.

Meijer, A., Grimmelikhuijsen, S., & Bovens, M. (2021). De Legitimiteit Van Het Algoritmisch Bestuur. *Nederlands Juristenblad, 18*, 1470–1478.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*, 501–507.

Moerel, L., & Prins, C. (2016). *Privacy for the Homo digitalis: Proposal for a new regulatory framework for data protection in the light of big data and the Internet of Things*. Wolters Kluwer. Available at: https://doi.org/10.2139/ssrn.2784123

Moore, M., & Tambini, D. (reds.). (2018). *Digital dominance: Power of Google, Amazon, Facebook, and Apple.* Oxford University Press.

Mols, B. (2019). *Internationaal AI-beleid. Domme data, slimme computers en wijze mensen.* WRR Working Paper.

Nemitz, P. (2018). Constitutional Democracy and Technology in the Age of Artificial Intelligence. *Philosophical Transactions of the Royal Society A, 376*(2133), 20180089. Available at: https://doi.org/10.1098/rsta.2018.0089

Nemitz, P., & Pfeffer, M. (2020). *Prinzip mensch. Macht, Freiheit und Democratie im Zeitalter der Künstlichen Intelligenz*. Dietz.

Passchier, R. (2020). *Artificiële Intelligentie En De Rechtsstaat. Over Verschuivende Overheidsmacht, Big Tech En De Noodzaak Van Constitutioneel Onderhoud*. Boom.

Pasquale, F. (2016). *The Black Box society: The secret algorithms that control money and information.* Harvard University Press.

Poon, M. (2016). Corporate capitalism and the growing power of big data: Review essay. *Science, Technology, & Human Values, 41*, 1088–1108.

Prins, C. (2018). Urgenda en Digitalisering. *Nederlands Juristenblad*, 2018/1098, 22.

Purtova, N. (2018). The Law of everything. Broad Concept of Personal and future of EU Data Protection Law. *Law, Innovation and Technology, 1*, 40–81.

Raad van State. (2018). *Ongevraagd Advies Over De Effecten Van De Digitalisering Voor De Rechtsstatelijke Verhoudingen*. Raad van State. W04.18.0230/I. Available at: https://www.raadvanstate.nl/@112661/w04-18-0230/

Raad van State. (2020a). *Jaarverslag 2020*. Raad van State.

Raad van State. (2020b). *Ongevraagd advies over ministeriële verantwoordelijkheid*. Raad van State. Available at: https://www.raadvanstate.nl/@121396/advies-ministeriele-verantwoordelijkheid/

Raad van State (2021). *Digitalisering. Wetgeving en bestuursrechtspraak*. Raad van State. Available at: https://www.raadvanstate.nl/publish/library/13/digitalisering_wetgeving_en_bestuursrechtspraak.pdf

Rechtbank Amsterdam. (2019). ECLI:NL:RBAMS:2019:4799, ruling 4 July 2019. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2019:4799

Reed, C. (2018). How should we regulate Artificial Intelligence? *Philosophical Transactions of the Royal Society A*, *376,* 20170360. Available at: https://doi.org/10.1098/rsta.2017.0360

Select Committee on Artificial Intelligence. (2018). *AI in the uk: Ready, willing and able?* (Report of Session 2017-19, hl Paper 100). Authority of the House of Lords. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Smits, J. (2015). Wetgeving En Andere Normenstelsels: Zes Aanwijzingen Aan De Nederlandse Wetgever. *RegelMaat, 30*(5), 357–359.

Smuha, N. (2019). From A 'Race To AI To A 'Race to AI regulation': Regulatory competition for Artificial Intelligence. *Law Innovation and Technology, 13*(1), 57–84.

Smuha, N., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). *How The Eu can achieve legally trustworthy AI: A response to the European Commission's proposal for an Artificial Intelligence Act*. Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3899991_code3594902.pdf?abstractid=3899991&mirid=1

Solove, D. (2011). *Nothing to hide. The False tradeoff between privacy and security*. Yale University Press.

Staatscourant-2017-69426. (2017). *Besluit Van De Minister-President, Minister Van Algemene Zaken, Van 22 December 2017, Nr. 3215945, Houdende Vaststelling Van De Tiende Wijziging Van De Aanwijzingen Voor De Regelgeving*. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Stikker, M. (2019). *Het Internet is Stuk*. De Geus.

Svantesson, D. (2020). Is International Law ready for the (Already Ongoing) digital age? Perspectives from Private and Public International Law. In M. Busstra, W. Theeuwen, Y. Buruma, and D. Svantesson (reds.), *International Law for a Digitalised World* (Royal Netherlands Society of International Law, Collected Papers nr. 147) (pp. 113–155). T.M.C. Asser Press.

Tijdschrift voor Toezicht. (2020). Aflevering 1. *Boom Juridisch Tijdschriften*. Available at: https://www.bjutijdschriften.nl/tijdschrift/tijdschrifttoezicht/2020/1

TNO. (2021). *Het Technologische Ecosysteem Van AI In Nederland* (WRR Working Paper nr. 47). Wetenschappelijke Raad voor het Regeringsbeleid.

van Asselt, M., Voss, E., & Fox, T. (2010). Regulating technologioes and the uncertainty Paradox. In M. Goodwin, E. Koops, and R. Leenes (reds.), *Dimensions of technology regulation* (pp. 261–286). Wolf Legal Publishers.

van Dijck, G. (2020). Algoritmische Risicotaxatie Van Recidive. Over De Oxford Risk of Recidivism Tool (*OXREC*), Ongelijke Behandeling En Discriminatie In Strafzaken. *Nederlands Juristenblad, 25*, 1784–1790.

van Eck, M., Zouridis, S., & Bovens, M. (2018). Algoritmische Rechtstoepassing In De Democratische Rechtsstaat. *Nederlands Juristenblad, 40*, 3008–3017.

van Ginkel, J., & Strijp, P. (2020). Van Beleidscyclus Naar Datacyclus. *iBestuur*. Available at: https://ibestuur.nl/podium/van-beleids-naar-datacyclus

van Gool, E., de Bruyne, J., & Fierens, M. (2021). De Regulering Van Artificiële Intelligentie (Deel 2). Een Analyse Van Buitencontractuele Aansprakelijkheid. *Rechtskundig Weekblad, 84*(26), 1003–1024.

van Heukelom-Verhage, S. (2020). Maatwerk Bieden In Een Gedigitaliseerde En Datagedreven Samenleving #Hoedan?. In L. van den Berge, M. Vermaat, M. Lurks, N. van Renssen en S. van Heukelom-Verhage (reds.), *Maatwerk in het bestuursrecht.* Boom.

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*, 1–6.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International, 22*(4), 97–112.

Verhey, L. & Verheij, N. (2005). De Macht Van De Marktmeesters: Markttoezicht In Constitutioneel Perspectief. In A. A. Rossum, L. F. M. Verhey, and N. Verheij (red.) *Toezicht* (Vol. 135, Handelingen der Nederlandse Juristen-Vereniging) (pp. 135–332), Kluwer.

Vetzo, M., Gerards, J., & Nehmelman, R. (2018). *Algoritmes en Grondrechten*. Boom Juridisch.

Wallach, W. (2015). *A dangerous master. How to keep technology from Slipping beyond our control*. Basic Books.

Winner, I. (1983). Techne and Politeia: The technical constitution of society. In P. Durbin en F. Rapp (reds.), *Philosophy and Technology* (Boston Studies in the Philosophy of Science) (pp. 97–111). Springer.

Wolswinkel, J. (2020). *Willekeur of Algoritme? Laveren Tussen Analoog En Digitaal Bestuursrecht*. Tilburg University.

WRR. (2008). *Onzekere Veiligheid. Verantwoordelijkheid Rond Fysieke Veiligheid*. Amsterdam University Press.

WRR. (2011). *iOverheid*. Amsterdam University Press.
WRR. (2013). *Naar Een Lerende Economie*. Wetenschappelijke Raad voor het Regeringsbeleid.
WRR. (2016). *Big data In Een Vrije En Veilige Samenleving*. Amsterdam University Press.
Wu, T. (2020). *The Curse of bigness. How corporate giants came to rule the world*. Atlantic Books.
Yeung, K., & Lodge, M. (reds.). (2019). *Algorithmic regulation*. Oxford University Press.
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Chapter 9
International Positioning

The final overarching task we have identified concerns a country's international positioning in the field of AI. This task is slightly different from the previous four because it plays out at a level influenced to some degree by all of them. Myths about AI also exist in the international domain, which are addressed in the international media and by international companies and research institutes. Several issues of contextualization also have an international dimension. For example, the global discussion about the rollout of 5G and European ambitions to develop a common data infrastructure (such as Gaia-X) affect the development of the technical ecosystem for AI. Stakeholder engagement can affect the local or national situation, but also has an international dimension; for example, in the role played by global scientific associations and NGOs. To a large extent the international arena is also where applicable regulation is implemented, such as treaties that govern dangerous applications of technology, ethical guidelines or the EU's ambitions to regulate AI.

Even though there is a strong interrelationship with the tasks previously discussed, it makes sense to address the international dimension of AI's integration into society separately. Firstly, because it involves a specific category of actors. Countries are represented in international bodies by specific parties who negotiate and co-operate with other international players. These are not only state actors but also international organizations, multinationals and even individuals.

There is also another reason to look separately at the international field. That is because of two issues specific to it. The first has to do with the competitiveness or earning power of a given country. What competitive advantage does a nation have? Can its position be strengthened? How does this relate to the capacities of other countries? The second, also with an eminently international dimension, is security.[1] The most extreme example of this issue is war. New technologies have a great impact on how armed conflicts are fought and how they can be won. In this context AI is often discussed in relation to autonomous weapons, although its influence on

[1] We say an 'eminently international dimension' because, of course, security is also a domestic issue and the two are increasingly intertwined. See WRR, 2017.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_9

warfare is in fact much broader. Security also plays a role in less extreme situations, becoming implicated in such activities as foreign influence and the export of ideologies, as well as sabotage and industrial espionage. Issues of this kind arise not only between countries with hostile relations, but even between allies. One example here is the dimension of 'flow security', which is about safeguarding all the flows of all manner of goods: food and medicines, for instance, but also data, capital and people.[2]

The issues of competitiveness and security can also become intertwined. In the discussion about 5G technology the Chinese company Huawei is seen not only as an economic competitor but above all as a security risk. The economic arguments in the US-China trade conflict go hand in hand with questions of national security.[3] The debate in Europe about the power of America's 'big tech' companies was also initially about competitiveness, but is now increasingly being interpreted in terms of the demand for strategic autonomy and digital sovereignty.[4] A growing list of publications on 'geo-economics' emphasizes the strong connection between economics and competitiveness on the one hand and geopolitics and security on the other.[5] In this chapter we first examine the two issues separately and then discuss what connects them. Figure 9.1 reveals how those themes relate to competitiveness, national security and the underlying geo-economic situation.

A final reason to consider the international field separately is the question of the level at which some tasks should be addressed. A number of domains, such as the global financial system, are so internationally intertwined that certain challenges cannot be addressed adequately at the national level. In our region the European Union has become the level at which rules and agreements in many areas are established, but in others global organizations like the United Nations (UN) or alliances such as NATO play a more important role. In part therefore, the international field therefore needs to be examined separately to establish the best level at which to tackle certain issues.

The central question in this overarching task is, 'what is our international position?' We first discuss international positioning in relation to competitiveness (Sect. 9.1), then specifically examine national AI capacities, the phenomenon of AI strategies and the often-associated idea of a global race to establish AI dominance. After that we look at international positioning in relation to security (Sect. 9.2). As well as examining the rise of autonomous weapons, we also examine other ways in which AI can influence warfare. Finally, we address broader security issues between countries and the rise of a 'digital dictatorship'.

---

[2] WRR, 2017.

[3] Daniel Drezner reveals how, under Donald Trump, the economy was deployed more emphatically as a weapon in strategic situations (Drezner, 2019).

[4] Until recently the concept of strategic autonomy was used mainly by France to describe the military domain and by India during the Cold War. In recent years a host of European politicians, from Emanuel Macron to Peter Altmaier, have been applying it to the domain of digitalization (Timmers, 2019).

[5] The term 'geo-economy' was coined in Luttwak, 1990. A recent review of the literature published since then can be found in Scholvin & Wigell, 2018.

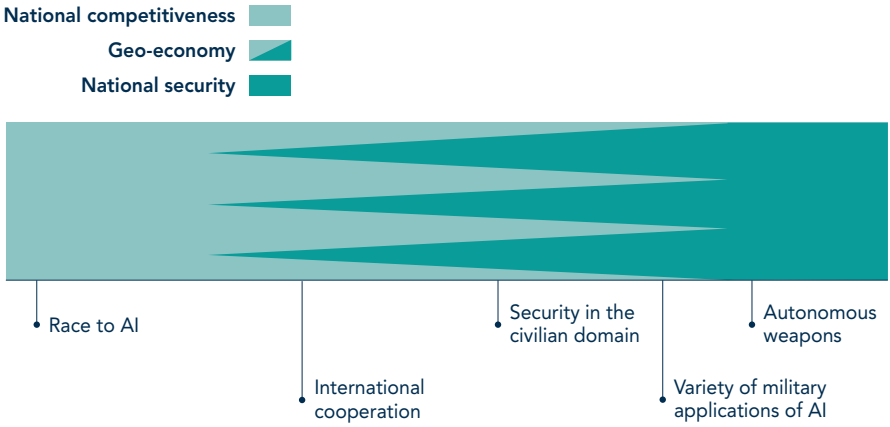


Fig. 9.1 Issues related to national competitiveness, national security and geo-economics

9.1 AI and Competitive Advantages

9.1.1 AI Capacities

Like previous technological revolutions, AI will change the relative competitive positions of countries and there are great expectations about the economic value it could generate. In Chap. 3 we mentioned PwC's prediction that AI will contribute US$15.7 trillion to the global economy in 2030.[6] So what does the international economic domain currently look like? As discussed in previous chapters, AI is a complex phenomenon with various dimensions. As such it is impossible to explain this domain based on a single criterion. There are indices for the economic value of AI activities, the number of AI actors per country and the number of AI patents, and there is now also an AI index.[7] None, however, explains the entire picture. Authors Jeffrey Ding and Kai-Fu Lee have attempted to come up with a classification that can be used to estimate a country's AI capacities.[8] If we combine the various approaches, we arrive at five relevant dimensions.

1. **The quality of fundamental research.**
2. **The availability of data.**
3. **The required hardware.**
4. **A dynamic private sector to commercialize the technology.**
5. **An enabling government.**

[6] Rao & Verweij, 2017.

[7] Stanford University publishes an annual AI Index Report on global trends in this domain.

[8] Ding, 2018; Lee, 2018.

The first three of these have been discussed in Chap. 6 as aspects of the technical ecosystem (requirements for a functioning AI ecosystem). The private-sector dimension includes both large technology companies and innovative start-up culture, as well as the AI investments made by major companies in other sectors. A government can enable development through investment, but also by implementing legislation that is clear at the very least and perhaps even creates room for experimentation.

Taking these five dimensions as a starting point, it becomes clear that the US and China are the two great world leaders. Both are strong in all the dimensions. Both have access to huge amounts of data, due to a combination of their sheer size and relatively lenient legislation. Both also have a wide variety of technology companies: the US has big tech in Silicon Valley, with firms such as Alphabet, Amazon, Facebook, Microsoft and Apple, while China boasts giants like Baidu, Alibaba, Tencent (a trio sometimes abbreviated as 'BAT') and Huawei. In addition, both have enterprises specializing in the application of AI that are rapidly growing into major market players, such as Uber and Netflix in the US and Bytedance and Hikvision in China.

The adoption of AI on consumer platforms in China is very high; voice recognition software is widely used and consumers can make payments using facial recognition.[9] In the field of fundamental research the US is still in the lead but China is well on the way to narrowing that gap.[10] A study of research institute citations between 2012 and 2016 showed that China is currently in second place behind the US, and that Tsinghua University scored even higher than Stanford University for the total number of AI citations in the 'elite institutes' category.[11] Another study of academic papers presented at major AI conferences showed a decrease in the proportion of authors from American institutes (from 41% in 2012 to 34% in 2017) and an increase in Chinese authors from 10% to 24%.[12]

Anecdotal information also confirms this trend. The Chinese start-up Face++ dominated an international image recognition competition in 2017, beating teams from Google, Microsoft and Facebook. In the field of speech recognition China's iFlytek has overtaken America's Nuance and both companies can now be called international leaders. Former Google CEO Eric Schmidt warned in 2017 against complacency towards Chinese AI capabilities and predicted that the country would match the US in five years.[13]

---

[9] Lee, 2018: 118.

[10] Stanford's most recent AI index reveals that China passed the EU in 2017, both in the number of scientific publications and the percentage of total publications (it had already surpassed the US in 2008). China passed the EU for citations of scientific articles in 2016 and the US in 2019. However, both the US and the EU still lead China when the 'weight' of the citations is taken into account (Zhang et al., 2021: 18–30).

[11] Another study examined citations in the top 100 AI journals and conferences between 2006 and 2015. This revealed that the proportion of papers by authors with Chinese names increased from 23.2% to 42.8% (Lee, 2018: 89).

[12] Leung, 2019: 250.

[13] Lee, 2018: 90, 105.

Jeffrey Ding and Kai-Fu Lee have different ideas about their respective competitive positions, however. According to Ding, the US is still the world leader by some distance and will remain so for the foreseeable future. Lee on the other hand is betting on China. Their divergence has to do with the weight they accord the various domains involved. Ding explains that the US has a particularly strong position in the field of hardware, most notably specialized chips. China remains dependent on these physical components and the US is making it harder for it to gain access to them. Lee, by contrast, points to the fact that China has access to far more data and is relatively unhindered in what it does with it, a factor he believes will be of paramount importance in the application phase of AI. He also points out the role of government; China has a far more ambitious AI strategy than the US.

China's *New Generation Artificial Intelligence Development Plan* (AIDP) was published in July 2017. It sets out the nation's precise goals for the nature and scale of AI in the coming years: to be on par with the most advanced countries in the field by 2020, to be the world leader in certain areas by 2025 and to be the world's primary AI innovator by 2030. In addition to these ambitions, the plan also prompts local authorities to establish their own plans and funds and sets out the key policy instruments that will be deployed to achieve the goals.

The emphasis on establishing technical standards is striking: this point is mentioned no fewer than 24 times in the AIDP. We return to standardization later in this chapter. The Chinese plan further emphasizes the importance of international cooperation in regulation and ethical standards for AI.[14] Meanwhile, the technology is now being used widely in China. The tech platform Tencent has launched a healthcare system called Miying to assist medical professionals with diagnoses. The police use facial recognition, and even software that analyses body positions. Funds are available for applications in education and business. In Hangzhou Alibaba is building City Brain, an AI system to improve traffic management and the response times of emergency services.[15]

It is impossible to say whether Ding or Lee will be right. What is clear, however, is that these two countries are undisputedly the world's AI superpowers. So what about the rest of the world? As a whole, the EU is not in a bad position. It is ahead of China and comparable with the US in fundamental research. There is less data available here, though, in part because of the national diversity within the EU but also due to stricter legislation.[16] The EU has a strong position in hardware for AI. European countries are dependent on the US for specialized chips but have unrestricted access to them thanks to their friendly relations. Furthermore, the EU is making progress regarding government support. In addition to national AI strategies, there is now also one at the European level.[17] Total investment remains

[14] Ding, 2019: 43–44.

[15] Creemers, 2019: 130.

[16] The European Commission has since formulated far-reaching ambitions for data sharing, including substantial financial resources and a Data Governance Act. A Data Act is also in the pipeline.

[17] European Commission, 2018.

relatively limited but there is a growing momentum in the EU to promote AI as a strategic technology. The COVID-19 pandemic also seems to be contributing to this momentum. Twenty percent of the €670 billion allocated to the EU recovery plan has been earmarked for the wider development of digitalization and AI will inevitably benefit from this. The same applies to the funds being allocated for the EU4Health programme, the Connecting Europe Facility – Digital (to finance infrastructure) and the Digital Europe Programme.[18]

The EU's biggest weakness lies in the business environment for AI. Companies in other sectors, such as infrastructure and energy, are developing their AI capacities. There are the traditional technology companies as well, like SAP, Dassault, ASML and TomTom, and a number of tech start-ups have also grown to become major players, amongst them Spotify, Zalando and Adyen. But there are still no large, diversified technology platforms in the EU of the kind found in both the US and China, and many European start-ups are tied to the US market through acquisitions or invested capital.

The United Kingdom, which is no longer part of the EU post-Brexit, has the most developed AI ecosystem in Europe. That nation is particularly strong in fundamental research, which dates back to the early work on AI by scientists such as Alan Turing, after whom the major national AI institute is named. British research was behind the development of DeepMind, the advanced AI lab acquired by Google in 2014. DeepMind has been responsible for many algorithms that have caused controversy in recent years, including AlphaGo. Other European countries with strong AI capabilities are Germany and France. Germany is particularly strong in robotics and uses AI for smart applications in factories. France also has industrial applications but focuses strongly on AI in healthcare and defence.

Another country with an internationally competitive position in AI is Japan. There too, the technology is closely linked with the industrial sector and specifically with car manufacturers. AI has also developed strongly in Canada. As in the UK, this is based on a thriving ecosystem for fundamental research driven in part by the presence of prominent scientists Geoffrey Hinton, Yann LeCun and Yoshua Bengio, whose work has been funded by the Canadian Institute for Advanced Research (CIFAR).[19]

Russia's AI capacities are relatively limited, especially in terms of research investment.[20] At the same time, though, it does have a strong position in specific domains within AI. In 2015 the United Instrument Manufacturing Corporation announced a major research project in the field of AI and semantic data analysis. Russia's answer to Google, Yandex, has been using AI for search results for years. ABBYY focuses on text recognition. VisionLabs specializes in facial recognition for banks and the retail sector. N-Tech.Lab won first place in a global facial recognition competition in 2015 with its FaceN algorithm.[21] The software developed by this

[18] Trommel, 20 April 2021: 28.

[19] From an interview with Geoffrey Hinton in Ford, 2018: 92.

[20] See Mols, 2019.

[21] Bendett, 2019: 171–172.

firm linked images of Russian citizens mined from various data sources and platforms. A conference in 2018 set out a path for a Russian AI strategy that focuses on developing expertise, training and education programmes, identifying global developments and the use of AI in war games.[22]

One important conclusion we can draw from this brief overview is that many of the countries mentioned are committed to developing a national version of what we have called an 'AI identity' (see Chap. 6). Countries that invest in distinctive AI capacities and specialize in specific domains can use this strategy to strengthen their competitive position, allowing even relatively small nations to hold their own in the international AI arena. Several relatively small economies appear to be very successful in AI when you consider the number of relevant actors in this field. Israel, South Korea and Singapore are particularly notable in this respect.[23] As another relatively small country, the Netherlands excels in fundamental research in AI [24] and education.

Key Points: AI Capacities

- AI is a complex phenomenon, but we can estimate a country's capacities based on five dimensions: the quality of fundamental research, the availability of data, the required hardware, the business ecosystem and an enabling government.
- The US and China are the two world leaders. Both score well in all five domains, but interpretations of their relative positions vary. The EU also scores well, except in that it lacks an ecosystem for the commercial production of AI applications.
- The UK, Germany, France, Japan, Canada and Russia are medium-sized actors that excel in specific domains or applications and therefore have an 'AI identity'.
- Smaller countries can also be relevant actors on the world stage if they have specialized in a particular area of AI or fundamental research.

9.1.2 National AI Strategies

The field of AI is highly dynamic. This applies not only to private-sector players but also to governments. As we have seen in Chap. 3, many countries have presented AI strategies in recent years. But while these address various AI-related issues, such as

[22] Bendett, 2019: 174.

[23] Mols, 2019: 16.

[24] Rathenau, 2021a, b.

its use by government and ethical principles, they are often aimed primarily at strengthening a country's competitiveness.

We can distil a number of patterns from these documents. Canadian researcher Tim Dutton compared several of them and identified the following general themes: 'research', 'talent', 'industrial strategy', 'ethics', 'the future of work', 'data', 'AI use by government' and 'inclusion'.[25] Of these, research, industrial strategy and talent are the most commonly mentioned. Ethics also appears relatively often, but the paragraphs mentioning it are generally generic. Considering the consequences for ethics and broader civic values in relation to AI seems to lag some years behind the publication of the national strategies.

Investment in research and talent is addressed in many strategies. For example, the German one announced the establishment of twelve R&D centres and a hundred professorships. The American university MIT reported an investment of a billion dollars in an 'AI college'. Co-ordination and co-operation in research are also cited regularly. The French document provides for four interdisciplinary AI institutes, and in Canada CIFAR is working with several institutes to co-ordinate research. The Alan Turing Institute was established in the UK in 2015, with a growing number of research centres affiliating with it.

Another pattern revealed by comparing the various documents is that many place AI in a broader perspective of technological development. The Chinese strategy, for example, is linked to other plans for key technologies such as 'Made in China 2025'. The Japanese one positions AI within the 'Fourth Industrial Revolution'. The same applies to the South Korean version, which also speaks of an 'intelligent information society'. As mentioned in Chap. 3, the Dutch action plan for AI has also been incorporated into the government's broader digitalization strategy.

In line with the idea of an AI identity, it is salient that many countries link the development of AI in their strategy to sectors and domains in which they are already competitive. In Germany the federal government's *Artificial Intelligence Strategy* emphasizes the implementation of AI in heavy industry. This is in line with previous strategic initiatives such as 'Industry 4.0', which focused on robotics and smart manufacturing. As a major producer of machinery, infrastructure and transport technology, Germany wants to lead the 'smartification' of these sectors.

The Japanese strategy emphasizes three areas, one of which is mobility. With companies like Toyota, Nissan, Honda and Mitsubishi, Japan has much to gain from implementing AI in that market. The same applies to France, home to major car manufacturers such as Peugeot, Renault and Citroën. In a report that mathematician Cédric Villani wrote for the French government, he highlighted four areas for developing AI in France. Mobility is one, defence another (also a sector in which the French economy is strong). A third is health, in which a data hub is being established to combine information from healthcare providers, hospitals, health insurers,

[25] Dutton, 28 June 2018.

pharmaceutical companies, laboratories and other relevant parties.[26] Again this project – and others in the health domain – is building on national strengths. The French economy is characterized by a high degree of centralization ('dirigisme'). In healthcare the country has huge, centralized databases that can be further developed to serve as a basis for AI projects. That is not the case in many other countries.

Other nations that are strong in defence are also focusing on that sector. Israel does not yet officially have an AI strategy, but it does have strong capacities and has expressed an ambition to become a leader in AI in the fields of defence and cybersecurity. As mentioned earlier, part of Russia's strategy includes organizing AI war games. Moreover, governments in countries like Russia and China have a great deal of control over their people. Not surprisingly then, both are very strong in AI applications in the field of facial recognition. We return to this in Sect. 9.2.

A further pattern in many of the strategies is to focus on areas where there are major social issues that AI can provide an answer to. This seems to be why the domain of healthcare has been included in the Japanese strategy. Japan is the world's fastest ageing society and therefore faces an increasing demand for healthcare services. AI could help meet this need. The Indian strategy is entitled *AI for All* and its explicit goal is 'inclusion'. This is a major challenge for a nation faced with huge economic and social inequality. Several of the country's recent digital strategies, such as a programme for financial inclusion and services using biometric data, aim to achieve a more inclusive society. India's strategy for AI can thus be seen within the context of this ambition. Alongside the three domains already mentioned, a fourth in France's AI strategy is ecology – another area in which that nation's economy is strong, particularly regarding energy. Moreover, the Paris Agreement on climate change has made ecology a global challenge and an area in which the French are keen to build their global standing.

A final pattern in various strategies is the development of policies aimed at applying AI research in practical and commercial contexts. One of the five pillars of the UK's 'AI Sector Deal' is the establishment of an AI Council whose task is to improve co-operation between universities and industry. Another approach is Canada's 'Scale AI', part of the national 'superclusters' policy. This brings together companies in retail, manufacturing, transport, infrastructure and ICT to develop smart logistics chains using AI and robotics, and so improve the competitive position of Canadian business. Then there is Singapore's '100 Experiments'. In this innovative programme companies are invited to submit problems that could be solved using AI-based products, but where none is currently generally available. One condition is that it has to be possible to build such a product within nine to eighteen months. Applying firms are paired with Singaporean AI researchers, who receive special funding from government.

[26] Villani, 2018.

Key Points: National AI Strategies
- More than 60 national AI strategies have been published since 2017. A number of patterns can be discerned in these documents:
- A number of patterns can be discerned in these documents: they emphasize investment in research and talent; they place AI in a broader perspective of technological development; they highlight links with sectors in which a country is already competitive; they identify challenges AI could help overcome; and they encourage the commercialization and practical application of research results.

9.1.3 An International AI Race?

Many of the national strategies place a strong emphasis on strengthening the international position of the country concerned. Some appear to be in a race to become AI leaders and so increase their competitive advantage. A lot of the documents contain passages that reflect this in some way. The Chinese one refers to developing a "first-mover advantage in the development of AI", while the US version mentions "accelerating American leadership in AI".[27] Many authors also use the metaphor of the global race.[28] Kai-Fu Lee sees an analogy with the 'space race' during the Cold War. The victory in go over Lee Sedol can be seen as China's 'Sputnik moment', and the presentation of its national AI strategy a few months later as the equivalent of President Kennedy's speech calling on America to put a man on the Moon.[29] The Soviet Union's launch of Sputnik I led to the founding of NASA and DARPA, the innovation arm of the US military.[30] Like space then, AI is now the focus of numerous innovation plans.

Many of these developments can indeed be interpreted as a race. Countries are competing to attract talent. Germany's policy of establishing more professorships in AI could draw talented researchers away from Dutch institutions. As mentioned in Chap. 3, this 'brain drain' is a frequent topic of discussion – including in the Netherlands.[31]

The broader policy of competitiveness can also be understood as a race. Embracing AI too late can lead to a loss of earning power. There is also a risk that phenomena like network effects and dependencies will make it hard to catch up. In

[27] Smuha, 2019: 2.

[28] See, for example, Walch, 2020; Harari, 2019.

[29] Lee, 2018: 98.

[30] Weinberger, 2019.

[31] Hueck, 2 September 2018; De Rijke, 8 April 2019. A Dutch study by the Rathenau Institute has revealed that there is no net outflow of researchers from the Netherlands yet (Rathenau Instituut, 2021a, b).

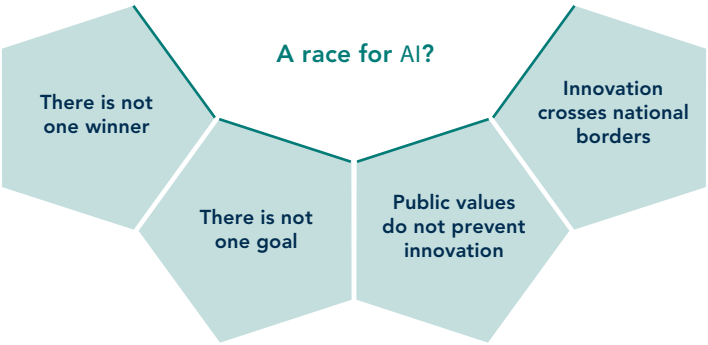


Fig. 9.2 Four shortcomings of the metaphor of an AI race

the specific case of AI, access to large amounts of useful data leads to better algorithms, creating a vicious circle that is hard for other parties to break out of. This is often referred to as the 'winner-takes-all' dynamic of AI.

Also consider the impact of AI on specific sectors. If American companies achieve success thanks to investment in AI for self-driving cars, that could be very detrimental for German economy's huge motor industry. The focus of AI strategies on sectors in which a country already excels involves both opportunities and risks: countries with large car manufacturers have expertise and data that can give them an advantage in the development of self-driving vehicles, but if they fail to grasp these opportunities their industry may fall behind and lose market share. The same applies to the Dutch agricultural sector. China is investing in innovative agricultural technology, and American companies like John Deere have access to data on Dutch agriculture and horticulture through the machines they sell. So, there is an opportunity here for the Netherlands in the field of agricultural and horticultural AI, but at the same time also a threat that the country's traditionally strong position in this sector could be weakened.

Finally, the metaphor of a global race could also be applied to the military domain: AI can give a specific country a strategic advantage over others. We discuss this in more detail in Sect. 9.2. But this metaphor also has serious shortcomings. Some of its implicit assumptions about the global development of AI are incorrect. Figure 9.2 reveals these issues, which we discuss further below.

The first shortcoming of the race metaphor is that it suggests that there can only be one winner. The development of AI is presented as a 'zero-sum' situation. This may be the case with some specific goals in other domains, such as being the first to reach the Moon, but it does not apply to the development of a system technology. Kai-Fu Lee – who makes the analogy with space, as we noted earlier – likewise suggests that a race is not the right metaphor. The development of AI, he says, is ultimately more like the Industrial Revolution or the rise of electricity than the space race. As with electricity, there is no such thing as 'zero-sum' in AI. One country's gain is not necessarily another's loss. Technologies spread all over the world and

lead to progress and more prosperity in all sorts of places.[32] We could even say this of the space race. Even though only one country could be the first to land on the Moon, the innovations that made space travel possible, such as satellites and GPS technology, benefited people worldwide.

But some countries certainly gained more than others – for example, because their innovative companies exported industrial products, energy resources or space technology to other countries. However, it is important to emphasize that the benefits of those technologies were also shared by citizens elsewhere. This perspective helps shift the focus from the development of the technology to its diffusion.

In AI the focus is often on competition at the most advanced level, in which only the leading companies based in the richest countries can participate. But the less advanced forms of AI are much more widespread, and they are having a far greater impact on the world. To stay with the analogy of electricity, only a few countries are able to develop nuclear power because this form of energy generation is highly advanced, risky and requires huge investments. However, there are many other ways to generate electricity that benefit citizens and create markets for businesses all over the world. In other words, while electricity innovations are often concentrated in only a few places, electricity production is much more widely distributed and its use even wider still.[33]

The focus on diffusion is also important because this aspect raises different questions than when focusing on the 'technological frontier' of AI, the place where innovation takes place. Developing world-leading laboratories not the same as ensuring that AI is widely embedded in society. The US economy is at the forefront of technology in many areas, but the general population benefits less from this than people in other countries.[34] Conversely, countries that were not the developers of a new technology can still be very successful at implementing and disseminating it.[35]

Another implication of the focus on diffusion is that the technical expertise of the inventors and major laboratories becomes less important than that of the people responsible for maintaining the technological infrastructure. For example, a very large proportion of the people who work on electricity or IT systems do so as repair or maintenance staff. This important form of technical expertise will also need to be given due attention in the development of AI.

If we reason in terms of diffusion rather than the 'technological frontier', it becomes apparent that there is a real concern that certain groups in society will not be able to keep up with this development. The idea of a global race focuses on a struggle between countries and compares their dominant companies, but neglects to examine the effects on their wider populations. It can thus overshadow the issue of AI inequality in a country.

---

[32] Lee, 2018: 227.

[33] Edgerton, 2008: 80.

[34] Hall & Soskice, 2001.

[35] The Netherlands is a good example of this (WRR, 2013).

A second problem with the metaphor of a global race is that it suggests that everyone is aiming for the same goal. In the first place it is unclear what that goal should be. As has become apparent in this report, AI is a complex phenomenon with many potential areas of application. Its goal, therefore, is very difficult to define. For example, the goal of 'leadership in AI' is much harder to quantify than the first Moon landing. Success in AI can occur in several domains, which makes the idea of a single finishing line problematic.

Countries can also follow very different paths to achieve their AI goals. In Chap. 4 we saw that the development of electricity in continental Europe involved different applications and different organizational models than in the US. While countries such as Canada and the United Kingdom are focusing their AI strategies on fundamental research, other nations have chosen specific sectors or areas of application. The idea of an 'AI identity' in which countries are distinguished by their AI capacities is incompatible with the metaphor of an AI race. That suggests that everyone is on the same path, and in so doing blinds us to the various ways in which AI can be successful and make countries more competitive.

Perhaps even more important than the previous objections, is the fact that the idea of a 'global race' suggests a conflict between competitiveness on the one hand and the protection of civic values on the other. Based on the idea that some countries are becoming too dominant, it is often argued that those which are lagging behind should make haste and not be held back too much by discussions about the protection of rights because this will only increase the distance between them and the dominant players. So restricting access to data for privacy reasons or reining in experiments with surveillance to protect the freedom of citizens would be detrimental to a country's competitive position. Nick Bostrom emphasizes that the dynamics of an AI race could come at the expense of caution and safety.[36]

Although such tensions certainly exist in practice, it is dangerous and unjustified to try to place competitiveness and the protection of fundamental rights in opposing camps. Firstly, it is not clear whether economic competitiveness that violates fundamental rights can be sustainable in the long term. The benefits of implementing extreme surveillance to reduce crime do not weigh up against the costs for individual freedom. Such innovations will encounter particularly fierce resistance in countries with strong democratic traditions, however effective they may be. It should also be emphasized that innovations can have more economic success if civic values are safeguarded. We have seen this with previous system technologies. At first there was resistance to safety measures in cars because critics feared that they would push up prices and stifle innovation. But in fact, such measures eventually led to people having more confidence in cars and using them more.

[36] "This is one of the concerns with a racing dynamic, where you have a lot of different competitors racing to get to some kind of finish line first – in a tight race you are forced to throw caution to the wind. The race would go to whoever squanders the least effort on safety, and that would be a very undesirable situation" (Ford, 2018: 113).

This is exactly the argument that the European Union uses in its 'ethical' approach to AI. The European Commission states, "Building on its reputation for safe and high-quality products, Europe's ethical approach to AI strengthens citizens' trust in the digital development and aims at building a competitive advantage for European AI companies." Pekka Ala-Pietilä, chair of the Commission's High-Level Expert Group on AI, has said, "Ethics and competitiveness go hand in hand. Businesses cannot be run sustainably without trust, and there can be no trust without ethics. And when there is no trust, there is no buy-in of the technology or enjoyment of the benefits that it can bring."[37] Later in this chapter we look into the role of the European Union with regard to legislation and ethics. A few points of criticism aside, we support the European Commission's conclusion that the goals of competitiveness and civic values do not have to be at odds with one another.

The EU's approach is sometimes criticized for emphasizing ethics over competitiveness in AI. The EU is working to improve the latter and there is a public discussion as to whether this is happening to a sufficient extent. Another justified criticism of this policy is that the EU is a little too eager to position itself as the world's ethical leader, with the market-driven US and state-driven China as two opposite extremes. In so doing it disregards the activities of other countries in this area. Nations like Japan, Canada, Dubai, Singapore and Australia have also developed their own ethical guidelines for AI; and even China, which is often portrayed as having little regard for ethical standards, published the *Beijing Principles for Ethical AI* in 2019.[38]

A similar idea to the race with a single winner is the notion that dominant countries can contain the development of a technology within their own borders. As we have seen with previous system technologies, this has never actually been the case; the development of revolutionary technologies has always been a global affair driven by researchers and companies in various countries. Efforts by governments to nationalize innovation have never succeeded and often even backfired: their companies lost their leading market positions because export controls and nationalistic government policies only encouraged the competition abroad.

This dynamic now seems to be affecting the Sino-American trade dispute, which also involves AI. The US government is trying to repress China's rise in this area. One of its main weapons is to ban chip manufacturers like Intel from selling their advanced products to China. It is also pressuring European companies to restrict their exports to China. It is quite possible, however, that this will only boost China's ambitions to develop its own chip sector. The US policy could also serve to strengthen ties between China and other countries in Europe or Asia. This trade conflict is now forcing European countries to take a stand in this global arena.

[37] Smuha, 2019: 15.

[38] Smuha, 2019: 19.

Key Points: International AI Race

- A number of international developments in AI, such as access to scientific talent and strengthening national security, could be considered part of a race dynamic. However, the metaphor of a race also has serious shortcomings.
- The metaphor incorrectly suggests a zero-sum situation and ignores the importance of diffusion.
- Moreover, not all countries have the same goal and so there cannot be just one 'winner'.
- The race metaphor also wrongly suggests friction between competitiveness on the one hand and civic values on the other.
- Finally, it is impossible to contain AI innovations within a country's borders.

9.1.4 From Competition to Co-operation

The clear historical lesson is that no country will be able to contain and develop AI completely within its own borders – not even the US or China, despite their advanced ecosystems. This applies even more to smaller countries, which can only benefit from open international co-operation.

In fact, focusing on international co-operation – between the European member states, for instance – could actually strengthen a country's competitive position. Nations should invest in co-operation to strengthen their international profile. This will require an integrated policy of 'AI diplomacy'. We have distinguished five areas on which this policy could focus: fundamental research, commercial applications, regulation, ethical guidelines and standards.

CLAIRE hopes to achieve these goals by creating a network of regional centres of excellence across Europe, with or without specializations, and a central hub. The project is intended to do for AI what CERN has done for particle physics: establish a central European institute with state-of-the-art infrastructure for AI research. In 2020 the head office was established in The Hague.

ELLIS, the European Laboratory for Learning and Intelligent Systems, is a pan-European AI 'network of excellence' that focuses on fundamental science, technical innovation and societal impact. It builds upon machine learning as the driver for modern AI and is aiming to secure Europe's sovereignty in this field by creating a multi-centric 'European AI Lighthouse'. There are currently 34 ELLIS units in 14 countries, either located at existing AI research institutes or created from scratch. Through them ELLIS aims to create new working environments for outstanding researchers and to enable combinations of cutting-edge research with the creation of start-ups and industrial impact.

Fundamental research partnerships are the clearest form of co-operation in AI. Encouraging and attracting international groups to their shores is a way for

Box 9.1: CLAIRE and ELLIS
CLAIRE and ELLIS are prominent AI research partnerships. CLAIRE, the Confederation of Laboratories for AI Research in Europe, is an alliance of AI scientists founded by Dutch professor Holger Hoos, Philipp Slusallek from Germany and Morten Irgens from Norway. Their vision document has been signed by more than 550 AI experts. The goals of the alliance are to strengthen European excellence in all areas of AI, with a human-centred focus. According to the founders, a strong European research organization is needed for the development of AI; if Europe is allowed to fall behind, this could lead to negative economic consequences, an academic brain drain, less transparency and increasing dependence on foreign technologies.

countries to directly strengthen their own AI development programmes. There are a number of such projects in Europe, of which ELLIS and CLAIRE are the two most prominent. Both are involved in broad AI research, and specifically in machine learning. CLAIRE has also been described as a 'CERN for AI' (see Box 9.1).

A second area in which co-operation can improve competitiveness is commercial projects. This could involve establishing new services and organizations as part of international partnerships, as well as strengthening and co-ordinating the AI activities of existing companies. In pursuit of greater strategic digital autonomy, much has been achieved in the EU in this area in recent times.[39] The Gaia-X project for cloud and data infrastructures is part of this ambition (see Box 9.2). There are also similar European projects to advance cybersecurity.[40]

Commercial projects could also involve encouraging more co-operation with existing companies active in the field of AI or a related discipline, such as the Scandinavian suppliers of telecom infrastructure Nokia and Ericsson. The Dutch chipmakers ASML and NXP are also examples. To achieve greater strategic digital autonomy, countries can contribute to protecting and strengthening European industries in specific domains and encourage further European co-operation if desired. Although there are legitimate reservations about such industry-focused policies, Europe has shown in the past that they can be successful. By working together European countries were able to create the aviation giant Airbus and the Galileo satellite navigation system, despite their initially weaker position.

A third way a country can improve its competitive position in AI is through co-operation in the domain of international legislation. This is an area in which the European Union can add real value. Regulatory enforcement is often seen as a form

[39] Sheikh & Timmers, 2020.

[40] This is translated into the Dutch context in Timmers & Dezeure, 2021 (commissioned by the national Cyber Security Council).

Box 9.2: Gaia-X
Gaia-X started life as a project in Germany, which has since been joined by France and other European countries. The project charter was presented in October 2019, and on 15 September 2020 a group of 22 organizations signed an 'incorporation paper'. They included German firms such as Bosch and Siemens and French ones like Orange and Atos. Full details of the project have yet to fully crystallize, but its basic goals are to strengthen European data sovereignty, reduce dependence on foreign players (through lock-in clauses in contractual arrangements), make cloud services more attractive and create an ecosystem for innovation. To achieve this the project aims to build a cloud infrastructure. This is intended not so much to provide an alternative to American cloud services as to make it easier for European parties to compete with these services, so that European data can stay under European control. The infrastructure will be governed by common rules, standards and technology. The project has identified various application domains for AI, such as 'sustainable finance' and 'ambient assisted living'.

of soft power that Europe is able to wield effectively.[41] One obvious recent example is the *General Data Protection Regulation* (GDPR), which has set a worldwide standard. But it has also come in for some criticism. According to some it puts the EU at a disadvantage compared with China and the US because it hampers innovations that use personal data.[42] With its introduction, however, the EU has not only established a global benchmark for data protection but also opened up debate on this issue.[43] Several other countries and a number of US states have adopted its underlying principles, and the GDPR has even influenced Chinese personal data protection policy.

At the 2019 meeting of the G20 in Japan, Angela Merkel stated that the challenge for the next European Commission was to draw up legislation for 'trustworthy AI' similar to the GDPR.[44] Researcher Nathalie Smuha speaks of 'regulatory competition', an international process of co-operation and competition in the regulation of AI, including legislation.[45]

[41] In this context Jan Zielonka goes so far as to describe the EU as a 'regulatory empire' (Zielonka, 2008: 474). Anu Bradford calls the global influence of the EU 'the Brussels effect'. She believes that the EU is in a position to implement regulations that can be embedded in the legal frameworks of global markets without needing to involve international institutions or seek international co-operation. The resultant effect is the 'Europeanization' of many aspects of global trade (Bradford, 2020).

[42] Smuha, 2019.

[43] Lee Bygrave explains exactly how this worked for the GDPR and refers to Bradford's work. Bygrave also emphatically emphasizes European expertise and the role played by the Council of Europe in the whole situation (Bygrave, 2021).

[44] Smuha, 2019: 17.

[45] Smuha, 2019: 26.

Such processes not only involve legislation, they also affect the fourth domain of international co-operation: guidelines[46] and ethical principles. A great deal is happening in this area but a few prominent projects stand out. In May 2019 the OECD adopted a set of common ethical principles for AI. This was the first such set of intergovernmental guidelines in this area. A month later they were formally adopted by the G20. There is also a UNESCO project to develop a global code of ethics for AI. Finally, the Council of Europe has established an ad hoc committee for AI, CAHAI, with the aim of drawing up binding rules to govern the protection of human rights, democracy and the rule of law in co-operation with all member states.[47]

Because international co-operation typically receives less attention than the other domains, we discuss it in more detail here. Our analysis of earlier system technologies has revealed how important it is to establish expert forums for the development of technical and other standards. Behind the scenes, much is going on in the forums where standards for AI are being developed. In the previous chapter we considered standards from the perspective of their role in regulating a new technology. For example, they are a way to enhance the interoperability of a technology. This does not concern its functions, but rather the dimension of international co-operation and its influence on a country's competitive position. The ability to set standards has a major impact on a country's competitiveness because the national industries that implement those standards thereby gain a 'first-mover advantage'. A country that develops its own standards rather than adopting those compiled elsewhere has more control over them and is also able to influence other countries that have to follow those standards and so potentially benefit from lock-in effects.

As we saw in Part I of this report, when it comes to the interpretation of standards the historical development of system technologies has always been characterized by a certain friction between technocratically oriented experts on the one hand and national politicians and officials on the other. So, what is the current international situation regarding standards and their implementation?

It is noteworthy that Europe is playing an internationally leading role here, and even has what is also called 'standard power'.[48] This is connected to a long track record of development in this area and of processes established to integrate standards across national borders – initially within Europe, but now also internationally. But the specific European model of standardization, characterized by public-private partnerships, also plays a role here. Standards are developed by private organizations licensed by governments. Each country has established separate bodies for specific purposes. For general standards these are NEN (the Netherlands

[46] These are not the 'directives' that serve as instruments of EU legislative policy, but rather a set of less formal rules.

[47] Smuha, 2019: 21.

[48] Kaiser & Schot, 2014.

Standardization Institute) in the Netherlands and DIN (the German Institute for Standardization) in Germany. There are also entities that focus on specific sectors, such as the German DKE for electrical engineering.

Furthermore, Europe has a clear hierarchical structure in place. The national bodies fall under the umbrella of European organizations: NEN is a member of CEN, the European Committee for Standardization, and DKE a member of CENELEC, the European Committee for Electrotechnical Standardization. In turn these are part of global organizations – respectively the ISO (International Organization for Standardization) and the IEC (International Electrotechnical Commission). International agreements have established that the higher its level in the hierarchy, the more priority a body has in setting standards.[49] Those lower down the chain describe the requirements imposed by the relevant standards in more specific detail.

In this respect the European model differs substantially from its American counterpart, which has a much greater commercial orientation. That means that a multitude of standardization bodies are often found competing in a particular domain. This lack of co-ordination makes the US system less influential globally than Europe's. Standards for AI are mainly developed in three forums: a joint initiative of the ISO and IEC, the IEEE Standards Association for engineers and the *International Telecommunication Union* (ITU), a UN specialist agency.[50]

A number of things stand out regarding AI standardization in the international domain. The first is that China is playing an increasingly important role. As we have seen, exerting an influence over global standards is part of that country's broader strategy. It has contributed high-ranking officials to ISO, the IEC and the ITU, and the proportion of Chinese appointees to the committees and working groups of these organizations is growing. In absolute terms it still has fewer representatives than many other countries, but that is changing.[51] Its so-called 'Belt and Road Initiative', a major international project, also has an explicit standardization component.[52] Finally, the Chinese system is due to be reformed as a consequence of the exploratory study *China Standards 2035*.[53] With regard to AI specifically, China is using its growing sway to shape approaches to facial recognition and surveillance in particular. Firms like ZTE, Dahua and China Telecom have submitted proposals for international standards to the ITU. In this way their home country is gaining

[49] Rühlig, 2020: 11.

[50] Cihon, 2019: 10.

[51] Rühlig, 2020: 22.

[52] In 2015 the Chinese government published its 'Action Plan for Harmonization of Standards along the Belt and Road'. The first step involved the translation of 500 Chinese national and industry standards into other languages (Rühlig, 2020: 24).

[53] Rühlig, 2020: 18.

influence in emerging economies in Africa, Asia and Latin America where these standards are often adopted, and thus strengthening its access to important markets for the technologies they govern. The Chinese proposals for surveillance standards, for instance, coincide exactly with those applied in the design of ZTE's Smart Street 2.0 traffic light.[54]

China is not the only country responsible for the 'geopolitization' of standards. The US is playing its part, too. Indeed, telecommunication standardization has become one theatre in the wider trade dispute between the two nations,[55] part of what has become known as the 'connectivity wars'.[56] For example, they have been fighting for leadership of the subcommittee for AI standardization of the ISO/IEC Joint Technical Committee.[57]

As a result of these developments, the European role in and approach to setting standards – a relatively technocratic process driven by commercial parties with specific expertise – is coming under pressure.[58] These developments call for a European response to the new politics of standardization.

Key Points: From Competition to Co-operation

- Since it is clear that no nation will be able to contain and develop AI completely within its own borders, individual countries can only benefit from international co-operation, as within the EU.
- Countries can strengthen their competitiveness in five specific domains by engaging in international co-operation and an integrated policy of 'AI diplomacy'.
- One of those domains is fundamental research. Projects such as ELLIS and CLAIRE are working to improve Europe's position in this regard.
- Countries that co-operate in the area of regulation can benefit from market influence and establish a regulatory 'first-mover advantage'. The individual countries in the partnership are then able to profit indirectly from this.
- Countries can also play a more active role in international organizations in the development of ethical guidelines and principles.
- Finally, Europe needs to respond to the current 'geopolitization' of standardization processes.

[54] Gross et al., 1 December 2019

[55] See Hillman (2019) for an analysis of the significance of standards in infrastructure projects.

[56] Leonard, 2016.

[57] Smuha, 2019: 21.

[58] Cihon, 2019.

9.2 AI and National Security

Positioning a country in the field of AI is not just a matter of improving its competitiveness. The international dimension also involves issues of national security and sovereignty. As we have already seen, competitiveness and security need not be mutually exclusive. In this section we focus primarily on issues of national security (so not on the security of AI applications for individual citizens, for example). This theme has been developed in more detail in the WRR report *Security in an Interconnected World*.[59]

The influence of AI on the international balance of power, and hence national security, is widely recognized. As a famous quote by Russian President Vladimir Putin in a speech to students and scholars in 2017 has it, "The country that leads in AI will become the ruler of the world." This is also the American view. The US military has historically had strong ties with big tech through project funding by organizations like DARPA and the Department of Defense.[60] In 2015 the Defense Innovation Unit (DIU) was established to harness Silicon Valley technologies for use by the military. The Secretary of Defense at the time, Jim Mattis, wrote a memorandum advocating an integrated national strategy for AI in 2018. Later that year President Trump signed the National Defense Authorization Act (NDAA), which also established the National Security Commission for AI. The Pentagon went on to launch the Joint Artificial Intelligence Center (JAIC).[61] In March 2021 the National Security Commission on Artificial Intelligence (NSCAI), chaired by former Google CEO Eric Schmidt, published a hefty final report.[62]

We have questioned the idea of a global AI race in terms of economic competitiveness. This metaphor does, however, seem more appropriate when it comes to national security. Unlike economic gains, military capabilities do offer zero-sum advantages to individual countries. This has been noted around the world: while there were fewer than 300 online hits for 'AI arms race' before 2016, in 2018, that number grew to 50,000. Newspapers like *The Guardian* and the *Wall Street Journal* now write extensively about the phenomenon.[63]

As far as the impact of AI on national security is concerned, the first application that often comes to mind is autonomous weapons. This is currently a much-discussed issue. We therefore begin by examining this particular phenomenon

[59] WRR, 2017.

[60] DARPA's ARPANET project was responsible for the precursor to the internet and has contributed towards the development of all manner of advanced weapons. See Weinberger (2019) for a historical overview of the agency's development and its success and failures.

[61] Leung, 2019: 266.

[62] National Security Commission of Artificial Intelligence, 2021.

[63] Leung, 2019: 263.

and then broaden our scope to consider other, lesser-known ways in which AI affects security.

9.2.1 Autonomous Weapons

Autonomous weapons appeal to the imagination and have been the subject of many dystopian novels. The theme of 'the rise of the machines' often takes the form of robots deciding to attack humankind. In Chap. 5 on 'demystification' we have demonstrated why the fear that robots will become conscious entities and decide that humans are their enemies is really quite unrealistic. However, we have also seen that robots do not have to become conscious to pose a threat to us. What is currently happening in the field of autonomous weapons, and what implications does this have for the international order and thus for the national security of individual countries?

It is actually not easy to define an autonomous weapon. Before we can focus on this issue, it is a good idea to address some of the existing technologies in this area, because there is a huge diversity of them. Israel's Harpy drone flies autonomously in search of enemy radar systems and is programmed to attack them without having to ask permission (Box 9.3). China has built its own version of this device using reverse engineering. South Korea has deployed a robotic weapon in the demilitarized zone on the border with North Korea that can shoot at moving objects autonomously. Russia is building armed ground robots for conflicts on the European plains. At least 30 countries have defence systems that, when activated, can autonomously intercept incoming threats such as missiles. The US has the Aegis system for ships and the land-based Patriot. Other examples are the German Mantis, the Israeli Trophy and the Russian Arena systems.[64] In 2017 sixteen countries possessed weaponized drones. Ninety percent of international sales involved Chinese technology.

So a lot is happening in the development of autonomous weapons. The news reports have created a real stir, and there have been several international campaigns to have such systems banned (see Chap. 7). But there are various obstacles to

Box 9.3: Drones and Warfare
Drones were first used in warfare in Vietnam but really took off after 11 September 2001. The US Army used them extensively in Afghanistan. In 2018 Syrian rebels carried out a major attack on a Russian airbase with thirteen drones. Later that year Russia brought the heavily armed Uran-9 system to that same conflict. In August 2018 an attempt was made to assassinate President Maduro of Venezuela with a drone.[65]

[64] Scharre, 2018: 45–46.

[65] Scharre, 2018: 364.

achieving this.[66] Several of these are illustrative of broader problems with the global regulation of AI and thus the task of positioning, and so we discuss them in more detail here. They concern the issues of definitions, the dual-use nature of AI and motivating countries to participate in a ban.

We have seen in Chap. 2 saw how difficult it is to define AI. Although autonomous weapons form a fairly specific application of the technology, they are also very hard to define, and this creates difficulties when it comes to establishing regulations governing them. Paul Scharre explains that there is considerable ambiguity about what constitutes an autonomous weapon because there are three different dimensions of autonomy.[67] The first concerns the nature of the tasks involved. Some are undertaken by humans and others by a machine. A fully autonomous vehicle will be able to do everything itself, but many current cars can already perform many of these tasks itself, such as cruise control, parking or braking. The latter ability enables the car to take over control from the human driver to prevent an accident. But when is a weapon autonomous? What if it can navigate of its own accord and avoid other objects, but not actually fire a weapon?

Specifically regarding the task of firing weapons, we can question whether a system is still autonomous if a human first has to press a button to activate its 'attack mode' and only then is the robot able to deploy deadly force itself. There are also technologies whereby a human marks an object for destruction and the robot subsequently proceeds to attack it until it is destroyed. Is this an autonomous weapon? One way of delimiting autonomy is to allow it only for tasks of a defensive nature. There are systems that can automatically shoot incoming missiles out of the air, a feature that many people will see the benefit of. But is it still a defensive response if that system fires back at the source of those missiles? If this system is moved into the opponent's territory, it becomes even more difficult to distinguish offensive from defensive use of an autonomous weapon.

The second dimension of autonomy is the role of the human being. For this purpose we use the same framework for the military domain as already discussed in Chap. 6. Semi-autonomous systems are the equivalent of 'human in the loop' where a machine – having completed a task – waits for human input to continue. In supervised autonomous systems a human remains 'on the loop' and can intervene or stop the process. Finally, in fully autonomous systems humans are 'out of the loop' and have no role in the decisions the system makes.

The third dimension of autonomy is the level of intelligence. A traditional landmine basically acts autonomously by exploding when someone steps on it. But few people would call it an autonomous weapon. This is because the level of intelligence required is too low. The mine does not need to conduct any complex calculations before it explodes. It can therefore better be described as 'automatic'. At a higher level a number of variables need to be considered before action is taken. This is what we call 'automated'. Only when the activity becomes much more complex, and a system is able to determine independently *how* to achieve a goal can we speak of 'autonomy'.

[66] For more details, see Buruma, 2020: 69–112.

[67] Scharre, 2018: 27–32.

The fact that autonomous weapons can be defined in three dimensions makes it all the more difficult to reach international agreement on a definition and to determine how an individual country should position itself in this area. Unlike with other types of weapons, the definition here involves the question of whether autonomy should be granted solely to defensive systems, for example, or also to systems where humans provide general instructions only, to ones where a human only operates the controls or to ones with a low level of inherent intelligence.

In addition to definitions, there is another problem with the international coordination of autonomous weapons that also illustrates a broader AI issue: the dual-use nature of these systems. This refers to applications that can be used for both peaceful civilian purposes and as weapons in a conflict.

One example is DARPA's Fast Lightweight Autonomy (FLA). Videos of this technology went viral worldwide. Accompanied by the James Bond theme, a swarm of drones can be seen flying through windows into houses and carrying out all kinds of complex manoeuvres in the air. The films caused quite a stir and the movement against autonomous weapons subsequently targeted FLA. Thus far the system has not been equipped with weapons, but it is clear that highly advanced aerial drones that can move around objects and fly information will be of great interest to the military. As long as there are no weapons involved, the technology underlying FLA is similar to that in a self-driving car, involving localization, mapping, object detection and dynamic navigation at high speed.[68] But these same technological features do not make self-driving cars a security threat.

Another example is a drone's ability to target and track an object. The military could put this functionality to good use – to follow a moving truck, say. But this same technology also has civilian applications; several commercial drones already have this capability and are used, for example, to track and film wedding processions.

Another capability that can be applied to autonomous weapons is the ability to detect human faces or identify specific people. If this is combined with firepower, it becomes very dangerous indeed. But the software that enables it is already being used in many other applications. In fact, such software can be downloaded free of charge from large open-source software databases such as TensorFlow, complete with support for training the necessary algorithms.

So the technology underlying autonomous weapons, from navigation and object tracking to facial recognition, is already used in a variety of peaceful civilian applications, which makes some kind of general ban difficult to implement. A drone that recognizes someone's face and then delivers them their package uses almost exactly the same technology as the drone designed to shoot that same person. Moreover, if facial recognition is used to avoid killing people rather than to kill them, is it still reprehensible? Even if this reduces the likelihood of civilian casualties? The wide range of peaceful applications of AI thus makes it far more difficult to restrict in military contexts – especially by comparison with, say, chemical weapons or long-range missiles (Box 9.4).[69]

---

[68] Scharre, 2018: 70.

[69] So to address the threats, more attention should be paid to practical experiences with other dual-use technologies (Brundage et al., 2018).

Box 9.4: Views on the Impact of Autonomous Weapons
Various AI researchers anticipate that the use of autonomous weapons will actually make war less destructive. Nick Bostrom stresses the importance of removing people from the battlefield as much as possible and the prospect of fewer fatalities occurring thanks to the precision of autonomous weapons.[70] Rodney Brooks points out that, unlike a human being, a robot can afford not to shoot back until after it has been shot at itself.[71] According to Yann LeCun, the greater precision and potentially less lethal nature of these weapons is turning the military into something more like a police force.[72] This sounds like a positive development, but according to Frank Pasquale it also brings new dangers. If combat increasingly takes on the character of an international policing mission, with fewer casualties, politicians will also feel less pressure to spare the lives of soldiers and so wars may continue to smoulder and be less easy to end.[73]

The third obstacle in the way of unambiguous international agreements on autonomous weapons concerns the motivation of powerful countries. There is now some international consensus concerning the idea of 'meaningful human control'.[74] The Dutch government has mentioned this in a comprehensive position paper[75] and in late March 2018 reaffirmed that "meaningful human control is always necessary for the deployment of autonomous weapon systems".[76] Two ministers noted that "the Netherlands is one of the few countries to have formulated a comprehensive government standpoint on this subject, and has submitted a summary hereof as a non-paper" to contribute towards the debate on autonomous weapons systems as part of the UN Convention on Certain Conventional Weapons (CCW). Eleven principles for policy on lethal autonomous weapons systems (LAWS) were finally adopted under the CCW in 2019, with 'human responsibility' second on the list.[77]

However, it will still be difficult to prevent countries with technologically advanced armies from developing weapons with more and more autonomy. As

[70] From an interview with Nick Bostrom (Ford, 2018: 107).
[71] From an interview with Rodney Brooks (Ford, 2018: 440).
[72] From an interview with Yann LeCun (Ford, 2018: 137).
[73] Pasquale, 2020: 152.
[74] AIV & *Commissie van advies inzake volkenrechtelijke vraagstukken* 2015.
[75] *Kamerstukken II* 2015/16, 34300-X, no. 88.
[76] *Aanhangsel Handelingen II* 2017/18, no. 1645.
[77] GGE, 13 December 2019.

already mentioned, there are various strong international lobbies against such weapons. It is striking that these are largely driven by NGOs and less so by states. In fact, not a single powerful nation has yet supported a complete ban on these weapons. Those which have spoken out in favour include Ecuador, Ghana, Iraq and Pakistan, as well as states that have no armies like Costa Rica and the Vatican. In other words, there are no military superpowers among them, nor any of the front runners in the protection of human rights. Instead, these are mainly countries that fear what stronger nations might be able to do to them and hope to mitigate that threat with a ban.[78] Although some countries in Europe, such as Austria, are also moving in this direction, without the support of nations with a strong military it will be very difficult to actually implement such a ban.

Scharre also notes that an international treaty is not the most effective tool to counter the use of certain weapons. There are instances of weapons being used despite a treaty ban, and there are also examples of weapons that have not been used despite the lack of a ban. What is important here is whether countries expect reciprocity. The fear that another nation might also use the weapon acts as a deterrent. Countries that do not have this fear are less inhibited in the use of new technologies.[79] Another deterrent is transparency about the use of a particular weapon, which allows a country to be held accountable. This transparency is difficult to achieve with autonomous weapons. Their autonomous nature is determined not by their hardware or certain distinct physical characteristics, but by their software. This makes it a lot harder to provide transparency about what they do in a conflict situation.

The three issues of defining autonomous weapons, their interdependence with civilian technologies and the motivation of powerful states illustrate how complex it is to reach an agreement on the international co-ordination of this technology. However, this does not make that impossible. This case study is illustrative of some of the obstacles to successful international co-ordination of other AI-related issues. By co-operating, individual countries can reinforce their position in this power play and exert more influence over the international agreements for the use of autonomous weapons.

The discussion on the impact of AI on security is, as mentioned, often about autonomous weapons. Because these weapons appeal to the imagination, other applications are unfairly given less attention. However, AI can also influence warfare in other ways, which will be discussed in the following paragraphs.

[78] Scharre, 2018: 350.

[79] Scharre, 2018: 340.

Key Points: Autonomous Weapons

- When it comes to AI and national security, much attention is being paid to autonomous weapons since they capture the imagination and also meet with a lot of resistance. Several obstacles make it difficult to reach clear international agreements on this issue.
- Autonomous weapons can be defined in three dimensions, which prevents the establishment of a generally accepted definition and therefore makes them difficult to regulate.
- The dual-use nature of many of the technical applications makes it very difficult to restrict or ban the technology.
- Countries with a strong military are resisting such impediments.
- Individual countries must take these obstacles into account in the international power play that affects their security at home and abroad.

9.2.2 Other Military Applications

As mentioned, the discussion on the impact of AI on national security often focuses on autonomous weapons. Because these attract so much attention, other applications unjustly receive too little. In a study for NATO on the influence of AI on warfare, in addition to autonomous robotic systems Matej Tonin also distinguishes 'information and decision support'.[80] AI can increase the speed of analysis and decision-making in war by shortening the response time of defensive systems, providing more relevant information to decision-makers (giving them a potential advantage over rivals), enabling early detection of cyberattacks and helping identify attempts to spread disinformation. Not only can AI increase the speed of decision-making, moreover, it can also improve its quality. Tonin quotes a British officer who observed that, in a particular situation, his forces were "swimming in sensors, drowning in data and starving for insight". AI can help here by, for example, analysing surveillance data, highlighting abnormal patterns or picking up weak signals of potential threats.

It is relevant to note at this point that several very important revelations about military organizations in recent years have come from very basic data sources. Information gleaned from the running and cycling app Strava revealed the location of a secret US military base in Africa and a top-secret Chinese ship was revealed in the background of a picture taken by a tourist.[81] In 2021 researchers discovered silos holding nuclear weapons in China by interpreting satellite data. AI could potentially

[80] Tonin, 2019.

[81] Singer & Brooking, 2018: 58.

make a significant contribution to the analysis of such basic information sources for relevant military data.[82]

The use of AI for information provision also brings to light broader issues concerning its use in the defence domain. The first is 'novelty detection'.[83] This concerns the ability of AI systems to ascertain that certain input falls outside the range of what they have been trained to do. An algorithm trained to find dogs in a picture must be able to recognize a dog that it has never seen before. But when presented with an image of a dolphin it must be able to recognize that that is so different from any dog that it is something entirely new, not categorize it as the type of dog the dolphin most closely resembles. So, the algorithm has to distinguish between 'different in detail, but very similar' and 'highly dissimilar'. This is particularly important in the military domain, where different types of incoming missiles need to be detected accurately, but aircraft must never be identified as missiles.

A second issue concerns the manipulation of data to mislead an opponent. For example, data manipulators can look for 'edge cases', where a weakness in an algorithm leads to a totally different outcome. Laboratory researchers made subtle manipulations to images of buses so that they were identified as ostriches and did the same with turtles to make them become guns. In another study only a few pixels had to be manipulated for a neural network to identify a picture of an elephant as a car.[84] We saw in Chap. 6 how this may cause problems for self-driving cars, and such manipulation is even more dangerous in the military domain. Systems could be misled so that they fail to recognize attacks. Conversely, a non-threatening activity could be presented as a threat in order to provoke an attack. Extremely complex deceptions can be conceived where combatants could subtly manipulate data to make a hospital appear as a military facility so that it is targeted by the enemy. If the deception is small and subtle enough, it will be difficult to trace and so it will be nearly impossible to prove that it was not the attacker's intention to bomb the hospital.[85]

The impact of AI as a system technology on the military domain is not easy to predict. The technology can be applied in practice in myriad different ways. This is also one of the reasons why military forces are so interested in it: they do not want to miss out on a major strategic advantage. It is also why it is so important to look further than just the influence of autonomous weapons in this field. Moreover, military innovations do not form the only threats to national security and sovereignty; developments in civilian AI can also give rise to security issues, as the Strava example shows, and so these must also be closely monitored.

---

[82] Kate Crawford gives a striking example of how all manner of personal information can be derived from data. In 2013 a dataset of 173 million anonymized New York taxi journeys was released. Analysts were quickly able to de-anonymize the data and calculate the drivers' annual earnings, identify a number of famous people who had frequented strip clubs and deduce which drivers were Muslim from breaks taken during prayer times (Crawford, 2021: 111).

[83] Lin, 2019: 145.

[84] Libicki, 2019: 139–140.

[85] Lin, 2019: 149.

Key Points: Other Military Applications
- AI can contribute to the speed and quality of analyses and decision-making in the military arena.
- The use of AI to interpret and manipulate data throws up new technical challenges such as 'novelty detection' and 'edge cases'.
- AI can also be deployed to support all manner of military operations.
- It is ultimately impossible to predict what impact AI will have on warfare, which is why it is important to closely monitor the use and usefulness of this system technology in the military domain.

9.2.3 Security Beyond the Battlefield

Issues of national and international security are increasingly becoming intertwined.[86] In recent years there has been increasing attention for the impact of cyberattacks on vital infrastructure. Examples are the rise of ransomware and the major cyberattack on Ukraine in 2017. Such attacks have had ramifications all over the world, including effects for the container company Maersk that in turn lead to problems at the Port of Rotterdam. In 2019 the WRR published a report on this threat of digital disruption.[87] The rise of the use of sensors in physical objects (the 'internet of things') creates new vulnerabilities to AI attacks.[88]

In addition to the digital infrastructure, the flow of the digital information itself is now increasingly at the centre of conflicts and security concerns. We have already described the security implications that seemingly harmless smartphone apps and tourist photos can have. In fact, all manner of everyday information can play a role in countries' endeavours to outcompete each other. Take the details people share on social media. It is well-known that Russian secret services analysed President Trump's tweets to create a psychological profile of him.[89] The information that world leaders, officials or even ordinary citizens post on social media may contain valuable information for foreign rivals.

Not only can AI applications quickly analyse huge quantities of such data, they can also distil new patterns from it. Research by platforms like Facebook has revealed psychological insights about people. One claim is that the consistent use of black-and-white filters on Instagram and posting face-only photos are indicators of clinical depression.

[86] WRR, 2017.

[87] WRR, 2019.

[88] Brundage et al., 2018.

[89] Singer & Brooking, 2018: 61.

More and more research involves distilling medical data from video images of people.[90] Various nations are quite likely already collecting as much information as possible from the social media pages of other countries' citizens and running their algorithms on the data.

As well as gathering valuable information, the manipulation and sharing of data on social media have become the latest battlefield in what Singer and Brookings have termed the 'Like War'. The terrorist organization IS frequently used social media alongside more traditional military tactics. In fact, its campaign was fought online as well as on the battlefields of Iraq and Syria. It reported on military actions, recruited new members using professional action videos and sowed confusion in the cities it attacked through its own Twitter accounts. Much like the Nazis used the radio for fast communication and to spread confusion in France in 1940 (helping them bypass the impenetrable Maginot Line), so IS deployed social media as part of a twenty-first century Blitzkrieg.[91] Much of this manipulation was done by humans, but its huge impact was down to the way algorithms can make messages go viral. The physical military conflicts of today are fought simultaneously on a 'digital battlefield' where the belligerents try to frame each other and influence public opinion in their own country and beyond using social media.

AI, and more specifically machine learning, is being deployed in many ways in this 'information war' (see also text Box 9.5).[92] First of all through the method of microtargeting, which we have discussed in Chap. 3. This involves creating digital profiles of people who use social media to find out what messages resonate best with them. Sentiment analysis and natural language processing (NLP) are used to gain a greater understanding of specific populations so they can be sent targeted messages. One application of NLP is chatbots, which are becoming better and better at imitating people and can so be used to influence them.

Box 9.5: Modern Propaganda

Whereas propaganda used to be spread by broadcasting radio programmes into another country to influence opinions and feelings there, today it is spread through social media. People can now even be contacted directly by sending them friend requests and then bombarding them with information. By giving 'likes', citizens in far-away countries can participate in a conflict and contribute to the propaganda surrounding it. Online fundraising campaigns make this contribution even more tangible.

[90] The Chinese company Tencent has partnered with a British healthcare firm to develop software that can recognize people with Parkinson's disease in video footage (Ram, 7 May 2019).

[91] Singer & Brooking, 2018: 7.

[92] Kerr, 2019: 71.

The functioning of democratic institutions can be jeopardized by information wars. During the 2016 US presidential election, both domestic and foreign actors (amongst them the Russian state) attempted to influence public opinion through social media. Jeff Giesea worked for Donald Trump's online campaign. In a paper for a NATO journal, he described the tactics it used as 'memetic warfare', after the memes that go viral online.[93] He drew parallels between these tactics those employed by the propagandists of IS and Russia.

Another emerging application that deserves close attention is the deepfake. Deepfakes are falsified images or audio clips that are hard to distinguish from the real thing. In 2017 the company Lyrebird presented a hair-raisingly authentic-sounding recording of a conversation between Barack Obama, Hillary Clinton and Donald Trump. In the field of imaging, researchers have succeeded in converting a two-dimensional photograph into a three-dimensional face that could be given distinct expressions. Using ordinary cameras, other scientists were able to capture the facial expressions of an individual talking and then transfer them onto the face of another person (this is called 'deformation transfer'). Video and audio footage of Barack Obama, of which a great deal is publicly available, can now be used to create fake video clips in which he can be made to say anything the creator wants him to.[94] Lyrebird claims that it now only needs a few minutes of training data to create realistic deepfake audio fragments.[95]

It has also become possible to create completely new images from scratch using the generative adversarial networks (GANs) discussed in Part I. This technology is used in Hollywood films and computer games to create new objects and environments. But it can generate fake faces of non-existent people as well. These images are now so realistic that many observers are unable to distinguish them from photographs of real people. This means that it is now possible to create footage of events and human actions that never actually took place, but are indistinguishable from actual occurrences. So fake news, manipulation and polarization are new weapons in modern warfare. This is forcing defence forces and others to rethink their position.[96]

Thanks to the rise of technologies for producing fake material, technology researcher Aviv Ovadya foresees something he calls the 'infocalypse', a compound of information and apocalypse.[97] This will be brought about through the combination of deepfake images, chatbots that can imitate people very accurately ('laser phishing') and all kinds of other forms of manipulation that make it almost impossible to distinguish fake from real. Even if the material is recognized as fake with hindsight, realistic-looking falsified images can still cause plenty of damage if people initially believe they are real.

---

[93] Giesea, 2016.

[94] Singer & Brooking, 2018: 254–255.

[95] Schick, 2020: 148.

[96] Ministerie van Defensie, 2020.

[97] Warzel, 2018.

Current phenomena such as fake news and hacked Twitter accounts that spread false reports are but a small taste of what will become possible in the infocalypse. That could deepen social divisions by pitting groups against each other and augmenting distrust in institutions, but it could also lead to general apathy towards news reports when it appears that anything and everything can be manipulated. The philosopher Daniel Dennett even speculates that the end of the modern age of photographic evidence is approaching and that we will return to a world where people rely more on memory and trust than on incontrovertible proof.[98] These technologies can also be used to stifle critics. An Indian journalist who was highly critical of the government was inserted into a deepfake pornographic video that was distributed by politicians.[99] More and more countries are using disinformation as a weapon, too. According to researchers from the University of Oxford, while 28 nations carried out disinformation operations in 2017 that number had increased to 70 by 2020.[100] This phenomenon thus now forms a threat to the proper functioning of democracy.[101] Which brings us to the broader phenomenon of non-military threats involving AI, our next topic.

Key Points: Security Off the Battlefield
- In addition to the digital infrastructure itself, the information available through this infrastructure is increasingly a cause of security concerns.
- AI can be used to distil sensitive information from the increasing number of data sources in the civilian domain.
- State and non-state actors are engaged in manipulating and influencing people in an information war.
- All manner of applications of AI, such as microtargeting and deepfakes, are being used in the information war and are forming a growing threat to democracy.

9.2.4 Digital Dictatorship

The question of how a technology like AI affects different political regimes is also relevant to international security. For a long time, the answer was simple and reassuring: modern technology decentralizes and democratizes. For example, its complexity was seen as a factor in the failure of Soviet central planning. Centralized regimes were thought incapable of generating innovation and resolving the issues of co-ordination.[102]

As we have seen in Chap. 5, from the outset there was a strong ideological current that the internet was a force that could counter centralized power and free

[98] Dennett, 2019: 46.

[99] Schick, 2020: 125.

[100] Schick, 2020: 85.

[101] See also the interview with Dr Hans-Jakob Schindler in Semaan, 16 March 2020.

[102] See Hayek (1994) and the analysis by Fukuyama (1992) of the fall of the Soviet Union.

individuals from the grip of the authorities. The Arab Spring that began in 2011 has been partly attributed to the democratizing effect of internet platforms. However, this view of digital technology is changing. The author Evgeny Morozov argues that, while individuals and companies were the first to find their way to digital technology, governments have now discovered it too and they are proving that internet technologies are just as suitable for increasing centralization, control and surveillance.[103] This applies to the regimes of countries such as Iran, Russia and China, of course, but also to Western security services such as the American NSA.

There are even reasons to believe that AI technology can actually facilitate centralization. It allows governments to monitor their populations cheaply and on a large scale. The Stasi had to employ a huge network of human informants to spy on a section of the East German population. This limited the scope and extent of its activities. Now algorithms can do the job by analysing patterns in much larger quantities of data than were ever available before. This is another aspect of the dual-use nature of AI. It no longer takes huge additional investments in money and personnel to conduct mass surveillance, because to a large extent governments can simply build on the capacities already found in private-sector applications. This also makes it more attractive for them to use AI for such purposes.[104] The simple awareness that one might be under surveillance can also lead to 'chilling effects' and self-censorship by the public.

Previously decentralization was necessary in order to be able to co-ordinate the distribution of information. The free market generates an immense number of signals concerning supply and demand that need to be interpreted to set prices. No twentieth-century planning office could do it better. Looking at the example of navigation apps, it is now possible for a central organization to collect all the data they generate and use it to manage traffic and create the optimum flow. The founder of the Chinese platform Alibaba, Jack Ma, argues that AI thus enables better central planning. With sufficient information planners can better understand, predict and manage the economy.[105]

Technologies such as AI offer new opportunities to influence human behaviour and subtly nudge people in a certain direction. The fear is that not only computers will be 'programmed' in this way, but also people.[106] The authors of an article in *Scientific American* point to the danger of a computerized society with totalitarian tendencies – a digital version of George Orwell's Big Brother.[107]

Of course, technologies such as AI can also strengthen democracy. In addition to the older ideas of decentralization and participation, more and more parties are using the new instruments for democratic objectives. GVA Dictator Alert is an algorithm that scans flight data to warn when a dictator is landing at Geneva Airport. In

[103] Morozov, 2011.

[104] Wright, 2019a: 38

[105] Wright, 2019b: 28. The American company Predata, for example, also generates analyses based on web-monitoring to predict future hot spots and bottlenecks.

[106] Helbing et al., 2019.

[107] To denote the difference from older forms of surveillance, Shoshana Zuboff speaks of the 'Big Other' (Zuboff, 2019).

Chap. 7 we gave examples of civil society actors who use AI to promote equity and inclusion. There are several projects that are currently using AI to achieve the UN's Sustainable Development Goals (SDGs) – for example, to improve the position of small farmers in developing countries.[108] AI is also being used to monitor human rights violations.[109]

At the same time the momentum of the non-democratic effects of AI seems to be growing, driven partly by the rise of authoritarian countries. Political scientists speak of three historical waves of democratization: the first from 1820 to 1926, the second just after the Second World War and the third from 1975 onwards.[110] Every wave so far has been followed by a democratic reversal or countermovement. Nicholas Wright makes an interesting suggestion that is relevant here. Every setback for democratization was accompanied by a different form of dictatorship. The first wave was stalled by fascism. The second was followed by 'bureaucratic authoritarianism', a term for the kind of dictatorships found in Latin America and elsewhere from the 1960s onwards. Wright claims that the third wave could give way to 'digital authoritarianism'.[111]

According to the annual Freedom House study, however, this democratic reversal has been going on for a good while. For some time, the regimes of leaders like Rodrigo Duterte, Recep Erdogan, Vladimir Putin, Viktor Orbán and Jair Bolsonaro have been referred to as 'illiberal democracies', a concept in which new technology does not play a central role. It is possible, though, that digital technologies are now increasingly reinforcing such authoritarian governments.

Not only is a form of dictatorship emerging that relies on digital technology, but this model of governance is also increasingly being exported – and not just to relatively weak democracies in Africa or Latin America, say, but even to developed ones in Europe. Steven Feldstein has developed an AI Global Surveillance Index and revealed that at least 75 countries are using forms of AI surveillance in smart city platforms, facial recognition systems and smart policing. Chinese companies play a key role here, but firms from the US (Cisco, IBM), France (Thales, Teleste), Japan (NEC) and Germany (Bosch) are also contributing their expertise.[112]

A 2020 Amnesty International report on the export of European surveillance technology uncovered the activities of a Dutch company, Noldus Information Technology. It supplied the product FaceReader, which analyses facial expressions, to the Chinese Ministry of Public Security – according to the report, a heavy user of biometric data for mass surveillance.[113] The global proliferation of such technologies is a problem for the international order and the values that countries like the Netherlands want to uphold.

To shed more light on the nature and extent of digital dictatorship, we end this chapter with two case studies. In them we examine the phenomenon in more detail using the examples of China and Russia, the countries with the most advanced capabilities in this domain.

---

[108] Hirsch Ballin, 2021: 29–30.

[109] Isha Salian, 2019.

[110] Huntington, 1991.

[111] Wright, 2019b: 24.

[112] Feldstein, 2019.

[113] Amnesty International, 2020: 29.

Key Points: Digital Dictatorship
- Technologies such as AI can have a decentralizing and democratizing effect. But authoritarian regimes are also increasingly capable of using such technologies for their own ends, and we now speak of 'digital dictatorships'.
- The instruments of such dictatorships are increasingly being exported, putting pressure on democracies worldwide.
- Not only are authoritarian regimes contributing to this, but so too are Western companies.

9.2.5 Case Study: Digital Dictatorship in China

We saw in Sect. 9.1 how China has embraced AI and is pursuing global leadership through the AIDP. The country also uses this technology explicitly to monitor and control its own population.

It is important to realize that, while AI does play an important role here, it is being used as part of a much broader set of technologies and non-digital methods. One key example is the 'grid management system', where 'grid managers' are responsible for collecting information about a section of a neighbourhood. The Golden Shield Project develops broader digital technologies for governing the population and coordinating the actions of government. Data is also collected through a large network of sensors in the physical environment, part of the Internet Plus project.[114]

Another component of the Chinese data collection strategy is SkyNet (oddly enough, the same name as the malicious machine that turns against mankind in the film *The Terminator*), a programme to install a nationwide network of CCTV cameras. By 2010 it had already installed 800,000 cameras in Beijing, and in 2015 the police claimed they could now monitor 100% of the city. In the same year the National Development and Reform Commission (NDRC), the state planning body, announced plans to monitor all public spaces and leading industries with a surveillance system entitled Sharp Eyes by 2020.

AI is required to analyse so many sources of data, and companies such as Hikvision, SenseTime, Yitu and Megvii are developing smart cameras for this purpose. SenseTimewants to be able to monitor 100,000 high-resolution video feeds simultaneously and to identify and track individuals in real time using this technology. In 2018 the police used the system to identify and arrest a fugitive from among 60,000 concertgoers.[115] Facial recognition is in wide use in China.[116]

Pivotal to the aim of controlling the Chinese population is the famous 'social credit system'. This is not actually a single system, but comprises a number of

[114] Hoffman, 2019: 51–52.

[115] Polyakova & Meserole, 2019: 3–4.

[116] This technology is used, for instance, to identify drivers for the ride-hailing app Didi, to transfer money via Alipay, to collect train tickets and to gain access to tourist attractions (Agrawal et al., 2018: 219).

regional and national projects. At the national level there is the Xinyi+ Project, in which companies such as Ant Financial (financial affairs), Didi Chuxing (a 'Chinese Uber') and Ctrip (a travel agency) are co-operating in the field of transport and rental in order to exclude certain people and offer greater convenience to others, based on their scores. An example of a regional system is found in the city of Fuzhou, where the company JD Finance is using AI to develop a 'smart city credit platform'.[117]

More and more information is coming to light about the oppression of Muslim Uyghurs in China's Xingjiang province, particularly focusing on the 're-education camps' in which more than a million people have been imprisoned. Also relevant is the Strike Hard campaign launched in 2014, which also has a strong digital component and uses technologies such as AI. The numerous police checkpoints in the province are equipped with biometric sensors and iris scanners, and can monitor the CCTV cameras installed in the local area. The DNA of many Uyghurs has been collected and they are forced to install the Jingwang app, which not only enables the authorities to track and block their messages but also provides direct access to their phones. The police monitor the population to ensure they have actually installed these 'electronic handcuffs'.[118] All cars in the province are required to have navigation software installed that runs on BeiDou, the Chinese version of GPS, and drone swarms are used to monitor places where CCTV cameras cannot be installed. According to a paper by the Brookings Institution, in addition to physical prisons China also has "the world's largest open-air digital prison".[119]

China is thus a textbook example of how AI can be used for the goals of authoritarian regimes. It is all the more important to keep a close eye on these developments now that such technologies are increasingly being exported, including by state actors including the military (the People's Liberation Army) and the Ministry of Public Security, state-owned enterprises such as CEIEC and private companies like Huawei, ZTE and Tencent.[120]

These exports are ending up all over the world.[121] China's so-called 'Great Firewall' is being copied in Vietnam and Thailand. The company Yitu supplies portable cameras with AI for facial recognition to the Malaysian police and tendered to install facial recognition cameras in public spaces in Singapore.[122] Ethiopian security services use ZTE's telecommunication products to monitor journalists and activists. Zimbabwe and Angola have both signed AI deals to bolster their own regimes. In Venezuela ZTE has a contract to roll out a national ID card, a payment system and a 'homeland database' that will allow the regime to introduce the Chinese social credit system to the country. Surveillance systems are used in

[117] Ahmed, 2019: 57–59.

[118] Singer & Brooking, 2018: 101.

[119] Polyakova & Meserole, 2019: 5.

[120] Weber, 2019: 77–78.

[121] The export of digital authoritarianism is often also based on simple mass production by humans. The so-called '50 Cent Army', of which two million Chinese are said to be members, is named after the amount of money they are said to receive for each positive post about China.

[122] Wright, 2019a: 36.

government cameras in Ecuador, and in Dubai Chinese technology is used for the Police Without Policemen programme to fight crime with the aid of video surveillance and facial recognition technology.[123] It should also be mentioned that Chinese companies like Huawei and SenseTime are entering into partnerships with universities around the world, including some in the West.

9.2.6 Case Study: Digital Dictatorship in Russia

The other major developer and exporter of technologies in support of digital dictatorships is Russia. This activity is founded on a long tradition of controlling information that goes back to the time of the Soviet Union and began to be reinstated quite soon after the fall of that authoritarian regime. In 1995 a law was passed that allows the FSB, the successor to the KGB, to monitor all private communications. Since then, a series of acts has increased the government's grip on RuNet, as the Russian internet is called. These allow the authorities to block websites, register bloggers, store data and give the FSB access to encrypted data. The pressure put on VKontakte, Russia's largest social media platform, to provide access to information about opposition leader Alexei Navalny's presidential campaign (among other things) was enough to force the company's boss to sell his shares. He later founded the chat app Telegram, which also clashed with the Russian authorities because of the encryption it uses.[124]

Like China, Russia's instruments of digital dictatorship are also exported abroad (see Box 9.6). The importance of digitalization in conflict situations has long been recognized. The Russian general Valery Gerimasov is said to have emphasized the use of the asymmetrical possibilities offered by the internet for international competition. The FSB has since directed 75 educational and research institutions to study how information can be weaponized. A NATO researcher summarized the strategy as the '4Ds': "dismiss the critic, distort the facts, distract from the main issue and dismay the audience".[125] Traditional and online media channels such as *Russia Today*, *Sputnik* and *Baltica* are playing an important role in this information war. Russia's seizure of Crimea in 2014 has been dubbed 'Schrödinger's War' because of the way it exploited disinformation, confusion and hybrid warfare.[126]

There are clear differences between the instruments of digital dictatorship exported by China and Russia. While China's so-called '50 Cent Army' of online

[123] Polyakova & Meserole, 2019: 6.

[124] Kerr, 2019: 64–67.

[125] Singer & Brooking, 2018: 107.

[126] Singer & Brooking, 2018: 205.

Box 9.6: The Russian Toolbox

Russia uses a variety of instruments for internal control. One key surveillance system is SORM, the System for Operative Investigative Activities. Under this internet service providers are required to install a special device that enables the secret services to copy and monitor all their online traffic.[127]

In addition to hardware, Russia also carries out digital control using people; paid troll factories, so-called 'hacktivists' and the notorious Internet Research Agency (IRA) in St Petersburg are used to project online influence at home and abroad. AI plays an important part in the strategy, too. The country started using Safe City in 2015. This system recognizes faces and moving objects on video images captured by numerous cameras and shares the data directly with the authorities. Between 2012 and 2019 the country invested US$2.8 billion to equip all the host cities of the 2018 FIFA World Cup with the system. More than 100,000 cameras in Moscow are linked to facial recognition software provided by the company NTechlab.[128]

The Russian security services also use the Semantic Archive Platform supplied by the software company Analytical Business Solutions to collect, process and analyse open-source data.[129]

commentators is deployed to spread positive messages about their own nation, Russia is mainly concerned with spreading negative news in countries where it wants to sow discord. Nina Schick draws a parallel between this current policy and the 'active measures' during the time of the Soviet Union. Their aim was to change others' perception of reality to such an extent that they were no longer able to draw sensible conclusions about how to defend their own interests.[130]

A second difference is that the Chinese export product is technologically far more advanced and expensive, because it enables almost total control of the internet. Russia's tools rely more on specific hardware and the use of intimidation and legislation to control the population. According to a study by the Brookings Institution, the Russian product may appeal more to poorer regimes that lack the resources to control the entire internet in their country.

The SORM system is widespread in the countries that were formerly part of the Soviet Union, such as Kyrgyzstan, Belarus and Kazakhstan. The companies that export it, Protei and Peter-Service, also have telecom businesses in the Middle East and Latin America as customers. The Semantic Archive Platform is used in Belarus, Ukraine and Kazakhstan.[131]

---

[127] Soldatov & Borogan, 2015.

[128] Polyakova & Meserole, 2019: 8.

[129] Morgus, 2019: 92.

[130] Schick, 2020: 54.

[131] Polyakova & Meserole, 2019: 10.

One final characteristic of the digital dictatorship export drive concerns Russia's policy towards international institutions and forums. This attempts to blur the line between cybersecurity and information control, so that countries concerned about the former will also want to do more about the latter. Moscow has submitted documents to the UN proposing an 'International Code of Conduct for Information Security' that, if implemented, would pose a threat to human rights and international law. Russia also wants to bring the internet under the control of the ITU, and hence states. Finally, it wants the internet cable infrastructure between the BRICS – the emerging economies of Brazil, Russia, India, China and South Africa – to bypass the US.[132] We looked at the importance of such international forums in the first part of the chapter on competitiveness. The Russian policy just described illustrates the extent to which negotiations in those arenas are intertwined with security issues.

Key Points: Digital Dictatorships in China and Russia

- China and Russia are world leaders in digital dictatorship and the export of the instruments used to enable it. Each employs a different set of instruments to achieve authoritarian objectives, but both are exported all over the world.
- The Chinese model is technologically advanced and primarily aimed at encouraging positive coverage of the regime.
- The Russian model is less advanced and uses more hardware and analogue forms of intimidation. Abroad, it aims mainly to create confusion and conflict.
- The phenomenon of digital dictatorship is complex and multifaceted and deserves serious attention.

9.3 In Conclusion

The overarching task of international positioning all about a country's place and role in the international arena. This includes how it interacts with other nations, but also with non-state actors such as companies and criminal organizations. In this chapter we have seen that there are several international arenas where countries can and must take a stance on matters such as autonomous weapons, the regulation of AI and standardization, plus the challenges they all entail. In the final part of this report, we consider how this relates to the idea of 'AI diplomacy' and its implications for policy.

We have also seen how phenomena in which AI plays a role, such as the information war and digital dictatorship, are a cause of true concern. They pose a threat to

[132] Morgus, 2019: 93.

freedom and democracy worldwide, but also to national security. It is important to invest in a response to these phenomena and to formulate answers to them.

The two issues of competitiveness and national security are clearly intertwined in the context of AI. In this chapter we have highlighted the importance of AI diplomacy and raising awareness of the risks the technology poses to national security.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Ahmed, S. (2019). Credit cities and the limits of the social credit system. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 55–61). Air University Press.

Amnesty International. (2020). *Out of control: Failing Eu Laws for digital surveillance export*. Amnesty International.

Bendett, S. (2019). *The development of Artificial Intelligence in Russia.* In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 168–198). Air University Press.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G., Steinhardt, J., Flynn, C., HÉgeartaigh, S., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evens, O., Page, M., Bryson, J., Yampolskiy, R., & Amodei, D. (2018). *The Malicious use of Artificial Intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute.

Buruma, Y. (2020). International Law and Cyberspace. Issues of sovereignty and the common good. In *International Law for a Digitalised World* (pp. 69–111). knvir/T.M.C. Asser Press.

Bygrave, L. (2021). The 'Strasbourg Effect' on data protection in light of the 'Brussels Effect': Logic, mechanics and prospects. *Computer Law & Security Review, 40*, 105460.

Cihon, P. (2019). *Standards for AI governance: International standards to enable global coordination in AI Research & Development* (Technical Report). Future of Humanity Institute.

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Creemers, R. (2019). The international and foreign policy impact of China's Artificial Intelligence and big-data strategies. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 129–135). Air University Press.

Dennett, D. (2019). What can we do? In J. Brockman (red.), *Possible minds: Twenty-five ways of looking at AI* (pp. 41–53). Penguin.

Ding, J. (2018). *Deciphering China's AI dream* (Future of Humanity Institute Technical Report). University of Oxford.

Ding, J. (2019). The interests behind China's Artificial Intelligence dream. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 43–47). Air University Press.

Drezner, D. (2019). Economic Statecraft in the age of Trump. *The Washington Quarterly, 42*(3), 7–24.

Edgerton, D. (2008). *The shock of the old: Technology and global history since 1900*. Profile books.

European Commission. (2018, December 7). Member States and commission to work together to boost Artificial Intelligence "Made In Europe". Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6689

Feldstein, S. (2019). *The global expansion of AI surveillance*. Carnegie Endowment for International Peace.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

Fukuyama, F. (1992). *The end of history and the last man*. Simon & Schuster.

Giesea, J. (2016). It's time to embrace memetic warfare. *Defence Strategic Communications, 1*, 67–75.

Hall, P., & Soskice, D. (2001). *Varieties of Capitalism: The institutional foundations of comparative advantage*. Oxford University Press.

Harari, Y. N. (2019). Who will win the race for AI? *Foreign Policy Magazine*, Winter 2019. Available at: https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/

Hayek, F. (1994). *The road to Serfdom*. University of Chicago Press.

Helbing, D., Frey, B., Gigenrenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R., & Zwitter, A. (2019). Will democracy survive big data and Artificial Intelligence? In D. Helbing (red.), *Towards digital enlightenment* (pp. 73–98). Springer.

Hillman, J. (2019). *Infrastructure and influence: The strategic stake of foreign projects*. Center for Strategic and International Studies.

Hirsch Ballin, E. (2021). *Mensenrechten Als Ijkpunten Van Artificiële Intelligentie* (WRR Working Paper nr. 46). Wetenschappelijke Raad voor het Regeringsbeleid.

Hoffman, S. (2019). Managing the State: Social Credit, surveillance, and the Chinese Communist Party's Plan for China. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 48–57). Air University Press.

Huntington, S. (1991). *The third wave: Democratization in the late 20th century*. University of Oklahoma Press.

Kaiser, W., & Schot, J. (2014). *Writing the rules for Europe: Experts, Cartels and International Organizations*. Palgrave Macmillan.

Kerr, J. (2019). The Russian Model of digital control and its significance. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 62–74). Air University Press.

Lee, K. F. (2018). *AI Superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.

Leonard, M. (red.). (2016). *Connectivity Wars: Why migration, finance and trade are the geo-economic battlegrounds of the future*. European Council on Foreign Relations.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Libicki, M. (2019). A hacker way of warfare. In N. Wright (red.) Artificial Intelligence, China, Russia and the global order (pp. 137–142). Air University Press.

Lin, H. (2019). Escalation risk in an Artificial Intelligence-infused world. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 143–152). Air University Press.

Luttwak, E. (1990). From Geopolitics to Geo-Economics: Logic of conflict, grammar of commerce. *The National Interest, 20*, 17–23.

Ministerie van Defensie. (2020). *Defensievisie 2035*. Ministerie van Defensie.

Mols, B. (2019). *Internationaal AI-beleid. Domme data, slimme computers en wijze mensen*. WRR Working Paper.

Morgus, R. (2019). The spread of Russia's digital Authoritarianism. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 89–97). Air University Press.

Morozov, E. (2011). *The Net Delusion: How to Not liberate the World*. Penguin.

National Security Commission on Artificial Intelligence. (2021). *Final Report*. NSCAI.

Pasquale, F. (2020). *New Laws of Robotics: Defending human expertise in the age of AI*. Harvard University Press.

Polyakova, A., & Meserole, C. (2019). *Exporting digital authoritarianism: The Russian and Chinese models* (Policy Brief, Democracy and Disorder Series). Brookings.

Rao, A., & Verweij, G. (2017). *Sizing the Prize: What's the real value of AI for your business and how can you capitalise?* PricewaterhouseCoopers. Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Rathenau Instituut. (2021a). *Waardevol Gebruik Van Menselijke DNA-Data. Onderzoek Naar Het Borgen Van Publieke Waarden In De Waardeketen Van DNA-Data*. Rathenau Instituut. Available at: https://www.rathenau.nl/sites/default/files/2021-05/Waardevol_gebruik_van_menselijke_DNA%20data_Rathenau_Instituut.pdf

Rathenau Instituut. (2021b, June 7). *International mobility of AI scientists*, Factsheet. Available at: https://www.rathenau.nl/en/science-figures/international-mobility-ai-scientists

Rühlig, T. (2020). *Technical Standardisation, China and the future international order: A European Perspective*. Heinrich Böll Stiftung.

Salian, I. (2019, April 4). AI in the sky aids feet on the ground spotting human rights violations. blog, *NVIDIA*. Available at: https://blogs.nvidia.com/blog/2019/04/04/human-rights-watch-ai-gtc/#:~:text=AI%20in%20the%20Sky%20Aids%20Feet%20on%20the%20Ground%20Spotting%20Human%20Rights%20Violations&text=In%20a%20traditional%20human%20rights,collect%20hospital%20or%20autopsy%20records

Scharre, P. (2018). *Army of None: Autonomous weapons and the future of war*. WW Norton & Company.

Schick, N. (2020). *Deep fakes and the Infocalypse: What you urgently need to know*. Octopus Publishing Group.

Scholvin, S., & Wigell, M. (2018). *Geo-economics as a concept and practice in international relations: Surveying the state of the art* (Working Paper nr. 102). Finnish Institute of International Affairs.

Sheikh, H., & Timmers, P. (2020, December 3). Na Trump is Het Tijd Voor 'Make Europe Great Again'. *NRC*. Available at: https://www.nrc.nl/nieuws/2020/12/03/na-trump-is-het-tijd-voor-make-europe-great-again-a4022477

Singer, P., & Brooking, E. (2018). *LikeWar: The weaponization of Social Media*. Mariner Books.

Smuha, N. (2019). From A 'Race To AI To A 'Race to AI regulation': Regulatory competition for Artificial Intelligence. *Law Innovation and Technology, 13*(1), 57–84.

Soldatov, A., & Borogan, I. (2015). *The Red Web: The struggle between Russia's digital dictators and the new online revolutionaries*. Public Affairs.

Timmers, P. (2019). Challenged by "Digital Soevereignty". *Journal of Internet Law, 23*(6), 12–21.

Timmers, P., & Dezeure, F. (2021). *Nederlandse Strategische Autonomie En Cybersecurity* (onderzoek in opdracht van de Cyber Security Raad). Cyber Security Raad. Available at: https://www.cybersecurityraad.nl/binaries/cybersecurityraad/documenten/rapporten/2021/02/18/onderzoeksrapport-digitale-autonomie/Onderzoeksrapport+%27Nederlandse+strategische+autonomie+en+cybersecurity%27.pdf

Tonin, M. (2019). Artificial Intelligence: Implications for NATO's Armed Forces. *149 stctts 19 E rev. 1 fin.*

Villani, C. (2018). *For a meaningful artificial intelligence: Towards a French and European strategy. AI for Humanity.*

Walch, K. (2020, February 9). Why the race for AI dominance is more global than you think. *Forbes*. Available at: https://www.forbes.com/sites/cognitiveworld/2020/02/09/why-the-race-for-ai-dominance-is-more-global-than-you-think/?sh=3a34ad2b121f

Warzel, C. (2018, February 11). Believable: The terrifying future of fake news. *Buzzfeed News*. Available at: buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news

Weber, V. (2019). Understanding the global ramifications of China's information-control model. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 76–80). Air University Press.

Weinberger, S. (2019). *The imagineers of war: The untold history of DARPA, The Pentagon agency that changed the world*. Vintage.

Wright, N. (2019a). Global Competition. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 35–41). Air University Press.

Wright, N. (2019b). Artificial intelligence and domestic regimes: Digital authoritarian, digital hybrid, and digital democracy. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 21–34). Air University Press.

WRR. (2013). *Naar Een Lerende Economie*. Wetenschappelijke Raad voor het Regeringsbeleid.

WRR. (2017). *Veiligheid In Een Wereld Van Verbindingen. Een Strategische Visie Op Het Defensiebeleid*. Wetenschappelijke Raad voor het Regeringsbeleid.

WRR. (2019). *Voorbereiden Op Digitale Ontwrichting*. Wetenschappelijke Raad voor het Regeringsbeleid.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *The AI index 2021 annual report*. Stanford University, Human-Centered AI Institute. Available at: https://arxiv.org/ftp/arxiv/papers/2103/2103.06312.pdf

Zielonka, J. (2008). Europe as a global actor: Empire by example? *International Affairs, 84*(3), 471–484.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Part III
Agenda: Conclusions and Recommendations for AI Policy in the Netherlands

Chapter 10
Policy for AI as a System Technology

Artificial intelligence is not just another technology – it is a 'system technology' that will fundamentally change our society. That is the key message of this report. Government and society therefore need to be much more aware of and actively involved in AI's integration into daily life. The government in particular needs to focus on five overarching tasks to help shape the integration process, because only then will it be able to continue to protect the civic values affected by AI. Such a challenge demands a policy infrastructure that reflects both a political and an administrative commitment.

AI is to our century what electricity was to the nineteenth and the internal combustion engine to the twentieth. It is not a concrete technology that can be overseen and managed by a group of experts or policymakers from one or more ministries. AI is everywhere, it is continuously being improved and it generates complementary innovations, which makes it a very versatile but also unpredictable phenomenon. But unpredictability and uncertainty about how to integrate or embed AI into society cannot be used as an excuse to sit back and watch it take its course. Rather, the potentially unlimited value that AI could deliver calls for a most carefully considered approach to this process. Government must also consider the broader agenda in this respect so that it can continue to intervene in and adjust the process in the future.

By examining AI through the lens of previous system technologies, we can learn a great deal about how system technologies are embedded. The lessons that we learned from embedding earlier system technologies form the basis for the recommendations that the WRR presents in this final chapter. The key point is that considering AI as a system technology has implications for the way we look at public values. History teaches us, as we have argued in Chap. 2, that the impacts of system technologies on public values cannot simply be categorised on a list. After all, given that AI has the potential to be applied throughout our entire society and we are currently only at the beginning of its development, the impact of AI on public values will not only be broad but also unpredictable. In the previous chapters and the final analysis of this chapter, we therefore approach public values in a way that agrees with the dynamic nature of AI.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_10

10.1 Five Tasks as Lessons from the Past

From an analysis of the history of previous system technologies, we have distinguished five overarching tasks for the integration of AI into government and society: demystification of what it is and can do, contextualization of its development and application, engagement by various parties, regulation of the technology, its use and the social implications and, finally, its national positioning in relation to other countries and international organizations (Fig. 10.1). We have discussed these tasks in detail in Part 2 of this report and recap them briefly here, also indicating what civic values are at stake and what risks are involved if we do not face up to these tasks.

AI as a System Technology

There is a rich body of academic literature discussing technological revolutions, epochal innovations and technical eras. A recurring central concept in this corpus is that of 'general purpose technologies', those not used for a specific purpose but applicable broadly throughout society. Examples include the steam engine, electricity, the combustion engine and the computer. In Chap. 4 we revealed how AI has the three characteristics of a general purpose technology: it is (1) ubiquitous, (2) subject to continuous technical improvement and (3) enables complementary innovations in other fields.

In this report we have labelled AI a system technology. On the one hand this points to the fact that – like electricity and combustion engines – it is part of a wider system of other technologies, while on the other we use this term to emphasize the systemic effect such technologies have on society.

What Do We Mean by AI?

In this report we have adopted the definition formulated by the High-Level Expert Group on AI (AI HLEG) of the European Commission: "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."

The broadest definition of AI equates it with the use of algorithms, while the strictest views it as the imitation of all human skills ('artificial general intelligence'). The former stretches the concept of AI enormously while the latter defines it out of existence. The AI HLEG version is sufficiently specific, while at the same time – by admitting non-concrete phenomena such as deep learning –leaving room for new techniques and developments.

AI and Digital Technology

AI is strongly intertwined with other digital technologies such as computing and data but does not coincide with them. One of the fathers of computing, Alan Turing, was also the inventor of the so-called Turing test, which is used to assess AI systems. AI is dependent on huge amounts of data, and internet data. Current deep learning methods require large amounts of digital

(continued)

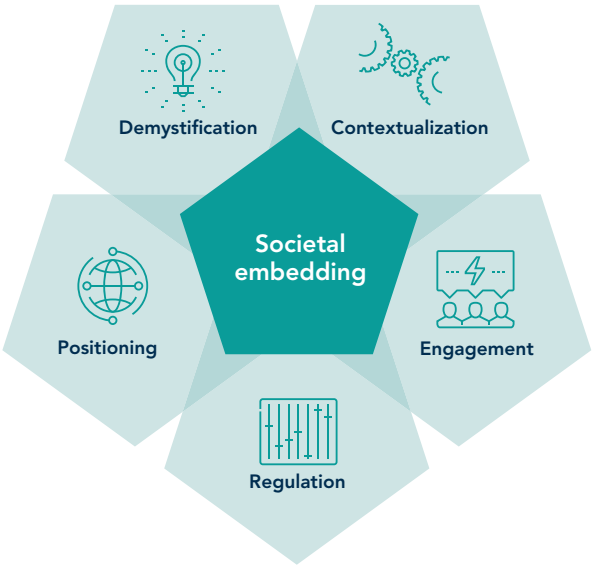


Fig. 10.1 Five tasks for the social integration of AI

information to work effectively. At the same time AI cannot be synonymized with these other technologies. We have outlined its development and the many ways it is linked to computers, data and the internet in Chaps. 2 and 3. But AI also has a separate scientific and historical background, with its own 'springs and 'winters'. While computers have been widely used since the Second World War, and the internet has been ubiquitous since the 1990s, the emergence of AI as a social phenomenon is a far more recent development. This is why it deserves individual attention.

10.1.1 Task 1: Demystification

The first overarching task – demystification – concerns preconceptions about AI as a technology. In fact, it is really about the question: *What is AI?* System technologies always go hand in hand with extreme preconceptions. Excessively high expectations lead to disillusionment and ill-considered applications, while exaggerated fears lead to rejection of a technology and unexploited opportunities. Clinging to such preconceptions will have a negative effect, particularly in the longer term. We argue that more realism is needed to be able to ask the right questions about societal integration and civic values. In the past we saw all manner of unrealistic expectations arise concerning the future of electricity and automobiles, driven by public demonstrations and races. Commentators thought that trains, the telegraph and later the internet would bring global peace by connecting the world. Conversely, the imagery surrounding earlier system technologies in the form of Frankenstein's

monster and words like 'electrocution' – which linked electricity to mortality – stirred up fears of these breakthroughs.

There are also numerous myths surrounding artificial intelligence. AI systems are said to be rational and objective, but also to work like an unfathomable 'black box'. It is thought that the technology could eventually match and even exceed all human capabilities, and even turn against humanity. In addition, there are all sorts of myths associated with digitalization in a broader sense, such as the idea – popular until quite recently – that the development of the internet should be ungoverned and, more importantly, unregulated. Another mistaken preconception is that there is no alternative to the current form of digital technology and that digitalization offers a solution to every problem.

If we do not address such ideas, society may come to rely too heavily on AI systems – with all manner of unwelcome consequences. They could also lead to AI being rejected and its benefits being reaped insufficiently, if at all. Finally, exaggerated preconceptions can prevent an open discussion on crucial questions surrounding the societal integration of a technology. Demystification primarily involves issues such as legal protection, the public's confidence in the technology, adequate provision of information and the quality of the public debate.

10.1.2 Task 2: Contextualization

The second task we have distinguished is contextualization. This concerns the application of AI and the question: *How will the technology work?* In other words, contextualization relates primarily to the technical ecosystem. System technologies do not function independently; they are dependent on other supporting technologies or their underlying facilities. An example is the car's dependence on the oil industry, petrol stations and a road network. Moreover, system technologies become connected over time to other emerging technologies, as the car is connected to electronics. In addition to the technical ecosystem, contextualization is also about the role of the social ecosystem. At a macro level, a lasting effort will be needed to adapt work processes, value chains and knowledge development. Only after this has been done will organizations be in a position to use the technology effectively and become more productive. At the micro level this will require behavioural change and effective interaction between the users and the new technology.

AI also requires various supporting technologies or facilities, such as data, telecommunication networks, chips and supercomputers. Furthermore, we are already seeing increasing connectivity between AI and other new technologies such as 5G networks, the 'Internet of Things' and quantum computing. As far as macro level developments are concerned, the expectation that AI will make human work redundant on a massive scale appears unfounded. Rather, a process of intensive training and practice will be required to make it an effective tool in the workplace. At the micro level the task is to achieve effective human-machine interaction. Here the relative autonomy of AI systems forms the main challenge.

Insufficient attention to supporting technologies and facilities (such as good quality, secure and readily available data and networks) will lead to poorly functioning AI systems, underutilization of opportunities or stagnation of development. Just as the road network was essential for the use of the car, so AI requires technical adaptations to the ecosystem. Attention to that aspect is particularly important in those areas where a country can benefit most from AI. For the Netherlands this means areas in which the country has traditionally had a strong international position (such as agriculture and services) and areas where AI can help address existing challenges (such as those in healthcare). Other countries will obviously make other choices, such as manufacturing in the case of Germany or defence in the case of France. Insufficient attention to the social ecosystem will also lead to poor implementation and to all manner of issues, or even rejection of the technology if the users of AI systems are not adequately equipped to deal with these issues. So not only are the quality and safety of AI applications at stake, but also the public benefits that can be gained in areas ranging from wider access to better quality healthcare and education to better government services.

10.1.3 Task 3: Engagement

The third overarching task, engagement, concerns the societal environment of AI and the question: *Who should be involved?* When new system technologies arise, large companies and governments have the means and interests to be early adopters. Civil society parties usually do not become involved until later. As such these new technologies initially only reinforce the existing balance of power in society. Consider how the deployment of the steam engine in factory production processes marginalized workers or how adapting the infrastructure for the automobile forced non-drivers (at that time mainly poorer people) off the roads.

Stakeholders' engagement in society can take a wide range of forms. At one extreme is violent resistance, while non-violent protests and calls for bans are also ways of restricting a new technology. At the other end of the spectrum, civil society can play its part in improving a technology – for example, by contributing its own expertise or by applying it in its own practices.

AI in its current manifestations also reinforces existing imbalances. Less affluent citizens, ethnic minorities and women are among the groups discriminated against by algorithms. Civil society is now mobilizing to protest against a number of controversial applications, such as autonomous weapons, facial recognition and the use of AI by the police. Much of this opposition takes the form of protests and calls for bans, but strikes are on the table too. But when it comes to more co-operative forms of engagement aimed at the useful social integration of AI – such as contributing expertise or using the technology to tackle challenges related to climate change, poverty, or human rights –much still remains to be gained.

What will happen if engagement lags behind? It is likely that existing imbalances will be reinforced and the balance of power between governments and large

companies on the one hand and citizens on the other further distorted. In particular, the rights of various weaker social parties will be threatened. So, if there is not enough engagement in AI, fundamental rights such as equality, privacy, non-discrimination and autonomy as well as democratic principles like participation, inclusion and pluralism will all be at stake. A regulatory framework is an important prerequisite for shaping this engagement, which brings us to the fourth task for government.

10.1.4 Task 4: Regulation

The task of regulation is relevant at the societal level, focusing on the question: *What frameworks are required?* When a new technology leaves the lab, it is initially difficult to oversee, adapt or develop the necessary frameworks. Much is still unclear about its nature and effects, and so as long as AI is not yet embedded across the full breadth and numerous contexts of society it is difficult to know what specific civic values it might compromise.

In the early phase, technology companies often promote self-regulation by the sector or argue that users themselves can be relied upon to safeguard certain values. Gradually, however, structural issues come to light that require a more active government role. Other system technologies were initially concentrated in the hands of a few companies, such as (in the US) GE and Westinghouse in the case of electricity or the 'big three' in Detroit when it came to automobiles. But other factors also contribute to the need for a more active government role. As technology becomes more deeply embedded in society, it increasingly touches upon civic values that fall under the responsibility of government. With time the broader social effects of a new technology become clearer, and so policy and legislation become less and less tentative. From this point government needs to develop a broader and more unified legislative agenda; separate dossiers no longer suffice.

With AI we saw an initial focus on self-regulation. Today the momentum has shifted towards more active government intervention (the European draft AI Act is a good example of this). At the same time structural issues are coming to light, which government will also have to address if it wants to manage the effects of the technology. These include the concentration of power in the hands of large companies, the growth of surveillance in society and increasing public-sector dependence on commercial businesses.

Of course, there are no panaceas or 'silver bullets' for the regulation of system technologies. Properly embedding a technology in society requires a broad set of measures developed over a long period of time. An example is the internal combustion engine that made the motor car possible: seat belts, insurance, number plates, airbags, driving tests, traffic rules and road signs were all steps that contributed towards its social integration – a process that continues to this day because the car and its environment are being developed continuously. It was impossible to foresee that all these measures would be necessary when the car was first introduced.

However, this does not mean that the legislator can endlessly vacillate about what the best approach might be. The task of regulation requires both a greater role for government and a broader legislative agenda. If government waits too long to develop its agenda, lawmaking will be left behind by the dynamism of the process. Meanwhile, other stakeholders will have taken control of the way AI is embedded to the extent that it will be almost impossible to reverse this development. Existing frameworks then lose their legitimacy and our social system based on shared civic values will come under threat.

10.1.5 Task 5: Positioning

The final overarching task we have identified is positioning. This relates to the international arena and is about the question 'what is our international position?' Firstly, this concerns the role that a new system technology can play in boosting national competitiveness. In the past technologies like the steam engine, electricity and the internal combustion engine helped many countries strengthen their competitive position in the international arena. They even influenced the nature and outcomes of international conflicts; railways were essential to Prussia's victory over France in 1870–1871 and the first computer code-breakers contributed towards vanquishing the Germans in the Second World War. These two dynamics feed the idea of a global race to dominate a new technology and some countries even try to develop and maintain such innovations completely within their own borders. However, history teaches us that system technologies always have a global character, and that international co-operation is in fact the best way to improve individual countries' competitiveness and security.

The same dynamic is involved in AI. There is much talk of an 'AI race', with the US and China setting the pace. Many countries have therefore developed AI strategies in recent years in order to join this race and to deploy AI to strengthen their competitiveness. But there is also a growing awareness of its impact in the areas of conflict and security. The most prominent application is so-called autonomous weapons. Several international initiatives have now been launched to control the development and extent of this new arsenal. But there are many other military and civilian applications of AI that can threaten national security.

If countries fail to develop their position in AI and pay too little attention to broader co-operation at the international level, they will miss out on opportunities to strengthen their competitiveness. Moreover, not enough consideration of their international position in AI will leave countries insufficiently aware of and prepared for the security risks the technology brings.

10.1.6 Five Tasks, Five Transitions

These five overarching tasks are thus critical to AI's successful integration into society. But it is also important to emphasize their interrelations. Demystification, for example, strengthens society's ability to engage with AI technology. So, although these tasks can be separated analytically, in practice a combined approach is needed. The stakes involved in integrating AI successfully are high (utilizing innovation potential, societal acceptance, etc.), and the process puts various civic values at risk – although it is impossible to predict in advance which will be affected, or how. We have argued elsewhere in this report that it is impractical to draw up an exhaustive list of civic values and analyse them all in the light of AI. The unpredictable nature of system technologies necessitates a more dynamic perspective. We therefore suggest that the debate on AI and its consequences for society be conducted on the basis of the five identified tasks. Many contemporary and future issues can be addressed within this broad framework.

With this cluster of five overarching tasks, we thus offer a long-term framework for AI's societal integration. This, however, does not answer the question of what needs to be done in the short term in the light of these tasks, particularly from the government point of view. In other words, what transitions are involved? Below we describe the transition associated with each task (Fig. 10.2) and then, in the next section, explain each transition with the help of concrete recommendations. The transitions are:

1. **From fiction to facts;**
2. **From abstraction to application;**
3. **From monologue to dialogue;**
4. **From reaction to action; and,**
5. **From nation to network.**

10.1.7 A Broad Agenda for AI

The five transitions represent an AI agenda for the years ahead. Our first observation in this respect is that the breadth of this agenda implies that national governments cannot be solely responsible for its implementation. Across all five tasks a variety of actors in society have a role to play and responsibility to take. For example, academics will be needed in the transition from fiction to a more facts-based approach of AI. Ordinary citizens can help shape this transition too, by informing themselves about AI or by following the 'national AI course'. The media also have an important role to play in informing people who are unmotivated or unable to find out more for themselves. Meanwhile, much of the transition from abstraction to its application will fall to industry. Government bodies may later become major users of AI, but initially all manner of private-sector players will need to answer the question of how

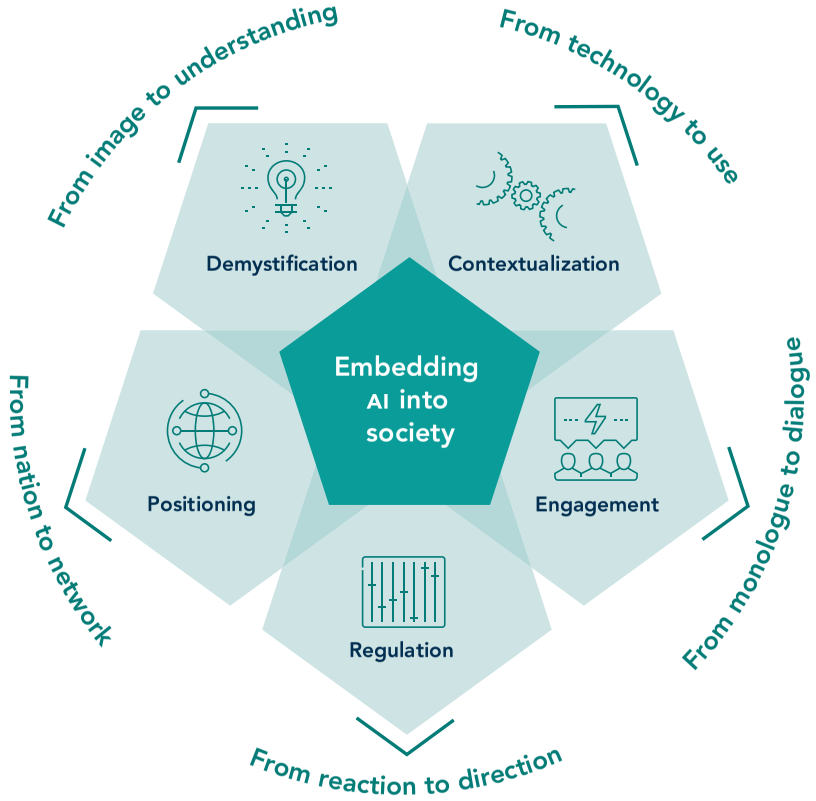


Fig. 10.2 Every task requires a transition

it can be used in practice. In short, all the tasks and the associated transitions will require a collective effort by various actors.

A second observation is that not all steps towards achieving the tasks will require the same effort. In fact, some things will happen automatically. As society collectively gains more experience with AI, for example, we can expect a degree of demystification and thus a more realistic awareness of its implications. Moreover, initiatives are already emerging in some areas. These include autonomous weapons and their effect on countries' international positions, which are receiving attention around the world. When it comes to regulation, not every new application of AI will require brand new legislation. Existing rules already provide the necessary framework for a variety of applications, and in some cases self-regulation by companies or other societal parties will suffice – for the time being, in any case.

In this report that was originally written for the Dutch government, we are therefore selective in the tasks we highlight: our recommendations concern only those areas in which the WRR believes the Dutch government should take more initiative. However, these recommendations may also apply to other governments. For each recommendation we suggest a number of concrete actions. We end by describing how these recommendations can be supported both institutionally and politically.

10.2 Transition 1: From Fiction to Facts

The task of demystification involves a transition from fiction to facts. This means that the current dominance of far-reaching preconceptions with utopian and dystopian outcomes must give way to a more rational understanding of the facts. In short, we need a more balanced picture of AI. The transition we are advocating here does not mean that government has to start telling society 'the truth' about AI. It does need to make learning about AI an integral part of its public function, however, and so evaluating AI and reflecting on its goals will likewise need to become central in that function. This also means that government will need to respond critically to parties with overoptimistic expectations, and likewise to those that only see risks. Our first recommendation, therefore, is to bring about this transition within government itself.

Recommendation 1
Make learning about AI and its potential applications an explicit goal of government's public function.

Two reflexes are typically observed when government uses a new technology. On the one hand there is 'technosolutionism'. A recent example of this in the Netherlands was the 'coronavirus tracing' app. Its introduction was announced early in the pandemic as an important part of the government's response to COVID-19, but the

stance adopted by the authorities automatically stifled discussion on its usefulness. No-one asked what the app actually contributed towards the response or whether – based on expert knowledge or the requirements of doctors and community health services – other, non-technical solutions would be preferable. Development of the tracing app was given high priority, but the stakeholders underestimated how long that would take. The outcome of a competition to build the app was that none of the entrants met the conditions set, and this was both a disappointment for the app's backers and a confirmation for those who had expressed misgivings.

The other side of the coin is a 'technophobic' reflex fuelled by failed or banned projects. Serious consequences have ensued recently from the inappropriate use of data by the Dutch government in two projects: the payment of childcare allowance by the tax authorities and the fight against fraud by local authorities using the System Risk Indicator (SyRI). The general public is now aware, albeit within a certain frame, that the government uses algorithms and that this potentially has negative consequences: privacy is undermined, and by extension fundamental rights are violated. As a result, the Dutch government has become more hesitant to use algorithms in general and AI in particular.[1]

Neither reflex is productive. Technosolutionism leads to sky-high expectations, whereby failure can lead to disappointment. Of late, however, heightened risk awareness in the Netherlands appears to be triggering the technophobic reflex. The result is that government is missing opportunities to improve existing practices. It is inevitable that mistakes will be made, but that does not mean that government should abandon the technology altogether. So how can the right balance be found?

As an emerging system technology, AI faces a lengthy process of use, practice and adaptation. It is not a simple tool or a magic wand that can be purchased and then left to perform its tricks. This is why learning must be an explicit goal of AI policy – more so than it is today. It also means that policy must take account of potential errors (without detriment to civic values). More explicit attention to learning will also allow executive agencies to experiment without immediately being held accountable in the political arena. For this to happen, such experiments also need to be recognized and supported by the industry at board level.

To build capacity in AI through learning, the WRR believes that government should first focus on attracting talent and training staff. AI will become a core component of organizations' primary processes. Technical and non-technical staff must be able to communicate at the same level to ask the right questions. Learning also implies organizing basic administrative tasks such as the timely and diligent archiving of information generated by AI processes, the transfer of knowledge to new staff members and access to databases and algorithms. On the latter point we must emphasize that the government must reach adequate contractual agreements with private IT suppliers regarding access to data, algorithms and other relevant AI

[1] Illustrative here is the ban on government use of discriminatory algorithms, which was adopted by a large majority in the Dutch House of Representatives in late May 2021. This ban also precludes the use of AI for positive discrimination.

information. This also requires government's own knowledge of AI to be up to scratch.

Such goals must be agreed explicitly before a large AI project is undertaken, not treated as incidental administrative burdens. So, this approach will also have implications for the way government works. Today it is common to allocate substantial budgets to large IT projects with fixed delivery dates. With AI, however, a more iterative process involving smaller projects is preferable. The required capabilities can be built through learning and evaluation, after which the projects can be scaled up.

Moreover, it is not only the government's executive agencies that will benefit from a learning approach based on progressive insights into AI. So too can political, legislative, supervisory and legal actors. How this might work in practice is illustrated by an example from the Dutch Council of State, the nation's supreme court: after having formulated a transparency assessment framework for the valuation of real estate it later issued a second ruling further clarifying the requirements, inspired in part by what had been learnt from the first.

Concrete Actions for Recommendation 1
- Work on building knowledge and capacity and on preventing dependency.
- Start with smaller ambitions and projects and then scale up.
- Explicitly allow room for mistakes and work with short evaluation cycles.

Wider society will also benefit from demystification and thus form a more realistic understanding of AI. This requires more than just knowledge, though; practical skills and an understanding of how to implement AI in different contexts are also needed. By analogy with the widely used term 'media literacy', in this context we refer to 'AI literacy'. Developing this skill is essential to enable society to adopt a realistic approach to AI and the changes it brings. Clarity about the facts of AI and its use is an important prerequisite here. Our second recommendation follows on from that.

Recommendation 2
Stimulate the development of 'AI literacy' amongst the general public, beginning with the establishment of algorithm registers.

Various actors have a role to play in the process of demystification. Journalists, academics and industry can all contribute towards the genesis of myths or help debunk them. Some degree of demystification will therefore occur automatically over time, without the need for state intervention. A basic level of AI literacy will eventually help citizens be more critical of overly optimistic, overly pessimistic or simply false representations, although these can probably never be eliminated altogether. That said, a number of ongoing developments could require government to play a greater role.

Much of the media coverage of AI is sensationalistic. There is plenty of speculation about systems that will supplant people and disrupt society. This creates a need for more facts about what AI systems actually can and cannot do. Many of the reports also feed fears of various kinds, as described in Chap. 5. Finally, AI is becoming more and more associated with applications for surveillance and control, which puts the technology in a poor light.

To encourage more realistic perceptions and a better understanding of AI, a first step for government is to be more transparent about its own use of the technology. It can do this by establishing algorithm registers (see Box 10.1). In the Netherlands, the City of Amsterdam has already started such a register to provide citizens with details of where and how it uses algorithms. Utrecht and Rotterdam have now copied this initiative. In its progress statement entitled 'AI and algorithms' of 10 June 2021, the national government announced that it was to investigate "how an algorithm register could contribute towards increasing transparency on the use of algorithms by the government".[2] Three months later, on 6 September 2021, the government submitted its Dutch Digitalization Strategy (I-Strategie Rijk) for 2021–2025 to parliament. This states that the creation of an algorithm register is one of the ambitions all ministerial chief information officers and their executive organizations intend to work on in the coming years.[3]

Box 10.1: Algorithm Registers

Calls for the creation of algorithm registers are increasing. On 19 January 2021 the Dutch parliament adopted a motion proposing the establishment of such a register to keep track of the algorithms the government uses, their objectives and the data they draw upon (TK 2020–2021, 33510: 16). The motion was prompted by the parliamentary inquiry into the childcare allowances scandal.

Some cities have already launched algorithm registers of their own. In the Netherlands they include Amsterdam, and elsewhere Helsinki, Finland. The Amsterdam register describes algorithms for automated parking enforcement, for processing public nuisance reports, for actions against illegal subletting and for crowd monitoring. In each case the register reveals what data was used to train the algorithm, how it is deployed, how officials use its output and how distortions (bias) and risks are dealt with.

Algorithm registers are also being considered further afield, as revealed in a report from the Law Society of England and Wales, the British professional body for lawyers. This organization advocates a register for algorithms in criminal law, in each case recording key details such as its transparency, standard operation and data use. The European Commission seems to be anticipating the introduction of an algorithm register, too: its proposed AI Act requires the registration of 'high-risk' AI systems, including those for private use.

[2] See *Kamerstukken II* 2020–21, 26643, no. 765, action 5.

[3] See *Kamerstukken II* 2020–21, 26643, no. 779, theme 6, priority 2.

Helping citizens understand the different types of AI applications being explored or used by government is a necessary next step. In our view, however, the creation of algorithm registers will only bring real added value if it also encourages conversation about AI use among both those already using the applications and those who will be affected by them in the future. The success of such registers will depend very much on the quality of the information provided, society's capacity to use this information and the response of the responsible actors to any problems identified. The registers should therefore be reviewed periodically.

In addition, we advise government to be particularly aware of its own role in shaping public perceptions of AI. The way government uses the technology – and advertises that use – is bound to play a part in defining how ordinary people view AI and their emotional response to it. The government should start by investing more in AI applications that change society for the better or help tackle the major challenges it faces (such as reducing climate change and combating social inequality). In this report we have discussed the many ways AI can be put to good use – for example, to create healthier air, reduce energy consumption, improve diagnostic procedures, enhance medical assistance and ensure better animal welfare. We call upon government to encourage and facilitate more such applications of AI and to advertise their merits.

The titles of these projects should also be given more consideration. The Dutch anti-fraud system SyRI was initially named 'Black Box', which may unintentionally have contributed towards the preconception that 'AI must be incomprehensible for humans'. Terms like 'killer robots' for autonomous weapon systems and 'robot judges' for AI in law also influence people's perceptions. Designations of this kind evoke strong associations. Sometimes that is the intention – to paint a clearer picture, for example, or to focus a discussion. Such expressive terms can distract from the issues that really matter, though, so government must not underestimate the power of the words it uses.

Through public information and educational campaigns, government can help build a basic knowledge of and familiarity with AI as well as making people aware of its possible pitfalls. But it is essential that these activities are not limited to the classroom and the workplace, as was the case with the Dutch 2019 AI Action Plan. That could suggest that AI is something we only need to worry about later, or that it will only affect people in certain occupations, when in fact its potentially wide-ranging application means that all of society will need at least a basic understanding of this technology. It is important first of all that interested citizens, those who want to know, be informed about the use of AI applications. The algorithm register and, even more crucially, the discussion on the use of AI advocated by the WRR will only deliver added value if the whole of society has a basic understanding of the technology. Providing realistic information and education to improve the public's understanding of AI can increase confidence in it (see also Box 10.2). Particular attention should be paid to finding effective ways to impart this basic know-how to citizens who are less self-reliant.

Box 10.2: Current Programmes to Encourage AI Literacy
There are already several initiatives for citizens of the Netherlands interested in AI. The National AI Course is a good example that should be encouraged, as is the Dutch version of the Elements of AI course launched by NL AIC and Delft University of Technology. People looking for information on the use of AI can find it on government websites such as the Ministry of the Interior's knowledge database (Kennisbank). To complement these sources, a website should be created with an overview of the information people need to be aware of if they want to use AI or are confronted by it. This could be similar to the existing sites for homebuyers, for example, or consumer watchdog sites.

More than providing technical knowledge, developing AI literacy involves building the public's knowledge so they can put news reports about AI and its applications into perspective and develop a realistic idea of its potential and limitations. This can be seen in the same light as 'media literacy', which involves the competencies needed to participate in a media-dominated society.

Concrete Actions for Recommendation 2
- Establish a government algorithm register for AI applications, initiate the conversation about the use of AI and ensure periodic evaluations.
- Critically evaluate government's own contribution in shaping public perceptions of AI.
- Give greater priority to AI applications that benefit society and draw attention to them.
- Contribute actively to public information and educational campaigns about AI.

10.3 Transition 2: From Abstraction to Application

The transition required for the task of contextualization involves the step from AI as an abstraction to its application. By 'abstraction' we here mean AI as a technology confined to the intellectual domain of research labs and academic reflection, remote from 'real-life' contexts. There is currently a lot of focus on the fundamental characteristics of AI systems and on related issues of transparency and explainability. Broadening this to include their practical application means paying more attention to the contexts in which the technology is used, in particular the technical requirements of the relevant ecosystem and the way users interact with AI.

Regarding this transition, first and foremost we examine the broader technical ecosystem that AI forms part of. We have formulated the following recommendation for the Dutch government.

Recommendation 3
Explicitly choose to develop a national AI identity, then investigate what adjustments this requires to the technical ecosystem in relevant domains.

Our ecosystem approach reveals that a lot of technology is needed for a well-functioning AI system. Permanent attention must be paid to talent development, research into algorithms, network quality, access to chips, building databases, developing cross-sectoral standards and building a secure ecosystem for sharing data and datasets. It is also important to keep tabs on emergent technologies that can give AI a boost. The government has already launched initiatives in several of these areas, including the Growth Fund and the Intergovernmental Data Strategy.[4]

The WRR recommends focusing on one specific additional point, namely the technical adaptations required to facilitate the AI environment. We use the term 'enveloping' to describe how an environment is modified to allow a technology to function effectively within it, analogous with the construction of the road network to facilitate the motor car or the power grid for electrical appliances. These adjustments often cannot be left to the market alone. Moreover, the choices made may have far-reaching consequences for society. This means that government must be actively involved. Just as the development of the car in the twentieth century required the creation of a mobility infrastructure tailored to motor vehicles, so enveloping for AI means developing an environment 'readable' for that technology. AI systems need be able to analyse their surroundings to interact with them intelligently (see also Box 10.3).

Box 10.3: Examples of Enveloping
Take autonomous vehicles. These days they have more and more intelligence built in, but are still far from being able to move completely independently in a complex environment. Adjustments to road surfaces and markings, or even the construction of specific infrastructure reserved solely for these vehicles (as with the motorway for the traditional car), are all big steps forward in the use of the technology. This does not necessarily mean adding new lanes to roads, but could instead take the form of special signs and signals or specific zones, such as industrial estates and other controlled environments, where experiments can be carried out safely.

The same applies to all manner of home automation systems, and also to complex industrial robots, which today are still ill-equipped to deal with the complexity and unpredictability of human behaviour. Experimenting with and investing in environments that are more easily readable for these technologies could make an important contribution towards their effective functioning and hence their usefulness.

[4] *Kamerstukken II* 2020–2021, 26643, no. 765.

It is impossible for the Dutch government to support every effort to make the domains affected by AI more readable for the technology. Of necessity, therefore, it must focus on a number of specific areas. The WRR thus advocates developing what we call the 'Dutch AI identity'. This encompasses those domains on which our nation wants to focus in the development and deployment of AI. Within these parameters we as a country cannot risk failing to implement the necessary changes, whether because of co-ordination problems or other reasons, which is why this transition cannot and should not be left to the market alone.

This national AI identity could include those domains in which the Netherlands is traditionally strong or ones that are important drivers of the Dutch economy, such as certain segments of agriculture, horticulture, infrastructure and logistics. Developing AI here will help prevent Dutch industry losing market share or becoming too dependent on foreign suppliers, while at the same time it should generate new revenue models. In addition, the Dutch AI identity could include domains that embody important civic values and where the government has a specific responsibility to take a lead, like healthcare or effective governance. The so-called AI Coalition is already compiling plans to stimulate AI innovation in various sectors of the Dutch economy. By formulating a national AI identity, the government could help steer this process. One example of where such guidance is needed is agriculture, in certain segments of which a limited number of suppliers currently dominate sales of models, analytical tools, algorithms and information services. Another is healthcare, where there are ambiguities about the ownership and control of some data.[5]

The government can also support the Dutch AI identity through a strategic procurement policy. As a major economic actor, it is in position to stimulate markets by building demand for certain products. PIANOo, the Dutch Public Procurement Expertise Centre, is currently developing an innovation-focused procurement policy. In 2019 the government launched SBIR (Small Business Innovation Research) to encourage businesses to develop innovative AI applications for the public sector. The government could make more intensive use of these instruments. Moreover, procurement policy is fragmented in many areas, from education to local government. Central government can strengthen the development of the AI ecosystem and focus more on areas of application important for the Netherlands by targeting the use of procurement instruments and co-ordinating the underlying requirements and standards.

Concrete Actions for Recommendation 3

- Define the domains and focal areas of a national AI identity.
- Identify the technical requirements and opportunities in each of these domains.
- Help shape the national AI identity by adapting procurement policy accordingly.

[5] Cf. TNO, 2021a, b.

The transition from abstraction to application is not just about the technological context of AI, but also its behavioural and user contexts. We therefore make the following recommendation for this social ecosystem.

Recommendation 4
Strengthen the skills and critical capabilities of individuals working with AI systems by developing a suitable training and certification framework.

The WRR believes that more attention should be paid to human-machine interaction. Even where technical systems work properly and comply with ethical guidelines, a lot can still go wrong in practice. For example, because users do not know how to manage these systems or fail to critically evaluate their functioning. An important factor to consider here is that AI transforms existing working practices, changing the role of the human user and possibly rendering traditional safeguards inadequate. We may demand that a human user must always be responsible for decisions ('in the loop' or 'on the loop'[6]), but we must also ask if this is a meaningful and realistic stipulation.

In an autonomous vehicle, for instance, given how long it takes a human to respond, the driver cannot be expected to intervene in time to prevent an accident. The same applies to humans who are required to oversee, interpret and manage increasingly complex analytical methods. People accustomed to algorithms functioning correctly are disinclined to question their results (automation bias), especially under pressure. While a person is still responsible in name, and so the human factor is still present, their putative role no longer corresponds with what is actually happening in practice.

One specific issue of human-machine interaction is how to approach the fallibility of both the human and the computer. If they come to different conclusions, it can be difficult to judge which is right. A human may be able to rectify an algorithm's error, but an algorithm can likewise discover patterns that a human being will not consider or expect. So how might we organize the use of AI so that it is possible for a human to correct the machine and vice versa?

The interpretation of AI outputs also involves human-machine interaction (see Box 10.4). For example, users need to understand the nature of the information generated by the system. Which in turn requires knowledge of the difference between correlation and causality, of margins of error and of whether a specific algorithm generates more false positives or false negatives. Users must thus be provided with information about the capabilities and limitations of the systems they work with.

[6] Having a human 'in the loop' means that an AI system can only function in response to a certain human action. With a human 'on the loop' the system can function independently, but the human can intervene.

Box 10.4: Augmented Intelligence
Algorithms already exist to advise the police to patrol certain neighbourhoods and help teachers when streaming their pupils. But they can make mistakes. The algorithm in the Crime Anticipation System (a predictive policing tool used by the Dutch police) deployed officers to public parks to combat car theft. Its reasoning was that this crime tends to occur where people gather, and people gather in parks. The problem, of course, is that cars are not allowed in parks and so anyone with a modicum of common sense would reject that advice. But there are other cases where an algorithm may well discover a pattern of crime that humans have not yet thought of.

Similarly, teachers should not simply ignore the results of streaming algorithms but nor should they trust them blindly (automation bias).

So, we need to create a context in which the teacher or police officer is supported in their work while at the same time the fallibility of both human and machine are considered. In other words, rather than replacing human intelligence AI should instead augment and enhance it ('augmented' or 'hybrid' intelligence).

Various actors have a contribution to make here, with the various AI labs in the Netherlands in a good position to play a key role. The government can also help by being actively encouraging (as well as actually participating in some cases, as it does with the Police Lab). In particular, it needs to pay more attention to the dynamics of human-machine interaction in its own use of AI. But it should also consider the behavioural context, to the requirements for using AI in its internal audit and supervision processes and to the application of guidelines.

To ensure effective human-machine interaction and strengthen the skills of the people working with AI, a system of training and accreditation for both humans and machines should be established. This could include certification, licences and specific requirements for certain applications of AI. The European draft AI Law, which distinguishes various levels of risk, provides a good starting point for the necessary requirements. Licensing procedures could be established by analogy with the system of licensing and approval used by health agencies to safeguard how new drugs are brought to market. This also makes patient information leaflets compulsory, so that those prescribed the drugs can read about their side effects and possible risks. Certification is used in a wide variety of situations, from sustainable food production to compliance with standards for the use of chemicals (under the European REACH regulation for the registration, evaluation, authorization and restriction of chemicals). Organizations that meet the standards receive are certified by the competent body.

Effective human-machine interaction requires a system of certification not only at the product or organization level, but also for individual users. In various fields people who use certain technologies or have certain responsibilities are required to

be certified. Obvious examples include electricians qualified to work on wiring in a building and the registration of professionals in healthcare, but all manner of other professionals also require certificates. Chartered accountants, for instance. In addition, many jobs (in the public as well as the private sector) require their holders to prove that they satisfy certain continuing education requirements. These are all forms of documentation that attest to a person's proficiency in their work.

The WRR is not proposing that everyone involved with AI should be trained and hold a certificate or licence. Everyone is affected by electricity, another system technology, but only technicians with special responsibilities need to be certified to work with it. AI will likewise affect almost everyone, but only those who work actively with the technology or are responsible for its deployment should need to demonstrate they have acquired the necessary knowledge and skills. We also wish to emphasize that this is not just about possessing sufficient technical know-how, but also the ability to determine whether the necessary safeguards are being observed.

Concrete Actions for Recommendation 4

- Pay explicit attention to the behavioural context and human-machine interaction in audits, supervision and the use of guidelines.
- In addition to certification, licences and risk levels aimed specifically at AI systems and organizations, develop measures to guarantee that the people responsible for the technology possess the requisite knowledge and skills –a proficiency certificate or AI licence, for example (see Box 10.5).

10.4 Transition 3: From Monologue to Dialogue

Engagement, our third overarching task, requires a transition from monologue to dialogue. The monologue here is the current situation in which discussion of AI is dominated by a relatively monodisciplinary group of technical specialists when in fact all manner of other actors and organizations should also be involved. The great distance between the developers of AI systems and the social environment in which those systems are applied also has the characteristics of a monologue. Citizens and civil society actors have their own expertise to contribute, but in addition an important role in providing feedback on how AI systems function in practice. In short, the conversation about the design and application of AI must be joined by a greater variety of actors. The Dutch government is already undertaking political initiatives to involve civil society in the development and application of AI-based applications. Illustrative of this is its declared intention to "encourage the business community and consumer organizations to jointly draw up a code of conduct for the use of

Box 10.5: AI Licences
More research is required to determine how AI licences might work, who would need them and whether they should be made compulsory. Here we offer a number of points to consider.

- Look at existing forms of proficiency certification, such as the register of medical professionals, pilot's licences and the certification of mechanics, and whether similar approaches might be appropriate in AI.
- Who exactly needs to obtain certification: the developer, the deploying company or institution or the individual end user? This will vary according to the context; AI in the form of a healthcare robot will require a different approach than a purely algorithmic application.
- The relevant training programme should include a theoretical component. Its primary focus, however, should be AI in practice. How should it be used? What do users need to be aware of? How are the safeguards monitored? Above all, trainees should be given plenty of opportunities to practise. What can you do with the technology? Just as a diver certainly requires theoretical knowledge in order to be able to plunge safely into the depths, but first and foremost plenty of practical training, so AI certification should entail quite a lot more than the existing courses provides – which is mainly general knowledge and basic theory.
- Practical knowledge of AI should also include a set of procedures that need to be carried out in complex situations or in the event of an emergency, much as medical standards exist for specific procedures in healthcare. Furthermore, users of these systems need to know when they can and may resolve issues themselves and when they need to seek the help of an expert.
- Given AI's enormous dynamism, it is advisable to require some form of continuing education for all holders of AI certification.

consumer data and algorithms to influence purchasing behaviour".[7] However, consumer organizations and other bodies representing citizens' rights and interests can only fulfil their role if they have the capacity to do so. So, to effectuate the transition from monologue to dialogue, our first recommendation to government with regard to engagement is as follows.

Recommendation 5
Strengthen the capacity of civil society organizations to expand their work into the digital domain in general and AI in particular.

[7] See *Kamerstukken II* 2020–2021, 26643, no. 765, action 6.

A number of parties in civil society already have a good grasp of the issues surrounding AI. This obviously applies to organizations engaged explicitly with the digital domain. In the Netherlands these include Waag, Bits of Freedom and Privacy First. These groups are increasingly managing to reach the general public and to put issues involving AI on the political agenda. Major human rights organizations like Amnesty International are now also paying close attention to the impact of this technology. Unfortunately, the same cannot be said of most organizations that focus on the interests of specific groups (employees, patients, teachers, people in poverty, disadvantaged and discriminated groups and so on).

Bodies like trade union federation FNV, patient advocacy group De Cliëntenraad, anti-discrimination think tank Artikel 1 and tenants' union Woonbond do important work for specific groups in Dutch society. AI offers new opportunities for these organizations, but it could also threaten – and even damage – their position and that of the people they represent. Examples of such threats are the spectres of a 'digital poorhouse' to the detriment of impoverished people, a 'New Jim Code' that disadvantages people of colour and 'digital open-air prisons' that restrict the freedoms of minorities. It is therefore important that organizations of this kind be empowered to understand and address these effects. Moreover, their specific knowledge is indispensable for the further integration of AI into society. But that knowledge is currently absent from many discussions around this theme, one major reason for that being that these bodies tend to know little about the technology.

The government is responsible for upholding a strong democracy and so needs to ensure that diverse voices are heard on important issues. When a new system technology is introduced, civil society usually lags behind big business and government in its adoption. Yet grassroots voices are crucial when it comes to reporting abuses of the technology and finding new ways to exploit it on behalf of a whole variety of interest groups. The algorithm register mentioned earlier can mitigate this deficiency by making knowledge about AI use publicly available. In addition, it is important that the government actively approach and consult interest groups as part of its AI policy.

The government can also contribute to a more prominent role for civil society by providing grants and facilitating training programmes or partnerships. Nor should the formal and institutional mechanisms that engage particular interest groups in the democratic process be overlooked. In particular, we are referring here to the need to involve works councils and other codetermination bodies in AI-related decisions. Whilst Dutch law stipulates that employers need the consent of their works councils before processing employees' personal data, the specific workplace implications of AI – in the form of staff monitoring systems, for instance – have not yet been adequately addressed by those councils.

Concrete Actions for Recommendation 5
- Include AI literacy in funding policy and training programmes.
- Encourage co-operation between interest groups and suchlike organizations in the digital domain.
- Inform civil society stakeholders of the various ways they can engage with decision-making around the use of AI, such as through co-determination forums.
- Involve interest groups in political decision-making about AI policy and regulations structurally and from an early stage.

The second point in the transition from monologue to dialogue centres on the feedback loop between AI in practice and AI on the drawing board. A lot of attention is paid to the quality and reliability of data used in AI systems and to their analytical methods, their functioning and their transparency – that is, their input and processes – but much less to their outputs. In other words, whether AI does what it is supposed to and does it satisfactorily.[8] Integrating outcomes into the process by creating feedback loops to developers and other stakeholders would seem to be a logical requirement for AI systems, yet it is not an activity sufficiently rooted in practice. Consequently, our second recommendation in respect of engagement is as follows.

Recommendation 6
Make sure that effective feedback loops exist between AI's developers, its users and the stakeholders who experience it in practice.

There are various reasons why feedback loops receive relatively little attention. One is that real-life experiments are regularly conducted without the explicit consent of those involved. After the experimental phase, systems are implemented without first undergoing an evaluation of their effectiveness. Of particular relevance here is the fact that AI systems often draw on data about generic groups rather than bespoke information. As a result, the effectiveness of important legal safeguards, such as consent to use data and compensation in the event of malpractice, is significantly reduced.[9] Another problem is that such applications can engender discrimination against certain groups and yet leave them with few opportunities to defend themselves. Also, once a system's functionalities and operating instructions of a system have been agreed upon, changes may be required in response to the self-learning process (and the corresponding feedback) that necessitate a

[8] Cf. *Kamerstukken II* 2019–2020, 26643, no. 641, in particular section 3.2 concerning quality assurance.

[9] Kosta, 2020.

reassessment of the entire system. This is especially typical of the government. Such long and complex processes hamper the working of feedback loops.

In addition, as AI transitions from the lab to society the requirements for system feedback change. Systems are often extensively tested in the lab using carefully compiled sets of test data. When these systems are used in practice, in many cases the monitoring and feedback process is much less thorough than in that controlled research environment.

Another reason for a lack of feedback may be that the requisite information is difficult to obtain. For example, an employee recruitment algorithm will not integrate feedback on candidates who have been unjustly rejected as there is no data on how they would have performed had they actually been given the job. Algorithms that provide pupil streaming recommendations require feedback data that will only become available many years later, and even then, the results may be ambiguous (because eventually the student did not follow the recommended trajectory, for instance). If a student achieves better educational outcomes than the algorithm predicted, was it incorrect or did the student 'up their game' later in their schooling?

Finally, the commercial interests of developers or contractual agreements between them and user organizations may stand in the way of an effective feedback loop. To facilitate feedback while at the same time ensuring confidentiality, a limited number of persons within the organization could be authorized to monitor the factors relevant for the loop.

Effective feedback is crucial for the proper functioning of AI systems and the protection of civic values. The childcare allowances scandal is a tragic example of what can happen when there is not enough feedback and critical reflection on a system's output. As the implications of using algorithms for citizens and their legal position increases, it is crucial that feedback about those implications be processed actively. That feedback loop will need to be twofold. First there is a loop between the developer and the user (a GP, a police officer or a teacher, for instance). Barriers all too often exist between these two actors. But a second loop is also needed, taking in everyone affected by the system (the GP's patients, suspects arrested by the police, a teacher's pupils and so on). Both users and those affected are in a position to recognize errors, contribute expertise and suggest improvements. So rather than a one-way monologue, a dialogue is needed.

The government must therefore pay more attention to the way these feedback loops are organized and their scope, particularly in the public sector – including local government, executive agencies and especially those domains where decisions have a major impact on citizens. In the WRR's opinion, the development of a standard for feedback is a prerequisite here.

Concrete Actions for Recommendation 6

- Identify developers, users and citizens affected by AI systems in different domains and develop effective feedback mechanisms.
- Make feedback mandatory in government AI applications.
- Organize feedback in areas involving sensitive information in an indirect manner.

10.5 Transition 4: From Reaction to Action

An effective approach to regulation requires a transition from reaction to action. By 'reaction' we here mean a primarily passive, wait-and-see attitude to legislation, with new laws only introduced in face of acute, often specific issues. The risk here is that legislators both lose sight of the broader effects of AI on society and fail to consider its individual aspects as part of a bigger whole. Issues such as reliability, explainability and transparency are definitely important, but the decisive one is how to integrate AI in society. For the transition to an action-based approach, our recommendation for the short term is that the legislature assume a more active role, address relevant developments from a more integrated perspective and develop legislation relevant to an economic and social context in which AI is maturing. In addition to regulating the operation of the technology itself, lawmakers also need to focus on the other dynamics and economic forces associated with AI, such as the growing concentration of power in the hands of a limited number of (mostly private) parties and the consequences this has for AI's place in society. This transition thus requires that government play a more directive role in organizing the 'digital living environment'. So our first recommendation for the transition from reaction to action is as follows.

Recommendation 7
Link the regulation of AI to a discussion about the organization of the digital living environment and set a broad legislative agenda.

As we have seen in Chap. 8, various regulatory processes have been set in motion in recent years. These include both national and international initiatives, from European legislation on AI and data use and discussions around facial recognition and autonomous weapons to the regulations and guidelines drawn up by Dutch ministries. The European proposal for an AI Act, to which the Netherlands will eventually be bound, amounts to a concrete proposal for an AI system based on various risk categories. These regulatory processes mostly concern acute and relatively clearly defined issues such as the use of algorithms to combat fraud, bias and discrimination, as well as issues surrounding transparency and unreliable outcomes. Here the debate on regulating AI focuses mainly on the relevance of existing frameworks and whether new regulatory and supervisory institutions will be needed. One question that is not addressed sufficiently is what civic values we want to actively protect or develop, and what steps this will require as we integrate AI into society.

As AI becomes more embedded in our society, second and third-order issues will arise that require new rules for their management. A system technology always gives rise to questions about the effects of its concrete application, and even more so about the associated economic dynamics and their wider effects on society. Electricity and the advent of the motor car, too, forced legislators to consider developments from the perspective of their broader effects on society, in this case the

physical environment. Power cables had to be laid above or below ground, and a road network constructed that took account of the natural environment. Embedding AI will involve similar choices concerning the design of the digital living environment, a phenomenon that already encompasses many aspects of society. The WRR believes strongly that the government regulation of AI should involve more than just the technology itself and its applications (reliability, safety, transparency and so so), but also encompass the wider digital living environment.

The development of earlier system technologies teaches us that the role of government will grow as AI becomes more integrated into society. It would be prudent to take on this greater role sooner rather than later. The European proposal for an AI Act regulates the authorization of AI applications in the member states but, because it focuses on managing risks, leaves many matters unaddressed. The WRR recognizes that the potentially ubiquitous nature of AI will make it difficult to anticipate what frameworks will be threatened or otherwise require modification. In many cases this will only become clear over the course of time. But the legislature cannot afford to sit back and wait – the public and other interests at stake are too great. Lawmakers need to stay abreast of the latest developments to be able to respond in good time. To this end the government must not only invest in research and in monitoring those developments (as official regulatory bodies currently do), it should also dare to take concrete steps. The new and therefore somewhat uncertain nature of AI should not be overestimated. Already obvious ambiguities and tensions related to the existing legal frameworks can be rectified or eliminated fairly easily, which will benefit the ongoing process of embedding AI in society. Uncertainty about the applicability of the existing frameworks, after all, as well as points of legal contention such as what data may be used, currently pose obstacles to the technology's broader application. It is better to make these choices now, because otherwise we as a society could be faced with a fait accompli.

In the WRR's opinion, the most urgent task for government is to take the initiative and plan for the long-term development of AI and the management of its broader social effects. This involves issues such as the goals we want to pursue as a society and the question of where, for what purpose and under what conditions we want to use AI – including restrictions or even bans in certain domains (as also proposed in the European proposal for an AI Act). The opportunities society can derive from AI deserve particular attention here. Like electricity, AI is not only an economic good but can also benefit large groups in society or even the entire population. The introduction of electricity made the days longer, homes safer, cities cleaner and life more enjoyable in many ways. Similar advantages may be expected from AI. The challenge for government is to ensure that the technology is deployed where it can contribute the most, on a scale and for purposes congruent with the needs of Dutch society.

This discussion requires thorough consideration of sometimes conflicting civic values, a task that cannot and must not be left exclusively to technical experts and tech firms. Perhaps even more important for government than asking whether the existing frameworks are adequate for the challenges ahead or whether AI in fact requires new rules is the task of forming a clear picture of the organizational issues

Box 10.6: Legislating for the Information Superhighway
The 1998 policy paper on "legislation for the information superhighway" (*Nota Wetgeving voor de elektronische snelweg*) presented the then Dutch government's perspective on regulation of the internet. It was based on an extensive study of the internet's impact on the Dutch legislative environment. As well as exploratory technical, governance and legal surveys and a comparative international legal review of the internet, this also included a discussion of strategic themes such as internationalization and jurisdiction, reliability, markets and law enforcement.

The policy paper provided a framework of reference to give all actors in the process a better understanding of pertinent questions related to internet legislation, contained a series of proposals for new and amended statutes (as well as measures to repeal) and suggested possible Dutch input for international forums. To guide the implementation of these proposals, it also presented a prioritized plan of action.

involved in embedding this system technology in our society, including the role to be played by official regulation.

The WRR believes that a more strategic approach to AI should echo that used until recently for national land-use planning in the Netherlands. This is based on comprehensive long-term policy papers. In the case of AI, the government can also turn to its 1998 policy document on "legislation for the information superhighway" (*Nota Wetgeving voor de elektronische snelweg*), which set out a strategic vision for the internet by formulating a series of policy challenges and goals, a corresponding governance philosophy, a toolkit and an implementation plan (see Box 10.6).

Concrete Actions for Recommendation 7
- Accept that preparing legislation aimed at integrating AI into society will be a long and sometimes uncertain process. Adapt legislative instruments accordingly, but do not wait too long before acting.
- Draw up a broad and integrated legislative agenda for AI and the organization of the digital living environment, including specified policy goals, a corresponding governance philosophy, a toolkit and an implementation plan.
- Include in this agenda a list of legal provisions to explicitly regulate the implications of AI in the short term (covering, for example, automated decision-making, liability, archiving and the legal status of autonomous systems).
- Strengthen the monitoring role of relevant official regulatory bodies and create a feedback loop with policy and legislation. If necessary, process the results – along with those generated by other actors – in a separate monitor.

Our second recommendation for the transition from reaction to action concerns government's specific focus on regulating AI as a systemic phenomenon.

Recommendation 8
Use legislation to actively steer developments related to surveillance and data collection, the skewed relationship between public and private interests in the digital domain and concentration of power.

Treating the regulation of AI as a systemic issue – and hence an issue of AI's integration into society –reveals how the digital living environment needs to be organized accordingly. If government does not actively manage how and by whom AI is used in society, there is a risk that it will eventually be unable to control its development. This requires action in at least three areas.

First, it is important to reduce the public sector's dependence on private companies. While AI is finding increasing use in the private sector, government is less eager to adopt it due to unfamiliarity with the technology and growing concerns about its use. Some examples illustrate this gap. The police are required to adhere to strict rules when enforcing the law, but what if services or applications become available that allow individual citizens to use facial recognition software to identify criminals? Or take public space, where government is primarily responsible for overseeing the behaviour of individuals and businesses. But with the proliferation of other parties collecting information by means of cameras, drones and sensors, they potentially have access to more information about public space than the authorities themselves. As a result, government could lose some control of areas that fall under its responsibility, an issue that could be augmented by a brain drain of the requisite policymaking knowledge as more third parties use AI. In addition, this could lead to sensitive matters being outsourced – with the consequence that dubious practices are hidden from government view.

Secondly, the growth of mass surveillance, and with it the largely unfocused collection, use and reuse of data, needs to be brought to a halt. Here too, of course, the relationship between the social costs of surveillance and data use and their benefits could be examined on a case-by-case basis (as currently), and various safeguards could be put in place for individual applications – varying from facial recognition and influencing online behaviour to smart applications in homes.[10] But there is also a more structural component of this development: tracking people – including their behaviour and even emotions or unique DNA characteristics – has become an important part of the business model of numerous companies, including online platforms.[11] The internet economy is increasingly underpinned by various forms of

[10] De Conca, 2021.

[11] Rathenau Instituut, 2021a, b.

surveillance. AI can be seen as the next phase in this development, since it enables companies to track individuals, attach profiles to them and respond to their preferences. Also relevant is the strong increase in and distribution of digital devices that facilitate tracking, which is the reason why major technology companies are entering the market for smart consumer electronics or forming alliances with the manufacturing industry. Surveillance activities – including those by government itself – have a major impact on the use and perceptions of AI and raise questions about how companies utilize data and how the relationship between governments and citizens is affected. AI can never acquire a legitimate place in society if we cannot find a better way to protect civic values such as privacy, individual autonomy, security and democratic control.

Finally, another issue for the further development of AI is the far-reaching concentration of power within a limited number of technology companies – in particular, a small group of American 'tech giants' including Google, Facebook, Amazon, Microsoft and Apple. All of which also happen to be some of the biggest players in AI. Their power has only increased as a result of the COVID-19 pandemic and the growth of working from home and video conferencing. For example, a very small number of providers completely dominate the supply of certain crucial components to the Dutch higher education sector.[12] There is increasing worldwide resistance to the power these companies wield from their bases in Silicon Valley, and governments are now starting to act. The European Commission, the US Department of Justice and the UK government are amongst those to have described these firms as a threat to innovation, competition and privacy. In addition, the way they filter and disseminate information is increasingly seen as a serious political threat, not just to vulnerable democratic governments but even to established democracies such as the Netherlands.

The major technology companies have the capacity and resources to determine the direction in which AI is developed and used. Moreover, network effects allow them to play an important role in other sectors too. Activities driven not by democratic values but solely by commercial interests. Their position and power are particularly problematic when the services they provide become part of the social infrastructure. How the power of the big technology companies will be restricted is remains unclear, but the history of system technologies teaches us that monopolies are typically either broken up or forced to open up their infrastructures to others. Various proposals to this end are currently in circulation.[13] The most concrete to date have been tabled by the European Commission, which has drawn the contours of a coherent European internet law with its draft Digital Markets Act and Digital Services Act.[14] The WRR advises the Dutch government to contribute actively to these proposals and to provide input where necessary. The Netherlands can also take

[12] VSNU, 16 April 2021.

[13] CPB, 2021, p 16.

[14] Chavannes et al., 2021.

its own independent steps in this regard by adapting competition legislation and the regulation of data power. More effective use of public procurement policy (already mentioned as one of the concrete actions arising out of recommendation 3) could also be a means to encourage a greater diversity of suppliers of products and services.

In addition to these proposals aimed at limiting the power of the industry and ensuring a well-functioning market, there are also initiatives aimed at reducing dependence on private suppliers and developing alternatives with public funds. One example is the EU initiative described in Chaps. 5 and 9 to develop its own cloud services and AI centres (including in the Netherlands), as well as the AI4EU platform. There are also more far-reaching projects on a smaller scale, such as the establishment of digital utilities for electronic identification. A utility of this kind could also be considered for AI – for example, as part of the national AI identity mentioned earlier and its supporting technical infrastructure. An important facet of such initiatives is that their development can be rooted in civic values. This is particularly relevant for public sectors such as healthcare and education.

The WRR's primary concern regarding the transition from reaction to action is that government must realize that regulating AI alone will not be enough; it also needs to act in many other areas to ensure that the use of AI at least upholds, and preferably reinforces, a whole raft of civic values. If it remains insufficiently aware of this and fails to take up its broader task in good time, there is a risk that other interests and parties will take the lead in embedding AI in our society. It is unrealistic to think that that path can still be changed after the 'moment of closure' discussed in Chap. 8.

Concrete Actions for Recommendation 8

- Guarantee and secure government control over core digital facilities, if necessary, building them in-house, in critical domains for the Dutch AI identity and public sectors including healthcare and education.
- Review legislative policy on surveillance in light of the fact that AI is the next stage in the development of surveillance technology.
- Deploy available procurement instruments on a much larger scale to safeguard civic values. Ensure that such instruments do not favour the major technology companies.
- Actively contribute to European legislation and related initiatives for the regulation of AI and the wider digital environment.
- Accelerate the process of amending competition law, in particular where it affects the data economy and AI companies.

10.6 Transition 5: From Nation to Network

Finally, the task of positioning requires a transition from 'nation' to 'network'. What this amounts to is that we must not consider AI merely as a zero-sum competition with other countries but also need to work on building stronger ties with partner nations. This applies in particular to the member states of the EU. The transition here also involves considering national security not just as response to external threats, since it also encompasses the technologies citizens use in their daily lives. To fully understand the security threats we face, we need to shift our attention to the international network we form part of. The WRR proposes that rather clinging on to the idea that the Netherlands is in competition with other countries to build prosperity and power (as a nation), we should focus more on ties with other countries (as part of a network). As regards the economic component of this task, our recommendation is as follows.

Recommendation 9
Strengthen the competitiveness of the Netherlands through 'AI diplomacy' that focuses on international co-operation, in particular within the EU.

Governments and businesses worldwide are investing heavily in AI to strengthen their competitiveness. There is far-reaching international competition in all aspects of AI, not only in the form of large-scale public and private investment but also in the development and retention of talent. The Netherlands cannot afford to fall behind here, because many neighbouring countries are already making substantial investments in these activities.

However, the WRR does advise that, rather than simply 'doing enough to stay in the race', the Netherlands adopt a somewhat different role and position. More attention should be paid to strengthening competitiveness through international co-operation, by conducting 'AI diplomacy' instead of focusing on competition.

A first focal area here could be fundamental research. The European CLAIRE network has chosen to establish its head office in The Hague.[15] Strengthening partnerships like this could generate positive spin-offs for Dutch business. A good analogy is CERN in Switzerland, where Europe has become a leader in particle physics by pooling its research resources. It is worthwhile taking note of the conditions under which such research collaborations achieve success.[16]

Countries can also co-operate in the development of concrete AI applications. For example, France and Germany have initiated a European data and cloud service called Gaia-X.[17] The Netherlands joined later, and there is now also a Dutch hub

[15] See Box 9.1 in Chap. 9.

[16] See, for example, Smith, 1999.

[17] See Box 9.2 in Chap. 9.

representing our national interests at the European level. Critics may warn that such projects are unfeasible, but in fact Europe has a history of successful technological partnerships including Galileo (Europe's alternative to GPS) and the aircraft company Airbus. Here again, it would be wise to learn from past successes and failures.[18] Such partnerships clearly have the potential to strengthen the European position. Failure to participate would represent a lost opportunity to uphold Dutch interests at this level.

Collaboration to strengthen competitiveness could also take the form of more co-ordination between existing companies. The growing interdependence of economic and geopolitical objectives has led to trade disputes involving various digital technologies. Dutch firms including ASML and NXP, which supply important hardware for AI applications, already find themselves subject to the vagaries of US-Chinese trade relations. Similar situations may arise in the future and affect Dutch technology companies like Philips, KPN, TomTom or Adyen, and other European businesses such as Siemens, SAP, Ericsson, Nokia or Dassault. In the light of this contest between the global superpowers, European countries would do well to work together to strengthen their joint international position and so also improve their competitiveness as individual nations – and that of their own companies. Specifically, we should consider policies to protect key business from takeover bids (hostile or friendly) and unwarranted fines or sanctions imposed by trading partners.

Another way in which co-operation can strengthen Dutch competitiveness is through legislation and regulation. The EU is already active here when it comes to personal data (the GDPR) and the draft AI Law of April 2021.[19] In addition, the process of standardization is crucial. This technical domain has so far received relatively little attention in the AI debate, but is absolutely instrumental in strengthening countries' competitiveness.[20] Furthermore, as we have explained in Chap. 9 standardization is increasingly subject to geopolitical forces. China in particular is trying to have its own standards for AI accepted as the norm in international forums. The EU (including the Netherlands) needs to be very alert to this development and seek co-operation with other countries that subscribe to the same values.

While the EU is the appropriate forum for most areas of co-operation, in specific cases like-minded and pioneering third nations such as Canada, France, South Korea or Singapore could be approached as well. When it comes to issues of digitalization, we must be open to broad coalitions involving many countries.[21]

[18] Domini & Chicot, 2018.

[19] Anu Bradford's (2020) study on 'the Brussels Effect' reveals how EU legislation has set the tone on the global stage in various areas. As a regulatory power, the EU can influence the direction of the market.

[20] Veale & Zuiderveen Borgesius, 2021.

[21] See also WRR, 2015.

Concrete Actions for Recommendation 9
- Identify suitable domains and forums co-operation in AI.
- Explore opportunities to strengthening the Netherlands' position in each of these domains and forums as part of the Dutch 'AI identity'.
- Involve national and international actors such as standardization bodies and prominent academics in the policymaking process.
- Formulate specific goals for each domain, but also synergies across them – for example, between fundamental research and European projects for AI applications.
- Be alert to regulatory proposals submitted by other countries that could harm Dutch interests (AI diplomacy).

In short, the Netherlands can strengthen its competitiveness by co-operating internationally in the areas of fundamental research, establishing new services and co-ordinating industry legislation and regulations. The WRR therefore recommends developing an integrated AI diplomacy strategy to facilitate well-considered choices in these domains (including choices for the long term).

The transition from nation to network has a security dimension as well as an economic one. Our recommendation in this respect is as follows.

Recommendation 10
Develop the knowledge required to safeguard the defence of the Netherlands in the AI age. To this end strengthen the nation's capacity to defend itself in the 'information war' and against the export of 'digital dictatorship'.

The issue of AI's impact on security often focuses on autonomous weapons. These systems can indeed have far-reaching consequences for security and so the current efforts to control their use are certainly welcome. But AI influences the military domain in other ways as well, such as improving decision-making processes or enabling the analysis of more data. More and more attention is now being paid these aspects, not least within NATO. The WRR wishes to emphasize the importance of a broader perspective here. AI affects security not only in the military sense but also in civil society.

The far-reaching digitalization of society and the economy is making our country more vulnerable to non-military attack. Social media platforms, sensors in the infrastructure, operating systems, communication systems and various other 'networked' domains are all potential targets. Cybersecurity is a fast-growing policy domain. In a recent report the WRR argued that more urgent preparations are needed for the phenomenon of 'digital disruption'. In addition to the infrastructure and networks themselves, greater attention should be paid to the information that flows through

them.[22] Its influence and manipulation fall under what is termed 'information warfare'. In part this is being fought manually, but increasingly also by means of algorithms.

The WRR points out the need for an integrated approach to this risk. It was long assumed that digital technologies have an inherently democratizing effect. Although they can certainly help foster democracy, various authoritarian regimes have also proven very capable of using them for undemocratic ends. They deploy digitalization, and AI in particular, to strengthen their regimes – for example, by encouraging widespread, centralized and cheap surveillance. Moreover, countries such as China and Russia are increasingly exporting such technologies and so encouraging other states to move further down the road of authoritarianism. But the risks could ultimately affect the Netherlands as well. By using digitalization and AI as instruments of national security, such authoritarian countries have built up strong digital capabilities. The WRR believes that the Netherlands needs to be more aware of this. Moreover, the discussion should go further than only the rollout of 5G and the dangers of doing business with companies like Huawei. There are plenty of other risks, too, such as the import of technology like cameras with facial recognition, smart city technology for monitoring public spaces and new telecom hardware and software for public services. Another is the export of Dutch technologies to countries with authoritarian goals. Finally, campaigns to spread fake news, deepfakes and conspiracy theories in our country are also threats (see Box 10.7).

Several initiatives within the EU are addressing growing concerns about 'digital sovereignty'. In early 2021 the Dutch Cyber Security Council – an advisory body comprising representatives of the business community, the government and cybersecurity experts – explicitly called for a far more active stance by the national government to maintain its control over democracy, the rule of law and the economic innovation system.[23] The WRR agrees with the council's recommendation and advocates that the Netherlands work towards the development of a joint European strategy in this field. The Dutch initiative to collaborate with France and Germany

Box 10.7: AI as a Weapon of Information Warfare

'Microtargeting', 'sentiment analysis' and 'natural language processing' are all examples of techniques that are increasingly being used and threaten our national security. Deepfakes (faked video and audio recordings that are ever harder to distinguish from the real thing) are becoming more and more common. These activities entail risks for individual citizens and for society as a whole, because they encourage distrust, uncertainty and chaos. Such technologies could ultimately even pose a risk to democracy itself.

[22] WRR, 2019.

[23] Cyber Security Raad 2021.

on the establishment of an EU-wide regulatory body and gatekeeper empowered to monitor all mergers and takeovers by major digital platforms is a good first step in this direction.[24]

It is also important for the Netherlands itself to gain a better understanding of how foreign powers deploy information for their own purposes and how this can threaten our democratic system. We then need to strengthen our national capabilities – including in AI – to counteract that threat. It is not obvious how the information war can be won. What is clear, though, is that we have no time to lose: we must build the requisite expertise and make the necessary policy choices as soon as possible. A good first step in the short term is to focus more on threats of this kind in the annual Cyber Security Assessment compiled by the National Co-ordinator for Terrorism and Security.

Concrete Actions for Recommendation 10

- Identify how different forms of AI, such as microtargeting and deepfakes, are being deployed in the global information war.
- Prevent the import of technologies of digital dictatorship to the Netherlands and the export of Dutch technologies to countries where they will be used for dictatorial purposes.
- Further strengthen the digital sovereignty of the Netherlands as part of EU-wide efforts to this end.
- Systematically include information security risks in the annual Cyber Security Assessment

10.7 From Instruments to a Policy Infrastructure

The above recommendations concern the work that needs to be done to embed AI in society. Our final recommendation is about the way this work can be supported and focuses on the institutional aspects of government policy on AI.

As mentioned earlier, the history of system technologies teaches us that the role of government in AI will gradually increase in various ways. Railways were originally developed by private companies in the United Kingdom and the United States, but over time government took a more active role. First through regulatory legislation, and eventually in many European countries by becoming a public transport operator itself. The same occurred with electricity, where governments built the networks. So, while the nature of government's role varies, the extent of its involvement clearly increases. Each time this has happened, a policy infrastructure emerged to co-ordinate the new tasks and discharge the corresponding responsibilities. In the Netherlands, for instance, national public works agency Rijkswaterstaat was entrusted with managing the country's motorways and various new public bodies

[24] Rijksoverheid, 27 May 2021.

were created to oversee road use: the Netherlands Vehicle Authority to issue car registrations and driving licences, the Human Environment and Transport Inspectorate for the safety of taxis, buses and other forms of transport, the Central Office for Motor Vehicle Driver Testing and so on.

We expect a similar pattern to emerge for AI. This means that integrating it into society will require government to do more than only develop new instruments. In the coming years it will also have to build a policy infrastructure. For this transition, our final recommendation is as follows.

Final Recommendation
Build a policy infrastructure for AI, starting with a co-ordination centre that is anchored politically in a ministerial subcommittee.

The need for a policy infrastructure is becoming increasingly clearer. Like previous system technologies, AI will influence a variety of both sector-specific and generic civic values. In time both the risks and opportunities for those values will come into sharper focus. AI will also increasingly necessitate a debate about the goals we want to pursue as a society and the question of where, for what purpose and under what conditions we want to use this technology. Furthermore, it will require international co-operation, particularly within the EU. So the government will become increasingly involved in its development. In addition, the WRR notes ever greater recognition of AI's strategic importance – a factor that also calls for an active government role. These developments reveal the need for wide-ranging and generally available resources to support the underlying process of policymaking and legislation.

The discussion on a policy infrastructure for AI is in fact already underway in the Netherlands. For example, a Ministry for Digitalization[25] has been proposed that would include AI. There are also calls to establish a supervisory body for algorithms. Various countries have already passed the stage of conceptualization and have launched concrete initiatives to embed AI institutionally (see Box 10.8). The WRR advises the Netherlands to follow suit.

The governments of various countries are taking steps to develop a policy infrastructure for AI, but there is no one blueprint for this. Some differences have to do with the missions of the relevant bodies, their composition and competence, and above all how they are anchored in the government organization. Prior to the advent of AI, some countries had already established an agency (Denmark) or appointed a minister (Norway, Sweden, Germany, Italy) or undersecretary (France, Belgium) for digitalization or digital government – a responsibility that now also includes

[25] ROB, 2021.

Box 10.8: Countries Already Embedding AI Institutionally

Several countries, amongst them Belgium, the UK, France and Germany, have set up committees bringing together experts from academia, the industry and government to develop their national AI strategies. The argument posited for this broad composition is that AI will eventually affect all sectors of society, and hence also all ministerial remits.

Whereas these bodies were often initially temporary and external to government, some are now permanent organizations in the form of advisory committees (in Austria and Singapore, for instance), government task forces (Kenya, India and others) or initiatives entrusted with responsibility for AI (such as the National Robotics Initiative in the US, which is supported by various government organizations). The United Arab Emirates is the only country to have a Ministry for AI, as the country wants to be at the global forefront of AI in sectors like transport, healthcare, renewable energy and transport, and even build houses on Mars before 2117.

The UK has set up an Office for Artificial Intelligence to implement its own AI mission and the associated 'Data Grand Challenge'. This body falls under the Department for Digital, Culture, Media and Sport and the Department for Business, Energy and Industrial Strategy. Important achievements so far include its *Guidelines for AI Procurement* and the *Guidance on Building and Using Artificial Intelligence in the Public Sector*. The Office for Artificial Intelligence and the UK government in general are assisted in the development of AI policy by an independent Council for AI, whose members include AI experts and representatives from the industry, the public sector and academia. This body is also working to broaden public knowledge of AI.

AI. In any case, it is clear that the Netherlands should look to other countries for inspiration in creating a Dutch AI policy infrastructure (see Box 10.9).

But there is another development that puts the need for a policy infrastructure on the agenda: the EU's draft AI Law requires member states to designate one or more national competent authorities to supervise the application and implementation of AI and to designate a single national supervisory authority as the official contact point for the public and other actors. This authority will also represent the relevant member state on the European Artificial Intelligence Board, the body that will implement the law.

In short, come what may the Netherlands is going to have to develop a policy infrastructure to meet the EU requirements. As for the next step, the WRR considers it premature at this stage to advocate a separate ministry or a specific regulatory body for AI. Both options may prove valuable at a later stage, but at present it remains insufficiently clear what their added value might be. Especially as there is a real possibility of overlap with existing actors. It also takes a lot of time, resources

Box 10.9: Starting Points for a Dutch AI Policy Infrastructure

The Netherlands already has various forums that can be considered part of an AI policy infrastructure. For example, several ministries now have departments, directorates or separate units for digitalization, such as the Digital Government Directorate and the Digital Economy Directorate.

The Ministry of Economic Affairs and Climate Policy, the Ministry of the Interior and Kingdom Relations and the Ministry of Justice and Security recently formed a partnership for digitalization. These three departments were jointly responsible for the first Dutch Digitalization Strategy in 2018 and the updated versions of 2020 and 2021. Since the new national government took office in 2022, there is now also an undersecretary for Digitalisation at the Ministry of the Interior.

Meanwhile, an interdepartmental working group has been active developing the government's perspective concerning the impact of digitalization on civic values, human rights and the SAPAI (Strategic Action Plan for Artificial Intelligence). Another such group, for AI specifically, has been formed to bring together government inspectorates and market regulators.

The Netherlands also has a recently overhauled committee of chief information officers, focusing amongst other things on 'digital transformation and technology-driven innovation' throughout government. This has developed a generic action plan for information management and installed a government commissioner for that domain. Finally, a Permanent Committee for Digital Affairs was established in the House of Representatives in 2021.

and energy to establish such complex official bodies. Moreover, centralization can create unrealistic expectations with regard to designated tasks and responsibilities. For a central 'algorithm authority' to be able to decide what is and is not permissible, for instance, it would require a thorough knowledge of rules, practices and standards in numerous fields, ranging from healthcare to mobility and defence. A near impossible challenge. Given that AI it still in its early stages as a system technology, it remains unclear which issues will demand a more general, overarching approach from government.

However, none of this means that the WRR is content with the current status quo in policy surrounding AI. Many actors within government are currently faced with AI-related issues but have limited knowledge of how to deal with them. Although some do co-operate in their search for answers, there is no structurally co-ordinated approach.

In recent years a number of audits have been conducted of AI applications currently used in and outside government (both central and at other levels), and several exploratory and advisory studies have considered their relationships with civic values. Many of these exercises, however, only highlight the fact that we are dealing here with a fragmented landscape of participating and responsible bodies. A system technology such as AI requires that the process of knowledge development be

permanent and clearly structured, and that the information generated be widely shared and discussed. This is necessary to gain a proper insight into the way in which AI is being integrated into society and what infrastructural issues this raises for government.

The WRR therefore believes that the next step towards a policy infrastructure should be a co-ordination centre for AI, which should discharge a number of functions.

Possible Functions of an AI Co-ordination Centre

- *Platform*. The centre facilitates co-operation between government organizations at the policy, implementation and evaluation levels. It also serves as a contact point for international organizations, with a focus on the EU and the European Artificial Intelligence Board.
- *Knowledge*. The centre identifies AI initiatives and trajectories already under way within and outside government. This could take the form of an annual monitor of the state of AI in the Netherlands (analogous with the Dutch 'Monitor of Well-Being'), with the results used to set training priorities, identify bottlenecks and so on, and also reviewed annually by parliament (as the Monitor of Well-Being is).
- *Facilitation*. The centre plays a prominent role in facilitating our other recommendations for AI's integration in society. For example, it could work on the development of an AI licence for government employees and collect 'better intelligence' on regulatory issues.
- *Positioning*. The centre is an independent body, but in order to stimulate knowledge sharing, co-operation and a coherent policy it falls under one or more national ministries. At the same time, it is important that the centre be fed with knowledge from outside: from academia, industry and so on. To this end an external AI council of prominent experts could be established, which would meet periodically to inform and advise the centre and government in general.

With the further elaboration of these functions, the proposed co-ordination centre could provide policy directorates, supervisory bodies and executive agencies with a structure through which they can interact on a regular basis and on a variety of issues. Because different domains –healthcare, education, agriculture and so on – all have similar questions, they can benefit from learning from each other's solutions. A co-ordination centre could also help focus on those issues, opportunities and risks of AI most relevant for government. Although the centre need not necessarily focus on overall binding policy – its task initially will simply be to bring together what is happening in AI within government – it can play an important co-ordinating and facilitating role in establishing the broader legislative agenda advocated by the WRR in recommendation 7. The experiences gained can then form a

basis to facilitate policy preparation, and perhaps also policy formulation and implementation, in the next phase.

Although the proposed centre will not itself have policymaking authority (at least not initially), it will play a crucial role in this area. Its findings will need be acted upon, and it will be close to the political and public administration arenas. It is therefore important that the centre have political 'anchorage' so that policy can be made quickly if necessary, and that be political agreement and backing be available to this end. The Cyber Security Council has previously advocated the creation of a ministerial subcommittee for cyberresilience.[26] In line with this proposal, the WRR also advises that the government establish such a subcommittee to discuss substantive issues of digitalization that require integrated co-ordination. These can include cyberresilience issues, and certainly also AI. The fact that digitalization has become such a politically sensitive issue is another good reason to set up a ministerial subcommittee (Fig. 10.3).[27]

10.8 In Conclusion – The Internal Combustion Engine of the Twenty-First Century

Today the motor car is considered an integral part of our daily lives. It is thus hard to imagine what a revolutionary idea it once was. Let us try to imagine what the situation was some 100 years ago. The internal combustion engine had already been around for a while in 1921, but it was only a few years earlier that Henry Ford had proven his ability to mass produce cars. People did not understand what they were dealing with and called them 'horseless carriages'. There was also scepticism about the usefulness of motor vehicles, which was not surprising given the many defects they had. Horses continued to be more suitable for many purposes. Moreover, there was no reliable road network to allow the car to function at its best.

In time however, the car would change the face of town and countryside, and our whole way of life. A 'battle for the streets' ensued, in which cyclists, pedestrians and those who could not afford a car would eventually be barred from parts of the road network. But the development also contributed towards a new sense of freedom and individuality. Thanks to these changes, the car transformed the way society was organized – and that called for new rules, new measures and new institutions. In addition, the car demanded a new perspective on wider issues concerning the design of the public infrastructure. Both the individual measures and this broader perspective were also required to address all kinds of second-order effects, such as pollution and the risk of accidents. Automotive companies became symbols of

[26] Cyber Security Raad 2021.

[27] This therefore entails an addition to the Parliamentary Standing Committee on Digital Affairs recently established by the House of Representatives.

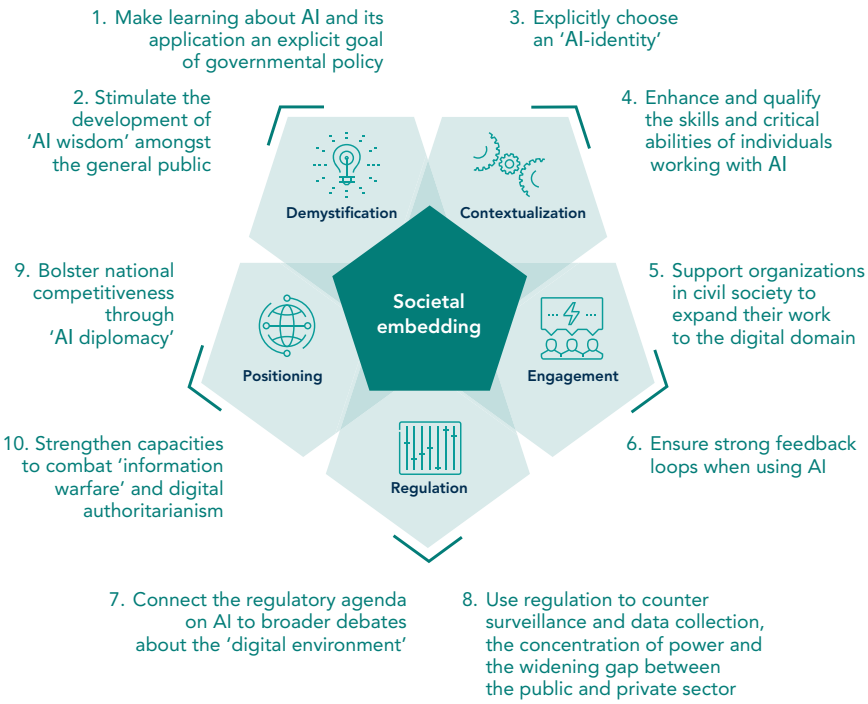


Fig. 10.3 Recommendations by task for AI's integration into society

progress and the national pride of various countries. During the Second World War the internal combustion engine made its mark on warfare in all kinds of vehicles.

These developments were impossible to foresee in 1921. In retrospect there is no simple answer to the question of how the motor car changed society, and whether that was a good or a bad thing. What is certain is that embedding the automobile in society was, and still is, a painstaking and lengthy process.

One hundred years from today we will take AI for granted just as we now take the car for granted. We cannot yet imagine what kind of world that will be, but once we are there it will be just as difficult to look back a century and imagine how AI began in the lab and then took decades to spread throughout society. We are now on the eve of that process. With the tasks we have identified in this report and the accompanying recommendations for government, the WRR hopes to help smooth the exciting path ahead.

Recommendations

Demystification

1. Make learning about AI and its potential applications an explicit goal of government's public function.
2. Stimulate the development of 'AI literacy' amongst the general public, beginning with the establishment of algorithm registers.

Contextualization

3. Explicitly choose to develop a national AI identity, then investigate what adjustments this requires to the technical ecosystem in relevant domains.
4. Strengthen the skills and critical capabilities of individuals working with AI systems by developing a suitable training and certification framework.

Engagement

5. Strengthen the capacity of civil society organizations to expand their work into the digital domain in general and AI in particular.
6. Make sure that effective feedback loops exist between AI's developers, its users and the stakeholders who experience it in practice.

Regulation

7. Link the regulation of AI to a discussion about the organization of the digital living environment and set a broad legislative agenda.
8. Use legislation to actively steer developments related to surveillance and data collection, the skewed relationship between public and private interests in the digital domain and concentration of power.

Positioning

9. Strengthen the competitiveness of the Netherlands through 'AI diplomacy' that focuses on international co-operation, in particular within the EU.
10. Develop the knowledge required to safeguard the defence of the Netherlands in the AI age. To this end strengthen the nation's capacity to defend itself in the 'information war' and against the export of 'digital dictatorship'.

Final recommendation

Build a policy infrastructure for AI, starting with a co-ordination centre that is anchored politically in a ministerial subcommittee.

References

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Chavannes, R., Strijbos, A., & Verhulst, D. (2021). Kroniek Recht en Technologie. *Nederlands Juristenblad, 2021*(16), 1350–1370. Available at: https://blog.chavannes.net/2021/04/kroniek-technologie-en-recht-2021/

De Conca, S. (2021). *The enchanted house. An analysis of the interaction of intelligent personal home assistants (IPHAS) with the private sphere and its legal protection*. Dissertation, Tilburg University.

Domini, A., & Chicot, J. (2018). *Case study report: From Concorde to Airbus, report for the European Commission*. European Commission. Available at: http://publications.europa.eu/resource/cellar/4940e0c9-2359-11e8-ac73-01aa75ed71a1.0001.01/DOC_1

Kosta, E. (2020). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*. https://doi.org/10.1111/rego.12331

Rathenau Instituut. (2021a). *Waardevol Gebruik Van Menselijke DNA-Data. Onderzoek Naar Het Borgen Van Publieke Waarden In De Waardeketen Van DNA-Data*. Rathenau Instituut. Available at: https://www.rathenau.nl/sites/default/files/2021-05/Waardevol_gebruik_van_menselijke_DNA%20data_Rathenau_Instituut.pdf

Rathenau Instituut. (2021b, June 7). *International mobility of AI scientists*, Factsheet. Available at: https://www.rathenau.nl/en/science-figures/international-mobility-ai-scientists

ROB. (2021). *Sturen of Gestuurd Worden? Over De Legitimiteit Van Sturen Met Data*. Raad voor het Openbaar Bestuur. Available at: https://www.raadopenbaarbestuur.nl/binaries/raad-openbaar-bestuur/documenten/publicaties/2021/05/25/advies-sturen-of-gestuurd-worden/Sturen_of_gestuurd_worden_Adviesrapport_2021_05.pdf

Smith, C. (1999). International collaboration in science and technology: Lessons from CERN. *European Review, 7*(1), 77–92.

TNO. (2021a). *Het Technologische Ecosysteem Van AI In Nederland* (WRR Working Paper nr. 47). Wetenschappelijke Raad voor het Regeringsbeleid.

TNO. (2021b, October 4). *Veiligere Europese Wegen Dankzij Doorbraak In Truck Platooning*. Retrieved from: https://www.tno.nl/nl/over-tno/nieuws/2021/10/veiligere-europese-wegen-dankzij-doorbraak-in-truck-platooning/

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International, 22*(4), 97–112.

WRR. (2015). *De Publieke Kern Van Het Internet. Naar Een Buitenlands Internetbeleid*. Amsterdam University Press.

WRR. (2019). *Voorbereiden Op Digitale Ontwrichting*. Wetenschappelijke Raad voor het Regeringsbeleid.

Appendix: Examples of AI Applications in the Netherlands

National government	Anticipating infrastructure maintenance projects
	Identifying which citizens need help due to unemployment
	Automation of contact with the public
	Predicting the likelihood of fraud using risk indicators
	Risk-based inspections
	Analysing documents for completeness
	E-mail security
	Automated operation of bridges and waterworks
	Translation of product information on imported goods
Municipalities	Categorizing reports of problems in the public space
	Managing traffic lights to give precedence to emergency services or cyclists
	Automated risk indicators for social security fraud
	Crowd management
	Controlling access to environmental zones via licence plate detection
Police	Anticipating crime patterns
	Automated assistance with online crime reports
	Assessing the likelihood of solving cold cases
	Matching photos of suspects with an existing photo database
Education	Digital invigilating during exams
	Adaptive learning tools for primary school students
Healthcare	ICU triage assistance
	Evaluations of medical imagery
	Automated reports of consultations
	Assistance with diagnostic procedures
Financial sector	Predicting price trends for investors
	Assessing the creditworthiness of customers
	Assessing potential discounts on insurance premiums
Agri-food	Monitoring animal welfare in barns
	Crop quality inspections
	Just-in-time machine maintenance
	Assessing soil quality using satellite data

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6

Retail trade	Predicting product shelf life
	Dynamic product pricing
	Personalized marketing
	Predicting purchasing behaviour
Media	Automated production of news articles
	Social media microtargeting
	Provision of personalized content
Law	Automated case law searches
	Automated online disputes mediation
Workplace	Predicting matches for vacancies
	Analysing employee performance
	Automated routing for drivers

Glossary

AI diplomacy An integrated international policy aimed at forming partnerships in AI. The policy focuses on five domains: fundamental research, commercial applications, regulation, ethical guidelines and standards.

AI identity A country or region's distinctive character in AI. An AI identity can be built by specializing in a particular type of AI or adopting a specific role within the international AI arena.

AI literacy The basic knowledge and competences needed to participate in a society dominated by AI.

Contextualization (Task 2) Contextualization concerns the embedding of technology in the social and technical ecosystem. The idea is that a technology will only work if it is in harmony with the technical and social context.

Demystification (Task 1) Demystification involves responding to overblown representations or incorrect perceptions of a technology to encourage a better understanding of what the technology actually means and what it can do.

Digital living environment The spatial planning and design of the environment in which a technology is used. In the digital living environment, rules and instruments are used to encourage the deployment of a technology there where it can contribute the most, and on a scale and for purposes that are congruent with those of society.

Engagement (Task 3) Engagement refers to how various groups in society are involved in the development of a technology to ensure that their interests are considered in the process.

Positioning (Task 5) Positioning is the way a country strategically deploys a technology in relation to the international arena. In addition to interaction with other countries, this also concerns the relationships with non-state actors such as industries and organizations (including criminal organizations).

Regulation (Task 4) Regulation is the development of frameworks to steer the development and use of a technology in the right direction. This can be done by

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6

establishing legislation, regulations, norms and standards, both at the national and international level.

System technology A multifunctional technology that becomes intertwined with the system of a society and functions as a part of a system of other technologies. Like general purpose technologies, system technologies are pervasive, lead to complementary innovations, and are subject to continuous technical improvement. The term 'system technology' emphasizes the systematic nature and the systemic effects of the technology. The societal embedding of a system technology involves a complex and lengthy process of adjustment and adaptation.

AI effect The effect of the development of computer skills on what we consider to be human intelligence. Once a computer has mastered a certain skill, people tend to see this not as intelligent behaviour, but as simple calculation.

AI winter A period with relatively little scientific progress in AI. To date there have been two AI winters: the first from 1969 to 1982 and the second from the late 1980s to the 1990s.

AI summer A period with relatively much scientific progress and activity in AI. We are currently experiencing an AI summer. This period began around 2012, when the scientific breakthroughs based on the neural network approach caused an explosion of activity in AI.

Algorithm A specific instruction for solving a problem or performing a calculation.

AlphaGo A computer program that can play the board game Go. The program was developed by the British research lab DeepMind (acquired by Google in 2014). In 2016, AlphaGo defeated the world champion Lee Sedol.

Artificial General Intelligence (AGI) Artificial Intelligence that can match human intelligence in all areas. This is also known as 'strong' or 'full' AI.

Artificial Super Intelligence (ASI) Artificial Intelligence that is superior to human intelligence in all areas.

Artificial Intelligence Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. This is the definition of the European Commission's AI HLEG.

Artificial Neural Networks (ANNs, see also 'Connectionism') An approach within AI that uses artificial neural networks to simulate the functioning of neurons in the human brain. These networks are fed with large amounts of data from which they distil patterns themselves, i.e., no rules are set in advance.

Automation bias A psychological mechanism that causes people to blindly follow a computer's suggestions, even if these are incorrect or go against common sense.

Backpropagation An algorithm used to improve pattern recognition in artificial neural networks (ANNs). Here, the algorithm looks at the results of the 'output layer' and traces information coming from the hidden layers (under the output layer), where individual units are identified that need to be modified to make the algorithm work more effectively.

Blockchain A decentralized system in which transactions are recorded in an invariable and public network. The term 'blockchain' refers to the structure of the database; a new transaction forms a 'block' with information about this new

transaction and about the previous transaction. If the transaction is approved, the block joins the other blocks to form an 'information chain'.

Central Processing Unit (CPU) A processor (piece of hardware) that receives, controls and executes the basic instructions of the program code.

Civil technology Technology developed by companies rather than (or only to a lesser extent) by the government or independent researchers.

Closure A moment in the development of a technology when a particular design or type of use becomes the norm, at which time the controversy surrounding the technology often goes away.

Collingridge dilemma This concerns the moment of regulation and the uncertainty that this entails. In the early stages of a new technology, it is easy to establish frameworks, but it is often unclear what regulation is needed. In addition, there is no need for regulation yet, because the effects are still unknown. At a later stage, it becomes clearer where rules are needed, but establishing regulation is much more difficult and costly, because standard practices and vested interests have also since been established.

Computer vision AI systems for observing, analysing and interpreting visual information such as photos, videos and the physical environment. One of the best-known applications of computer vision is facial recognition.

Connectionism An approach within AI that uses artificial neural networks to simulate the functioning of neurons in the human brain (see also Artificial Neural Networks). Connectionism and the 'symbolic approach' form the two main streams of AI.

Crime Anticipation System (CAS) A system that detects crime patterns and predicts where and when incidents are most likely. The police can use this information to anticipate crime, for example by monitoring a specific region more closely.

Data Information that can be stored by a computer. Usually, these are individual data without further explanation, which means that they have to be classified under a certain context or collection in order to be interpreted.

Deep Blue A chess computer developed by IBM. In 1997, Deep Blue beat chess grandmaster Garry Kasparov.

Deepfake An image or sound clip generated by AI that professes to be unmanipulated. Some deepfakes are so realistic that it is difficult to distinguish them from the real thing.

Deep Learning (DL) A form of machine learning that simulates the functioning of neurons in the human brain. Deep Learning uses multi-layer networks, hence the term 'deep'.

Dual-use technology Technology that can be used for both civil and military purposes.

Enveloping Adapting an environment to a technology.

Expert systems AI systems that follow pre-programmed rules based on expert knowledge. Expert systems fall under rule-based AI.

Federated learning A form of machine learning in which algorithms are improved by adapting their parameters to those of other datasets, without combining the

data in these datasets. The dataset does not need to be included in a central server. In effect, the algorithm is sent to the data instead of the other way round.

General Purpose Technology (GPT) Technologies that have a generic character (rather than a single, limited application) and can therefore be applied in countless forms and for a variety of purposes. Examples of earlier GPTs are the internal combustion engine and electricity.

Generative Adversarial Networks (GANs) A technique within AI that allows algorithms to improve each other. For example, an algorithm generates an image, and another algorithm tries to detect if the image is fabricated or authentic. The first algorithm continues to generate new images until the second algorithm is convinced that the image is authentic.

Geo-economy The domain at the interface of economy and competitiveness on the one hand and geopolitics and security on the other.

GPT-3 (Generative Pre-trained Transformer 3) Language processing software that can generate natural texts based on relatively little input.

Graphic Processing Units (GPUs) Processors (hardware) typically used for processing complex images and graphical data. Due to their immense computing power, GPUs are also suitable for the complex calculations required for advanced AI.

Human-in-the-loop A form of interaction between humans and machines whereby an AI system is involved in a process but the ultimate responsibility of any decisions rests with a human being.

Human-on-the-loop A form of interaction between humans and machines whereby an AI system can take independent decisions without any human intervention. The process is monitored by a human being who is able to intervene and make adjustments.

Human-out-of- the-loop A form of interaction between humans and machines whereby there is no human involvement, and the AI system functions completely autonomously.

Internet of Things (IoT) The network of digital connections between objects and devices in the physical environment, by means of sensors and the internet, which allows information to be exchanged between them.

Logical AI See Symbolic AI.

Luddites Name for the English factory workers who rebelled against the mechanization of labour in the early nineteenth century.

Machine Learning (ML) An AI application that can perform predictive analyses based on patterns in datasets.

Microtargeting Carefully aligning advertising with the interests and sensibilities of individual users for commercial or political purposes.

Model A formal description of a system, process or equation used to simplify a complex subject.

Natural Language Processing (NLP) AI for reading, analysing and generating human language. The goal is for these algorithms to learn our 'natural' language of communication well enough to be able to perform the tasks used to interpret text.

Moore's law The principle that the number of transistors on a chip roughly doubles every two years.

Moravec's paradox The phenomenon that certain things that are difficult for humans are easy for computers and vice versa.

Productivity paradox The phenomenon that, while there are often high expectations of new technologies, the actual impact of these technologies on economic productivity is usually disappointing in the short term.

Proxy (proxies) A proxy is a data point that is indicative of other data. Proxies can be used to reconstruct information that has not been directly measured or collected. For example, language use could be a proxy for gender, and a postcode could be a proxy for ethnicity.

Quantum computing Technology that works with quantum bits, or qubits. Quantum bits can take on multiple states simultaneously, allowing a much higher number of possible calculations than traditional computers that work with bits in a binary logic.

Rule-based AI See Symbolic AI.

Reinforcement learning A form of machine learning where the algorithm is trained to follow certain strategies through a system of positive and negative feedback.

Soft law Soft law comprises norms, codes and recommendations, which by their very nature have little coercive or binding force. However, soft law can be used to form a certain opinio juris.

Speech recognition The field of AI concerned with observing, analysing and interpreting spoken human language. This involves the use of algorithms to distinguish words and sentences in spoken language and convert them into text for analysis.

Strong AI See Artificial General Intelligence (AGI).

Supervised learning A form of machine learning in which a system is fed with data that has been pre-labelled by humans.

Symbolic AI An approach to AI based on logical rules and formulas. For example, 'if-then' rules are used to reason what the data could mean. This is also known as 'logical' or 'rule-based' AI. Connectionism and the Symbolic AI form the two main streams of AI.

System Risk Indication (SyRI) System Risk Indication (SyRI) is a legal instrument used by the Dutch government to combat fraud involving benefits, allowances and taxes, among others. SyRI was used by a number of Dutch municipalities until the court ruled in 2020 that the system is illegal because it violates the right to privacy.

Techno-chauvinism The belief that technology can provide a solution to any problem

Techno-determinism/technological determinism The notion that technology operates autonomously, and that society will simply have to adapt to it.

Techno-solutionism/technological solutionism The tendency to re-envision complex societal phenomena as issues to which technology has the answer.

Tensor Processing Unit (TPU) A processor (piece of hardware) designed specifically for machine learning applications.

Turing test An experiment devised by Alan Turing in 1950, in which a computer pretends to be a human being. A computer passes this test if a human is unable to establish whether the answers were provided by a human or a computer. Variants of this test are used to compare AI systems with various human skills, such as the use of language.

Unsupervised learning A form of machine learning in which the program is fed with unlabelled data and the algorithm must distil patterns from this data itself.

Watson Watson is a computer program developed by IBM to answer spoken questions on the game show Jeopardy! The program did this using large amounts of information from a database. In 2011, the program defeated the reigning human champions.

Weak AI (also called narrow AI) AI that focuses on a specific skill such as image or speech recognition, often within a specific context.

Bibliography

Aanhangsel Handelingen II 2017/2018, nr. 1645. (2018, April 4). *Aanhangsel van de Handelingen*. Available at: https://zoek.officielebekendmakingen.nl/ah-tk-20172018-1645.pdf

Access Now. (2020). *Europe's approach to Artificial Intelligence: How AI strategy is evolving*. Access Now. Available at: https://www.accessnow.org/cms/assets/uploads/2020/12/Europes-approach-to-ai-strategy-is-evolving.pdf

Ackerman, E. (2021, January 7). How Boston dynamics taught its robots to dance. *IEEE Spectrum*. Available at: https://spectrum.ieee.org/automaton/robotics/humanoids/how-boston-dynamics-taught-its-robots-to-dance

Ad Hoc Expert Group. (2020). *First draft of the recommendation of the ethics of Artificial Intelligence*. UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000373434

Afdeling Bestuursrechtspraak van de Raad van State. (2017). ECLI:NL:RVS: 2017:1259 (AERIUS I), ruling 17 May 2017. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RVS:2017:1259

Afdeling Bestuursrechtspraak van de Raad van State. (2018). ECLI:NL:RVS: 2018:2454 (AERIUS II), ruling 18 July 2018. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RVS:2018:2454

AFM en DNB. (2019). *Artificiële Intelligentie In De Verzekeringssector Een Verkenning*. Autoriteit Financiële Markten, De Nederlandsche Bank. Available at: https://www.afm.nl/~/profmedia/files/rapporten/2019/afm-dnb-verkenning-ai-verzekeringssector.pdf?la=nl-nl

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Agrawal, A., Gans, J. & Goldfarb, A. (reds.). (2019). *The economics of Artificial Intelligence: An Agenda*. National Bureau of Economic Research/University of Chicago Press.

Ahmed, S. (2019). Credit cities and the limits of the social credit system. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 55–61). Air University Press.

AIV en CAVV. (2015). *Autonome Wapensystemen: De Noodzaak Van Betekenisvolle Menselijke Controle* (Nr. 97 AIV/Nr. 26 CAVV). Adviesraad Internationale Vraagstukken en Commissie van Advies Inzake Volkenrechtelijke Vraagstukken. Available at: https://www.adviesraadinternationalevraagstukken.nl/documenten/publicaties/2015/10/02/autonome-wapensystemen

Ajami, S. (2016). Use of Speech-To-Text Technology for documentation by healthcare providers. *The National Medical Journal of India, 29*(3), 148–152.

Albert Heijn. (2019, May 20). *Albert Heijn Zet Kunstmatige Intelligentie In Tegen Voedselverspilling*, Albert Heijn Nieuws. Available at: https://nieuws.ah.nl/albert-heijn-zet-kunstmatige-intelligentie-in-tegen-voedselverspilling/

Algemene Rekenkamer. (2021). *Aandacht voor Algoritmes*. Algemene Rekenkamer.

Alliance for Artificial Intelligence. (n.d.). *Introducing ALLAI*. Available at: https://allai.nl/about-us/

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6

Amnesty International. (2020a). *We sense trouble: Automated discrimination and mass surveillance in predictive policing in The Netherlands*. Amnesty International. Available at: https://www.amnesty.org/en/wp-content/uploads/2021/05/EUR3529712020ENGLISH.pdf

Amnesty International. (2020b). *Out of control: Failing Eu Laws for digital surveillance export*. Amnesty International.

Amsden, A. (1989). *Asia's next giant: South Korea and late industrialization*. Oxford University Press.

Amsterdam Algoritmeregister. (n.d.). *Anderhalve Meter Monitor*. Available at: https://algoritmeregister.amsterdam.nl/anderhalve-meter-monitor/

Andersen, R. (2020, September). The Panopticon is already here. *The Atlantic*. Available at: www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/

Apple, C. (2020, June 24). Instant Gratification: The history of Instagram. *Spokesman*. Available at: https://www.spokesman.com/stories/2020/jun/24/how-instagram-hit-one-billion-users/

Association for Advancing Automation. (2018, January 25). Why AI won't overtake the world, but is worth watching. *Industry Insights*. Available at: www.robotics.org/content-detail.cfm/Industrial-Robotics-Industry-Insights/Why-AI-Won-t-Overtake-the-World-but-Is-Worth-Watching/content_id/6979

Austin, E. (2019, May 20). *Facebook Liegt Tegen Tweede Kamer Over Verkiezingsmanipulatie*. Bits of Freedom. Available at: https://www.bitsoffreedom.nl/2019/05/20/facebook-liegt-tegen-tweede-kamer-over-verkiezingsmanipulatie/

Autoriteit Persoonsgegevens. (2020, October 29). Pas Op Met Camera's Met Gezichtsherkenning. *nieuwsbericht*. Available at: https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-pas-op-met-camera%E2%80%99s-met-gezichtsherkenning

AWTI. (2020). *Krachtiger Kiezen voor Sleuteltechnologieën*. Adviesraad voor wetenschap, technologie en innovatie. Available at: https://www.awti.nl/binaries/awti/documenten/adviezen/2020/01/30/awti-advies-krachtiger-kiezen-voor-sleuteltechnologieen/Krachtiger+kiezen+voor+sleuteltechnologie%C3%ABn.pdf

Bakker, S. (2017). *From luxury to necessity: What the railways, electricity and automobile teach us about the IT revolution*. Boom Uitgeverij.

Bakker, S., & Korsten, P. (2021). *Artificiële Intelligentie Als Een general purpose technology: Strategische Belangen Van Verantwoorde Inzet In Historisch Perspectief* (WRR Working Paper nr. 41). Wetenschappelijke Raad voor het Regeringsbeleid. Available at: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Barkhuysen, T. (2021). Handhaving van de AVG: de AP kan het niet alleen. *Nederlands Juristenblad, 572*(8), 585.

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfield, D., & Squicciarini, M. (2020). *Identifying and measuring developments in Artificial Intelligence: Making the impossible possible* (OECD Science, Technology and Industry Working Papers 20(5)). OECD Publishing.

Bendett, S. (2019). *The development of Artificial Intelligence in Russia.* In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 168–198). Air University Press.

Benjamin, R. (2019). *The race after technology: Abolitionist tools for the New Jim code*. Polity Press.

Bennett, M., & Gollan, L. N. (2015). *The illusion of Newness: The importance of history in understanding the law-technology interface* (UNSW Law Research Paper, 2015-71). University of New South Wales.

Bennett Moses, L. (2007a). Why have theory of Law and Technological Change? *Minnesota Journal of Law, Science & Technology, 8*, 589–606.

Bennett Moses, L. (2007b). Recuring Dilemmas: The Law's race to keep up with technological change. *University of Illinois Journal of Law, Technology and Policy, Fall*, 239–285.

Bernstein, G. (2006). The Paradoxes of technological diffusion: Genetic discrimination and internet privacy. *Connecticut Law Review, 39*(1), 243–297.

Bertolini, A. (2020). *Artificial Intelligence and Civil Liability. Onderzoek in Opdracht van de Commissie Juridische Zaken van het Europese Parlement*. European Parliament. Available at: http://www.europarl.europa.eu/supporting-analyses

Bijker, W. (1995). *Of Bicycles, Bakelites, and Bulbs. Toward a theory of sociotechnical change*. MIT Press.

Bijlage bij de brief van staatssecretaris van Economische Zaken en Klimaat. (2019, May 17). *Discussienotitie: Toekomstbestendigheid Mededingingsbeleid In Relatie Tot Online Platforms*. Available at: https://www.rijksoverheid.nl/documenten/vergaderstukken/2019/05/20/bijlage-1-discussienotitie-toekomstbestendig-mededingsbeleid-in-relatie-tot-online-platforms

Bijlsma, M., Overvest, B., & Straathof, B. (2016). *Marktordening Bij Nieuwe ICT-Toepassingen. Vroegtijdig Ingrijpen Nodig* (CPB Policy Brief 2016/09). Centraal Planbureau. Available at: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-2016-09-Marktordening-bij-nieuwe-ICT-toepassingen.pdf

Bits of Freedom. (n.d.). *Gezichtsherkenning*. Available at: https://www.bitsoffreedom.nl/dossiers/gezichtsherkenning/

Black, J., & Murray, A. (2019). Regulating AI and machine learning: Setting the regulatory Agenda. *European Journal of Law and Technology, 10*(3). Available at: https://ejlt.org/index.php/ejlt/article/download/722/980

Blankena, F. (2013, September 9). Identiteitskaart Wordt Mogelijk Zonder Vingerafdruk. *Binnenlandsbestuur Digitaal*. Available at: https://www.binnenlandsbestuur.nl/digitaal/nieuws/identiteitskaart-wordt-mogelijk-zonder.9097480.lynkx

Boden, M. (2018). *Artificial Intelligence: A very short introduction*. Oxford University Press.

Boland, H. (2018, September, 2). Britain faces an AI brain drain as tech giants raid top Universities. *The Telegraph*. Retrieved from https://www.telegraph.co.uk/technology/2018/10/24/britains-ai-industry-must-avoid-brain-drain-us-mps-warn/

Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bousquet, A. (2009). *The scientific way of warfare: Order and Chaos on the battlefields of modernity*. Hurst.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Brand, S. (1995, March 1). We Owe it all to the Hippies. *Time Magazine*. Available at: http://content.time.com/time/subscriber/article/0,33009,982602,00.html

Bratton, B. (2016). *The stack: On software and sovereignty*. MIT Press.

Braun, C., & Stolk, R. (2020, March 4). Procederen Uit Naam Van Het Algemeen Belang. *Montesquieu Instituut*. Available at: https://www.montesquieu-instituut.nl/id/vl6ck1v35y97/nieuws/procederen_uit_naam_van_het_algemeen

Bresnahan, T., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics, 65*(1), 83–108.

Brooks, R. (2018, January 1). My dated predictions. blog, *Rodneybrooks*. Available at: https://rodneybrooks.com/my-dated-predictions/

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Brown, R. (2021). Property ownership and the legal personhood of artificial intelligence. *Information and Communications Technology Law, 2*, 208–234.

Brownsword, R., en Goodwin, M. (2012). *Law and the technologies of the twenty-first century*. Cambridge University Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G., Steinhardt, J., Flynn, C., HÉgeartaigh, S., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evens, O., Page, M., Bryson, J., Yampolskiy, R., & Amodei, D. (2018). *The Malicious use of Artificial Intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the modern productivity Paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans en A. Goldfarb (2019) *The Economics of Artificial Intelligence: An Agenda* (pp. 23–57). University of Chicago Press.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI Frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute. Available at: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.pdf?shouldIndex=false

Bui, P., & Liu, Y. (2021, May 18). Using AI to help find answers to common skin conditions. [Blog] Google.nl. Available at: https://blog.google/technology/health/ai-dermatology-preview-io-2021/

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.

Buruma, Y. (2020). International Law and Cyberspace. Issues of sovereignty and the common good. In *International Law for a Digitalised World* (pp. 69–111). knvir/T.M.C. Asser Press.

Business Insider. (2010, October 15). Mark Zuckerberg, Moving Fast and Breaking Things', *Businessinsider.nl*. Available at: https://www.businessinsider.com/mark-zuckerberg-2010-10?international=true&r=US&IR=T

Bygrave, L. (2021). The 'Strasbourg Effect' on data protection in light of the 'Brussels Effect': Logic, mechanics and prospects. *Computer Law & Security Review, 40*, 105460.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI now 2017 report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf

Cath, C., Wachter, S., Mittelstadt, B., Toledo, M., & Floridi, L. (2018). Artificial intelligence and the 'Good society'. The US, EU, and UK approach. *Science and Engineering Ethics, 24*, 505–528.

CB Insights. (2018, April 26). *Rise of China's Big Tech in AI; what Baidu, Alibaba, and Tencent are working on*. CB Insights Research Briefs. Available at: https://www.cbinsights.com/research/china-baidu-alibaba-tencent-artificial-intelligence-dominance/

CB Insights. (2021, June 24). *Despite a Pandemic Slump, the AI sector remains hot for acquirers.* CB Insights Research Briefs. Available at: https://www.cbinsights.com/research/artificial-acquisitions-trends-annual-deals/

CEG. (2018). *Digitale Dokters: Een Ethische Verkenning Van Medische Expertsystemen*. Centrum voor Ethiek en Gezondheid. Available at: https://www.ceg.nl/binaries/ceg/documenten/signalementen/2018/07/04/digitale-dokters%2D%2D-een-ethische-verkenning-van-medische-expertsystemen/webversie_CEG_Digitale_dokters_Een_ethische_verkenning_van_medische_expertsystemen.pdf

Centrale Raad van Beroep. (2019). ECLI:NL:CRVB:2019:1737, ruling 15 May 2019. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:CRVB:2019:1737

Chavannes, R., Strijbos, A., & Verhulst, D. (2021). Kroniek Recht en Technologie. *Nederlands Juristenblad, 2021*(16), 1350–1370. Available at: https://blog.chavannes.net/2021/04/kroniek-technologie-en-recht-2021/

Chen, S. (2017, October 12). China to build giant facial recognition database to identify any citizen within seconds. *South China Morning Post*. Available at: https://www.scmp.com/news/china/society/article/2115094/china-build-giant-facial-recognition-database-identify-any

Chen, Y., Casagrande, N., Zhang, Y., & Brenner, M. (2019). *Using Wavenet Technology to Reunite speech-impaired users with their original voices*, DeepMind.nl, 18 December. Available at: https://deepmind.com/blog/article/Using-WaveNet-technology-to-reunite-speech-impaired-users-with-their-original-voices

Chiusi, F., Fischer, S., Kayser-Bril, N., & Spielkamp, M. (2020). *Automating Society Report 2020*. AlgorithmWatch. Available at: https://www.ivir.nl/publicaties/download/Automating-Society-Report

Choi, W., van Eck, M., & Hukshorn, H. (2021). *Hoe Gemeenten Besluiten Over Algoritmen En Mensenrechten*, onderzoek in opdracht van het College voor de Rechten van de Mens. Hooghiemstra en partners.

Cihon, P. (2019). *Standards for AI governance: International standards to enable global coordination in AI Research & Development* (Technical Report). Future of Humanity Institute.

Claus, S. (2017, September 12) Een algoritme dat aan je gezicht ziet of je homo of hetero bent. *Trouw*. Available at: https://www.trouw.nl/nieuws/een-algoritme-dat-aan-je-gezicht-ziet-of-je-homo-of-hetero-bent~b7b99615/

Coeckelbergh, M. (2020). AI for climate: Freedom, Justice, and other ethical and political challenges. *AI and Ethics, 1*, 67–72.

Coglianese, C., & Lehr, D. (2019). Transparancy and algorithmic governance. *Administrative Law Review, 71*(1), 12–57.

College voor de Rechten van de Mens. (2021). *Handreiking: (Semi-)Geautomatiseerde Besluitvorming Door De Overheid*. College voor de Rechten van de Mens. Available at: https://publicaties.mensenrechten.nl/file/002e83bf-47f8-44fd-846e-46aff40f8a56.pdf

COMEST. (2017). Report of COMEST on Robotics Ethics. [Report] UNESCO. Geraadpleegd van: https://unesdoc.unesco.org/ark:/48223/pf0000253952

Committee on the Judicary. (2020). *Investigation of competition in digital markets*. *Majority Staff Report and Recommendations*, Subcommitte on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary. Available at: https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Crawford, K., Dobbe, T., Dryer, G., Fried, B., Green, E., Kaziunas, A., Kak, V., Mathur, E., McElroy, A., Sánchez, D., Raji, J., Rankin, R., Richardson, J., Schultz, S. W., & Whittaker, M. (2019). *AI Now 2019 Report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Creemers, R. (2019). The international and foreign policy impact of China's Artificial Intelligence and big-data strategies. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 129–135). Air University Press.

Crémer, J., De Montjoye, Y., & Schweitzer, H. (2019). *Competition policy for the digital era*. Europese Commissie. Available at: http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

CSR. (2018). *CSR Adviesbrief inzake opheffing 'Numeri Fixi'*. Cyber Security Raad. Available at: https://www.cybersecurityraad.nl/binaries/cybersecurityraad/documenten/adviezen/2018/07/26/csr-adviesbrief-inzake-opheffing-%E2%80%98numeri-fixi/CSR+Adviesbrief+inzake+opheffing+%E2%80%98Numeri+Fixi%27.pdf

CSR. (2020). *Naar Structurele Inzet Van Innovatieve Toepassingen Van Nieuwe Technologieën Voor De Cyberweerbaarheid Van Nederland* (CSR-advies 2020, nr. 5). Cyber Security Raad. Available at: https://www.cybersecurityraad.nl/adviezen/documenten/adviezen/2020/09/18/csr-advies-%E2%80%98naar-structurele-inzet-van-innovatieve-toepassingen-van-nieuwe-technologieen-voor-de-cyberweerbaarheid-van-nederland%E2%80%99%2D%2D-csr-advies-2020-nr.-5

CSR. (2021). *Nederlandse Digitale Autonomie en Cybersecurity* (CSR-advies 2021, nr. 3). Cyber Security Raad. Available at: https://www.cybersecurityraad.nl/documenten/adviezen/2021/05/14/csr-advies-nederlandse-digitale-autonomie-en-cybersecurity%2D%2D-csr-advies-2021-nr.-3

Daisy Intelligence. (n.d.). [Webpage]. Available at: https://www.daisyintelligence.com/

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accomodation. *Philosophy & Technology, 29*(3), 245–268.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in Judicial decisions. *Proceedings of the National Academy of Sciences, 108*(17), 6889–6892.

Das, D., de Jong, R., Kool, L., & Gerritsen, M. M. V. J. (2020). *Werken Op Waarde Geschat – Grenzen Aan Digitale Monitoring Op De Werkvloer Door Middel Van Data, Algoritmen En AI*. Rathenau Instituut.

Data Science Center Tilburg. (n.d.). *Data Science for Social Good* [ds for Social Good]. Available at: https://www.tilburguniversity.edu/research/institutes-and-research-groups/data-science-center/dssg

Dauverge, P. (2020). *AI in the Wild – Sustainability in the age of Artificial Intelligence*. MIT Press.

Davenport, C. (2019). *The Space Barons: Elon Musk, Jeff Bezos and the Quest to Colonize the Cosmos*. New York Public Affairs.

Davidson, D., & Delhaas, R. (2020, April 22). *Als De Politiek In Ieder Oor Een Andere Belofte Fluistert*. Argos. Available at: https://www.vpro.nl/argos/lees/nieuws/2020/microtargeting-in-Nederland.html

De Conca, S. (2021). *The enchanted house. An analysis of the interaction of intelligent personal home assistants (IPHAS) with the private sphere and its legal protection*. Dissertation, Tilburg University.

De Poorter, J., & Goossens, J. (2019). Effectieve Rechtsbescherming Bij Algoritmische Besluitvorming In Het Bestuursrecht. *Nederlands Juristenblad, 44*, 3303–3312.

De Ree, M. (2021, April 29). Onderzoek Naar Eerlijke En Uitlegbare Algoritmen. *CBS.nl*. Available at: https://www.cbs.nl/nl-nl/corporate/2021/17/onderzoek-naar-eerlijke-en-uitlegbare-algoritmen

De Rijke, M. (2019, April 8). Investeer In Kennisbasis AI of word Een Toeschouwer. *NRC*. Available at: https://www.nrc.nl/nieuws/2019/04/08/investeer-in-kennisbasis-ai-of-wordt-een-toeschouwer-a3956136

DeepMind. (2019). *Machine learning can boost the value of wind energy*, Deepmind.nl, 26 February. Available at: https://deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy

Delcker, J. (2018, June 27). *Merkel warns of AI brain drain to foreign tech companies*. Politico. Retrieved from: https://www.politico.eu/article/merkel-artificial-intelligence-warns-brain-drain-to-foreign-tech-companies/

DenkWerk. (2018). *Artificial Intelligence in Nederland: Zelf Aan Het Stuur*. Available at: https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf

Dennett, D. (2019). What can we do? In J. Brockman (red.), *Possible minds: Twenty-five ways of looking at AI* (pp. 41–53). Penguin.

Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. (2019). Facebook-Auftritte von öffentlichen Stellen des Bundes, brief d.d. 20 May 2019, 61924/2021. Available at: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Rundschreiben/Allgemein/2021/Facebook-Auftritte-Bund.pdf?blob=publicationFile&v=2

Dickson, B. (2020, March 2). Understanding the limits of CNNs, one of AI's Greatest Achievements. *TechTalks*. Available at: https://bdtechtalks.com/2020/03/02/geoffrey-hinton-convnets-cnn-limits/

Dignum, V. (2019). *Responsible Artificial Intelligence: How to develop and use AI in a responsible way*. Springer.

Dignum, V. (n.d.). *There is no AI – Race and if there is, it's the wrong one to run*. Available at: https://allai.nl/there-is-no-ai-race/

Dijksterhuis, E. (2006 [1950]). *De Mechanisering Van Het Wereldbeeld*. Amsterdam University Press.

Ding, J. (2018). *Deciphering China's AI dream* (Future of Humanity Institute Technical Report). University of Oxford.

Ding, J. (2019). The interests behind China's Artificial Intelligence dream. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 43–47). Air University Press.

Diogo, M., & van Laak, D. (2016). *Europeans globalizing: Mapping, exploiting, exchanging*. Palgrave Macmillan.

Domingos, P. (2017). *The master algorithm: How the Quest for the ultimate learning machine will remake our world*. Penguin Random House.

Domini, A., & Chicot, J. (2018). *Case study report: From Concorde to Airbus, report for the European Commission*. European Commission. Available at: http://publications.europa.eu/resource/cellar/4940e0c9-2359-11e8-ac73-01aa75ed71a1.0001.01/DOC_1

Dommering, E. (red.). (2000). *Informatierecht: Fundamentele Rechten Voor De Informatiesamenleving.* Otto Cramwinckel.

Dreyfus, H., & Dreyfus, S. (1986). *Mind over Machine*. The Free Press.

Drezner, D. (2019). Economic Statecraft in the age of Trump. *The Washington Quarterly, 42*(3), 7–24.

Dutton, T. (2018, June 28). An overview of national *AI* Strategies. *Medium*. Available at: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Edgerton, D. (2008). *The shock of the old: Technology and global history since 1900*. Profile books.

El-Dardiry, R., Overvest, B., Dinkova, M., & Albers, R. (2021). *Brave New Data. Databeleid In Een Imperfecte Wereld* (CPB Policy Brief, May 2021). Centraal Planbureau. Available at: https://www.cpb.nl/brave-new-data-databeleid-in-een-imperfecte-wereld

ELLIS. (2018). *Open letter*. ELLIS Society. Retrieved from: https://ellis.eu/letter

Elsevier. (2018). *ArtificiaI Intelligence: How knowledge is created, transferred, and used*. Elsevier. Retrieved from: https://www.elsevier.com/connect/resource-center/artificial-intelligence

EPRS. (2020). *Data subjects, digital surveillance, AI and the future of work*. European Parliament. Retrieved from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

EuroHPC JU. (2021, April 20). Vega Online: The EU First Eurohpc Supercomputer Is Operational. Eurohpc-ju.europa. Retrieved from: https://eurohpc-ju.europa.eu/press-releases/vega-online-eu-first-eurohpc-supercomputer-operational

European Center for Not-for-Profit Law. (2020). Being aware: Incorporating Civil Society into national strategies on Artificial Intelligence. Country Papers On Participatory Processes In Drafting National Ai Policies In The Czech Republic, The Netherlands, Australia And Canada. ECNPL. Available at: https://ecnl.org/publications/being-ai-ware-incorporating-civil-society-national-strategies-artificial-intelligence

European Commission. (2018a, December 7). Member States and commission to work together to boost Artificial Intelligence "Made In Europe". Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6689

European Commission. (2018b). *Annex to the coordinated plan on the development and use of Artificial Intelligence made in Europe*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017

European Commission. (2020). *A European Strategy for Data*, COM(2020) 66 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX:52020DC0066

European Commission. (2021a). *2030 Digital Compass: the European way for the Digital Decade*. Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118

European Commission. (2021b [2018]). *EU coordinated action plan on AI 2021 review*, COM(2021) 205 final. Available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

European Commission. (2021c). *Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Commission. (n.d.-a). *Destination Earth*. Available at: https://digital-strategy.ec.europa.eu/en/policies/destination-earth

European Commission. (n.d.-b). *Public Consultation: White Paper on Artificial Intelligence – A European Approach*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12270-White-Paper-on-Artificial-Intelligence-a-European-Approach/public-consultation_en

European Data Protection Board and European Data Protection Supervisor. (2021). *Joint Opinion 5/2021 on the proposal for a proposal for the regulation of the European Parliament and of the council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act), 28 June 2021*. EDPS. Available at: https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

European Parliament. (2020). European Parliament resolution of 20 October 2020 with recommendations to the commission on a civil liability regime for Artificial Intelligence (2020/2014(inl)). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html

European Political Strategy Centre. (2018). *The age of Artificial Intelligence: Towards a European strategy for human-centric machines* (EPSC Strategic Notes). EPSC. Available at: https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/age-artificial-intelligence-towards-european-strategy-human-centric

European Union Agency for Fundamental Rights. (2020). *Getting the future right. Artificial Intelligence and fundamental rights*. Publications Office of the European Union. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

Europese Commission. (2018, October 29). *Quantum Technologies Flagship kicks off with first 20 projects*. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/de/MEMO_18_6241

Feldstein, S. (2019). *The global expansion of AI surveillance*. Carnegie Endowment for International Peace.

Fiebig, T., Gürses, S., Gañán, C., Kotkamp, E., Kuipers, F., Lindorfer, M., Prisse, M., & Sari, T. (2021). *Heads in the Clouds: Measuring the implications of Universities migrating to Public Clouds*. Available at: https://arxiv.org/abs/2104.09462

Field, A. (2008). *Does economic history need gpts?* Available at: http://ssrn.com/abstract=1275023

Fierens, M., van Gool, E., & De Bruyne, J. (2021). De Regulering Van Artificiële Intelligentie (Deel 1) – Een Algemene Stand Van Zaken En Een Analyse Van Enkele Vraagstukken Inzake Consumentenbescherming. *Rechtskundig Weekblad, 84*(25), 962–980.

Fight for the Future. (n.d.). Interactieve Kaart. Available at: https://www.banfacialrecognition.com/map/

FLIR UGS. (n.d.). *Combat- Proven Robots.* Available at: https://www.flir.com/uis/ugs/

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (2021). The European Legislation on AI: A brief analysis of its philosophical approach. *Philosophy and Technology, 34*, 215–222. Available at: https://link.springer.com/article/10.1007/s13347-021-00460-9

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). Available at: https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Taddeo, M., & Turilli, M. (2009). Turing's imitation game: Still an impossible challenge for all machines and some Judges–An evaluation of the 2008 Loebner Contest. *Minds and Machines, 19*(1), 145–150.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

France. (2018). *AI for Humanity: French strategy for Artificial Intelligence.* President of the French Republic. Available at: https://www.aiforhumanity.fr/en/

Franken, H. (1993). Kanttekeningen Bij Het Automatiseren Van Beschikkingen. In H. Franken, I. TH. M. Snellen, J. Smit, and A. W. Venstra *Beschikken en automatiseren* (pre-advies Vereniging voor Administratief Recht, pp. 7–50). Samsom H.D. Tjeenk Willink.

Freeman, S. (2001). Illiberal Libertarians. *Philosophy & Public Affairs, 30*(2), 105–151.

Freeman, C., & Louçã, F. (2001). *As time Goes By: From the industrial revolutions to the information revolution*. Oxford University Press.

Freeman, K., Dinnes, J., Chuchu, N., Takwoingi, Y., Bayliss, S. E., Matin, R., Jain, A., Walter, F., Williams, H., & Deeks, J. (2020). Algorithm based smartphone Apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies. *British Medical Journal, 368*(127), m127.

Frenkel, S. (2018, June 19). Microsoft employees protest work with Ice, As Tech Industries Mobilizes Over Immigration. *The New York Times*. Available at: https://www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html#:~:text=%E2%80%9CWe%20believe%20that%20Microsoft%20must,the%20chief%20executive%2C%20Satya%20Nadella.&text=The%20policy%20has%20resulted%20in,parents%2C%20raising%20a%20bipartisan%20outcry

Frenken, K., & Fuenfenschilling, L. (2020). The rise of online platforms and the Triumph of the Corporation. *Sociologica, 14*(3), 101–113.

Fridman, L. (2019, August 16). *Elon Musk: What's outside the simulation?* AI Podcast Clips. Available at: www.youtube.com/watch?v=YIVf3P3zq7g

Fukuyama, F. (1992). *The end of history and the last man*. Simon & Schuster.

Fukuyama, F. (2021). Making the Internet safe for democracy. *Journal of Democracy, 32*(2), 37–44.

Fukuyama, F., Richman, B., Goel, A., Schaake, M., Katz, R., & Melamed, D. (2021). *Report of the working group on platform scale*. Stanford University. Available at: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/platform_scale_whitepaper_-cpc-pacs.pdf

Fussell, S. (2019, August 30). Why Hong Kongers are toppling lampposts. *The Atlantic*. Available at: https://www.theatlantic.com/technology/archive/2019/08/why-hong-kong-protesters-are-cutting-down-lampposts/597145/

Future of Life Institute. (n.d.-a). *National and International AI strategies*. Future of Life Institute. Available at: https://futureoflife.org/national-international-ai-strategies/?cn-reloaded=1&cn-reloaded=1

Future of Life Institute (n.d.-b). *An open letter: Research proprieties for robust and beneficial Artificial Intelligence*, Future of Life Institute. Available at: https://futureoflife.org/ai-open-letter/

Future of Life Institute. (n.d.-c). *Asilomar AI principles*, Future of Life Institute. Available at: https://futureoflife.org/ai-principles/

Gerbrandy, A., & Custers, B. (2018). Algoritmische Besluitvorming En Het Kartelverbod. *Markt en Mededinging, 3*, 101–109. Available at: https://www.bjutijdschriften.nl/tijdschrift/marktenmededinging/2018/3/MenM_1387-6236_2018_021_003_002

Gerechtshof Amsterdam. (2021a). ECLI:NL:GHAMS:2021:1560 – Gerechtshof Amsterdam, 01-06-2021 / 200.280.852/01. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHAMS:2021:1560

Gerechtshof Amsterdam. (2021b). ECLI:NL:GHAMS:2021:392. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHAMS:2021:392

Gerritsen, J., Hamer, J., Kool, L., & Verhoef, P. (2020). Beter beschermd tegen biometrie. *Beleid en Maatschappij, 47*(4), 451–466.

Gezondheidsraad. (2006). *Betekenis Van Nanotechnologieën Voor De Gezondheid*. Gezondheidsraad.

GGE. (2019, December 13). Meeting of the High contracting parties to the convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects. [Report] Convention on Certain Conventional Weapons, UN. Retrieved from: undocs.org/CCW/MSP/2019/9

Giesea, J. (2016). It's time to embrace memetic warfare. *Defence Strategic Communications, 1*, 67–75.

Giesen, I. (2007). *Alternatieve Regelgeving En Privaatrecht*. Kluwer.

Giesen, I. (2018). (Zelf)Regulering Van En In Het Privaatrecht; Op Zoek Naar Een 'Rel'? *Nederlands Tijdschrift voor Burgerlijk Recht, 5*, 135.

Goode, L. (2018, January 19). Google CEO says AI will be more important to humanity than electricity or fire. *The Verge*. Available at: https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer

Goossens, J., Hirsch Ballin, E., & van Vugt, E. (2021). Algoritmische Beslisregels Vanuit Constitutioneel Oogpunt. Tweedeling Tussen Algemene Regels En Concrete Toepassing Onder Druk. *Tijdschrift voor constitutioneel recht, 12*(1), 4–19.

Gordon, R. (2016). *The rise and fall of American growth: The U.S. standard of living since the Civil War*. Princeton University Press.

GPT-3. (2020, September 8). A Robot wrote this entire article. Are You Scared Yet, Human? *The Guardian*. Available at: https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Graef, I., & Prüfer, J. (2021). Governance and data sharing: A law and economics proposal. *Research Policy, 50*(9), 104330.

Greene, D., Hoffman, A., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A critical assessment of the movement for ethical Artificial Intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 2122–2131).

Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.

Gross, A., Murgia, M., & Yang, Y. (2019, December 1). Chinese Tech Groups shaping un facial recognition standards. *Financial Times*. Available at: https://www.ft.com/content/c3555a3c-0d3e-11ea-b2d6-9bf4d1957a67

Hage, J. (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law, 25*(3), 255–271.

Hage, J., & Verheij, B. (1999). Rechtsinformatica: De Stand Van Zaken In De Wetenschap. In A. Oskamp and A. Lodder (reds.), *Informatietechnologie voor juristen. Handboek voor de jurist in de 21e eeuw* (pp. 65–92). Kluwer.

Hagedoorn, P. (2021). *The digital challenge for Europe*. Peter Hagedoorn.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*, 99–120.

Hall, P., & Soskice, D. (2001). *Varieties of Capitalism: The institutional foundations of comparative advantage*. Oxford University Press.

Halpern, S. (2019, April 26). The terrifying potential of the 5G network. *The New Yorker*. Available at: www.newyorker.com/news/annals-of-communications/the-terrifying-potential-of-the-5g-network

Harari, Y. N. (2019). Who will win the race for AI? *Foreign Policy Magazine*, Winter 2019. Available at: https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/

Häußermann, J., & Lütge, C. (2021). Community-In-The-Loop: Towards pluralistic value creation in AI, or—Why AI needs business ethics. *AI and Ethics, 2*, 1–22.

Hayek, F. (1994). *The road to Serfdom*. University of Chicago Press.

Heest, F. (2020, March 2020). Nederland moet meer doen om talentvlucht te voorkomen bij kunstmatige intelligentie. *ScienceGuide*. Available at: https://www.scienceguide.nl/2020/03/een-belgische-postdoc-verdient-meer-

Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society, 34*(1), 1–14.

Helbing, D., Frey, B., Gigenrenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R., & Zwitter, A. (2019). Will democracy survive big data and Artificial Intelligence? In D. Helbing (red.), *Towards digital enlightenment* (pp. 73–98). Springer.

Helwig, P. (2020). Rekenen en Rekenschap. Algoritmes en de Archiefwet. *Tijdschrift voor Toezicht, 1*, 54–59.

Hern, A. (2019a, March 12). Tim Berners-Lee On 30 years of the world wide web: 'We can get the web we want'. *The Guardian*.. Available at: https://www.theguardian.com/technology/2019/mar/12/tim-berners-lee-on-30-years-of-the-web-if-we-dream-a-little-we-can-get-the-web-we-want

Hern, A. (2019b, July 30). Cambridge Analytica did work for leave. EU, Emails confirm. *The Guardian*. Available at: https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm

Hicks, M. (2018, October 12). 'Why Tech's gender problem is nothing new', *The Guardian.* Available at: https://www.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women

High-Level Expert Group on Artificial Intelligence. (2019a). *A definition of AI: Main capabilities and scientific disciplines*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

High-Level Expert Group on Artificial Intelligence. (2019b). *Ethics Guidelines For Trustworthy AI*. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on Artificial Intelligence. (2019c). *Policy and investment recommendations for trustworthy AI*. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

High-Level Expert Group on Artificial Intelligence. (2020). *Assessment List For Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

HiiL. (n.d.). *Supporting Justice Innovations*. The Hague Institute for Innovation of Law. Available at: https://www.hiil.org/what-we-do/the-justice-accelerator/

Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics. *The University of Toronto Law Journal, 68*(5), 12–35.

Hildebrandt, M., & Gutwirth, S. (reds.). (2008). *Profiling the European Citizens*. Springer.

Hillman, J. (2019). *Infrastructure and influence: The strategic stake of foreign projects*. Center for Strategic and International Studies.

Hirsch Ballin, E. (2021). *Mensenrechten Als Ijkpunten Van Artificiële Intelligentie* (WRR Working Paper nr. 46). Wetenschappelijke Raad voor het Regeringsbeleid.

Hirsch Ballin, E., Jaspers, A., Knottnerus, A., & Vinke, H. (2021). *De Toekomst Van De Sociale Zekerheid: De Menselijke Maat In Een Solidaire Samenleving*. Boom.

Hoeks, G. (2019, April 12). Europese Wetenschappers In Verweer Tegen Braindrain Kunstmatige Intelligentie. *Het Financieele Dagblad*. Available at: https://fd.nl/economie-politiek/1297131/europese-wetenschappers-in-verweer-tegen-braindrain-kunstmatige-intelligentie-lnl1ca8zhDaO

Hoffman, S. (2019). Managing the State: Social Credit, surveillance, and the Chinese Communist Party's Plan for China. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 48–57). Air University Press.

Hoge Raad. (1921). NJ 1921, 564 (Elektriciteitsarrest) ECLI:NL:HR:1921:186, ruling 23 May 1921. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:1921:186

Hoge Raad. (2018). ECLI:NL:HR:2018:1316, ruling 17 August 2018. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:2018:1316

Holoniq. (2020, April 9). *The 2020 AI strategy landscape: 50 National Artificial Intelligence strategies shaping the future of humanity*. Available at: https://www.holoniq.com/notes/50-national-ai-strategies-the-2020-ai-strategy-landscape/

Horowitz, M., & Scharre, P. (2015). *Meaningful human control in weapon systems: A Primer* (Working paper). Center for a New American Security. Retrieved from: https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf

Horowitz, M., Allen, G., Kania, E., & Scharre, P. (2018). *Strategic competition in an era of Artificial Intelligence. Center for a New American Security*. CNAS. Available at: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000en focal=none

Houwerzijl, M. (2018). Juridische Vraagstukken Rond Arbeid In De Klusseneconomie. *Beleid En Maatschappij, 45*(2), 208–216.

Hueck, H. (2018, September 2018). Kopstuk Van Kunstmatige Intelligentie Vertrekt Naar Zweden. *Het Financieele Dagblad*. Available at: https://fd.nl/ondernemen/1267790/kopstuk-van-kunstmatige-intelligentie-vertrekt-naar-zweden

Hueck, H., & van Wijnen, J. F. (2019, October 8). Kabinet Vaag over extra budget Voor Kunstmatige Intelligentie. *Het Financieele Dagblad*. Available at: https://fd.nl/economie-politiek/1319532/kabinet-wil-meer-geld-uittrekken-voor-kunstmatige-intelligentie

Huntington, S. (1991). *The third wave: Democratization in the late 20th century*. University of Oklahoma Press.

Huys, I., van Overwalle, G., & Matthijs, G. (2011). Gene and genetic diagnostic method patent claims: A comparison under current European and US Patent Law. *European Journal of Human Genetics, 19*(10), 1104–1107.

iBestuur. (2021, May 31). Overleg Met Informateur Mariëtte Hamer Over Digitalisering. *iBestuur*. Available at: https://ibestuur.nl/nieuws/overleg-met-informateur-mariette-hamer-over-digitalisering

Ihde, D. (2010). *Embodied technics*. Automatic Press/vip.

Ipsos. (2019). *Ipsos global poll for the world economic forum shows widespread concern about Artificial Intelligence*. Available at: www.ipsos.com/sites/default/files/ct/news/documents/2019-07/wef-ai-ipsos-press-release-jul-1-2019_0.pdf

Jeffries, A. (2014, April 15). 'This Anarchist collective is demanding $3 Billion from Google. The Counterforce, Kevin Rose, and the Fire Beneath San Francisco's growing class gap. *The Verge*. Available at: https://www.theverge.com/2014/4/15/5614652/deny-the-machine

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Johnson, C. (1982). *MITI and the Japanese Miracle: The growth of Industrial policy, 1925–1975*. Stanford University Press.

Joler, V., & Crawford, K. (2018). *Anatomy of an AI-system: An anatomical case study of the Amazon echo as an artificial intelligence system made of human labor*. Available at: https://anatomyof.ai/img/ai-anatomy-map.pdf

Juma, C. (2016). *Innovation and its Enemies: Why people resist new technologies*. Oxford University Press.

Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture and Society, 39*(2), 238–258.

Kaiser, W., & Schot, J. (2014). *Writing the rules for Europe: Experts, Cartels and International Organizations*. Palgrave Macmillan.

Kamerstukken II 2013/2014 26 643, nr. 298. (2013, December 13). *Vrijheid En Veiligheid In De Digitale Samenleving. Een Agenda Voor De Toekomst*, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-298.pdf

Kamerstukken II 2015/2016 34300-X, nr. 88. (2016, March 4). *Reactie op het advies 'Autonome Wapensystemen: De Noodzaak Van Betekenisvolle Menselijke Controle' Van De Adviesraad Internationale Vraagstukken (AIV) en de Commissie van Advies inzake Volkenrechtelijke Vraagstukken (CAVV)*, Kamerbrief. Available at: https://www.tweedekamer.nl/downloads/document?id=498139ae-0353-4c0e-9678-68f3eb2da7e9&title=Reactie%20op%20het%20advies%20%E2%80%98Autonome%20wapensystemen%3A%20de%20noodzaak%20van%20betekenisvolle%20menselijke%20controle%E2%80%99%20van%20de%20Adviesraad%20Internationale%20Vraagstukken%20%28AIV%29%20en%20de%20Commissie%20van%20Advies%20inzake%20Volkenrechtelijke%20Vraagstukken%20%28CAVV%29.pdf

Kamerstukken II 2017/2018 32 761, nr. 117. (2018). *Verwerking en bescherming persoonsgegevens*, Motie, 6 June. Available at: https://zoek.officielebekendmakingen.nl/kst-32761-117.pdf

Kamerstukken II 2018/2019 2018Z22902. (2018, December 4). *Mondelinge Vragen Van Het Lid Alkaya (SP) Aan De Staatssecretaris Van Binnenlandse Zaken En Koninkrijksrelaties Over*

Het Gebrek Aan Controle Op Het Gebruik Van Algoritmen Bij De Overheid, Mondelinge vragen. Available at: https://www.tweedekamer.nl/downloads/document?id=8bcf76b1-416f-4583-832a-71e72492ee75&title=Het%20gebrek%20aan%20controle%20op%20het%20gebruik%20van%20algoritmen%20bij%20de%20overheid%20%28Binnenlandsbestuur.nl%2C%2029%20november%202018%29%20.pdf

Kamerstukken II 2018/2019, 26643, nr. 570. (2018, October 9). *Brief Van De Minister Voor Rechtsbescherming*, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-570.pdf

Kamerstukken II 2018/2019 32 761, nr. 138. (2019). *Verwerking En Bescherming Persoonsgegevens*, Motie, 25 June. Available at: https://www.tweedekamer.nl/downloads/document?id=dd94a468-cf9d-4b5d-83a8-f608b99c2fc9&title=Motie%20van%20het%20lid%20Buitenweg%20over%20het%20voorkomen%20van%20indirecte%20discriminatie%20door%20besluitvorming%20via%20algoritmen.pdf

Kamerstukken II 2018/2019 32 761, nr. 152. (2019, November 20). *Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie*, Brief regering. Available at: https://www.tweedekamer.nl/downloads/document?id=1352f4b1-696c-499a-ab6d-67c2464ad6ff&title=Waarborgen%20en%20kaders%20bij%20gebruik%20gezichtsherkenningstechnologie.pdf

Kamerstukken II 2018/2019 33009, nr. 70. (2019, Aprila 26). *Aanpak Sleuteltechnologieën*, Bijlage bij Kamerbrief. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/publicaties/2019/04/26/aanpak-sleuteltechnologieen/Bijlage+2+Kamerbrief+Missiegedreven+Topsectoren-+en+Innovatiebeleid.pdf

Kamerstukken II 2018/2019, 35 134, nr. 2. (2019, February 5). *Initiatiefnota Van Het Lid Verhoeven Over Mededinging In De Digitale Economie*, Initiatiefnota. Available at: https://zoek.officielebekendmakingen.nl/kst-35134-2.pdf

Kamerstukken II 2018/2019 35 212, nr. 2. (2019, May 29). *Initiatiefnota Van Het Lid Middendorp: Menselijke Grip Op Algoritmen*, Initiatiefnota. Available at: https://www.tweedekamer.nl/downloads/document?id=06cee835-6c90-4bdc-b067-fd28dfe2da31&title=Initiatiefnota%20.pdf

Kamerstukken II 2019/2020 26 643 and 32 761, nr. 652. (2019, December 3). Beantwoording Schriftelijke Vragen AI Bij De Politie, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-652.pdf

Kamerstukken II 2019/2020 26 643, nr. 641. (2019, October 8). *Waarborgen Tegen Risico's Van Data-Analyses Door De Overheid*, Kamerbrief. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/10/08/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid.pdf

Kamerstukken II 2019/2020 26 643, nr. 642. (2019, October 8). *AI, Publieke Waarden En Mensenrechten*, Brief regering. Available at: https://www.tweedekamer.nl/downloads/document?id=1a5131a6-0f2e-4b5e-917f-8f3e2cf0a144&title=AI%2C%20publieke%20waarden%20en%20mensenrechten.pdf

Kamerstukken II 2019/2020 26 643, nr. 652. (2019, December 3). *Artificiële Intelligentie bij de Politie*, Kamerbrief. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/12/03/tk-artificiele-intelligentie-bij-de-politie/tk-artificiele-intelligentie-bij-de-politie.pdf

Kamerstukken II 2019/2020 26 643, nr. 681. (2020, March 12). *Verslag van een Algemeen Overleg*, Verslag. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-681.pdf

Kamerstukken II 2019/2020, 26643 nr. 641. (2019, October 8). *Brief Van De Minister Voor Rechtsbescherming*, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-641.pdf

Kamerstukken II 2019/2020, 26643, nr. 672. (2020, March 13). Brief van de Ministers van Binnenlandse Zaken en Koninkrijksrelaties en voor Rechtsbescherming, Kamerbrief. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-672.html

Kamerstukken II 2019/2020 30950, nr. 206. (2020, July 1). Rassendiscriminatie, Motie. Available at: https://zoek.officielebekendmakingen.nl/kst-30950-206.html

Kamerstukken II 2020/2021, 26 643, nr. 765. (2021, June 10). Voortgangsbrief AI en Algoritmen, Kamerbrief. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2021/06/10/kamerbrief-voortgang-algoritmen-en-artificiele-intelligentie/kamerbrief-over-algoritmen-en-artificiele-intelligentie-ai.pdf

Kamerstukken II 2020/2021, 26643, nr. 779. (2021). Informatie- en Communicatietechnologie (ICT); Brief regering; Nieuwe I-strategie Rijk 2021–2025, 7 September 2021. Available at: https://zoek.officielebekendmakingen.nl/kst-26643-779.html

Kamerstukken II 2020/2021, 28362, nr. 44. (2021, April 29). Motie van de leden Klaver en Ploumen: Reikwijdte van artikel 68 Grondwet, Motie. Available at: https://zoek.officielebekendmakingen.nl/kst-28362-44.pdf

Kasparov, G. (2018). *Deep thinking. Where machine intelligence ends and human creativity begins*. John Murray Press.

Kayser-Bril, N. (2019, December 11). *At least 11 police forces use face recognition in the EU, AlgorithmWatch reveals*. AlgorithmWatch. Available at: https://algorithmwatch.org/en/face-recognition-police-europe/

Keane, J. (2009). *Life and death of democracy*. Simon & Schuster.

Keane, J. (2011). Monitory democracy. In S. Alonso (red.), *The future of representative democracy* (pp. 212–235). Cambridge University Press.

Kelly, K. (2017). *The Inevitable: Understanding the 12 technological forces that will shape our future*. Penguin.

Kerr, J. (2019). The Russian Model of digital control and its significance. In N. Wright (red.), *Artificial Intelligence, China, Russia and the global order* (pp. 62–74). Air University Press.

Keymolen, E., Noorman, M., van der Sloot, B., Koops, B.-J., Cuijpers, C., & Zhao, B. (2020). *Op Het Eerste Gezicht: Een Verkenning Van Gezichtsherkenning En Privacyrisico's In Horizontale Relaties*. Wetenschappelijk Onderzoek- en Documentatiecentrum. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2020/04/20/tk-bijlage-wodc-rapport-op-het-eerste-gezicht/tk-bijlage-

Keynes, J. M. (1930). Economic possibilities for our grandchildren. In J. M. Keynes. (1932). *Essays in Persuasion* (pp. 358–373). Harcourt Brace.

Kleinberg, J. (2018). Inherent trade-offs in algorithmic fairness. In *Abstracts of the 2018 ACM International Conference on Measurement and Modelling of Computer Systems (sigmetrics '18)* (pp. 40). ACM.

Klincewicz, M., & Lily, F. (2020). *Consequences of unexplainable machine learning for the notions of a trusted doctor and patient autonomy*. Paper presented at the 32nd International Conference on Legal Knowledge and Information Systems, Madrid, Spain, 2020.

Kohlen, J., van de Sande, M., & Cox, M. (2021). 'Rebooting' Het Mededingingsrecht – Ook Het Mededingingsrecht Ontsnapt Niet Aan De Digitale Transitie. *Markt en Mededinging, 1*, 6–14.

Kool, L., Timmer, J., Royakker, L., & van Est, R. (2017). *Opwaarderen: Borgen Van Publieke Belangen In De Digitale Samenleving*. Rathenau Instituut.

Koops, B. (2006). Should ICT regulation be technology-neutral. In B. J. Koops, C. Prins, M. Schellekens, and M. Lips (reds.), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners* (pp. 77–108). T.M.C. Asser Press.

Koops, B. J., Lips, M., Nouwt, J., Prins, C., & Schellekens, M. (2006). Should selfregulation be the starting point?. In B. J. Koops, C. Prins, M. Schellekens, en M. Lips (reds.), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners* (pp. 109–149). T.M.C. Asser Press.

Kop, M. (2020). AI and intellectual property: Towards an articulated public domain. *Texas Intellectual Property Law Journal, 28*(1), 297–341.

Kosta, E. (2020). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*. https://doi.org/10.1111/rego.12331

Krupiy, T. (2020). A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective. *Computer Law & Security Review, 38*, 1–25.

Kuijpers, K., Muntz, T., & Staal, T. (2018, October 31). Privacy? Achterhaald. *De Groene Amsterdammer*. Available at: https://www.groene.nl/artikel/privacy-achterhaald

Kulk, S. (2020). 'Platformaansprakelijkheid – Van 'Notice And Takedown' Naar Algoritmisch Toezicht', *Nederlands tijdschrift voor Europees recht*, nr. 5/6*:* 132–140.

Kulk, S., & van Deursen, S. (2020). *Juridische Aspecten Van Algoritmen Die Besluiten Nemen. Een Verkennend Onderzoek*. Wetenschappelijk Onderzoek- en Documentatiecentrum.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.

Lancieri, F., & Sakowski, P. (2020). Competition in digital markets: A review of expert reports. *Stanford Journal of Law, Business & Finance, 26*, 65.

Lawton, G. (2019, October 2). AI can predict your future behaviour with powerful new simulations. *New Scientist*. Available at: https://www.newscientist.com/article/mg24332500-800-ai-can-predict-your-future-behaviour-with-powerful-new-simulations/

Le Maire, B., Altmaier, P., & Keijzer, M. (2021). *Strengthening the Digital Markets Act and its Enforcement*. Non-paper DMA. Available at: https://www.rijksoverheid.nl/documenten/publicaties/2021/05/26/non-paper-dma

Lee, K. F. (2018). *AI Superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.

Lee, R., & Vaughan, S. (2010). Reaching down: Nanomaterials and Chemical Safety in the European Union. *Law Innovation and Technology, 2*(2), 193–217.

Leeuw, F. (2020). *Van Legal Realism naar Legal Big Data: Ontwikkelingen In Empirisch-Juridisch Onderzoek Toen, Nu En Straks*. Boom Juridisch.

Leonard, M. (red.). (2016). *Connectivity Wars: Why migration, finance and trade are the geo-economic battlegrounds of the future*. European Council on Foreign Relations.

Lessig, L. (2006). *Code and other Laws of Cyberspace, Version 2.0*. Basic Books.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Lewis, P., & Hilder, P. (2018, March 23). Leaked: Cambridge Analytica's Blueprint for Trump victory. *The Guardian*. Available at: https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory

Lewis-Kraus, G. (2016). The Great A.I. Awakening. *The New York Times Magazine*. Available at: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Libicki, M. (2019). A hacker way of warfare. In N. Wright (red.) Artificial Intelligence, China, Russia and the global order (pp. 137–142). Air University Press.

Lin, H. (2019). Escalation risk in an Artificial Intelligence-infused world. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 143–152). Air University Press.

Liu, X., Faes, L., Kale, A., Wagner, S., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J., Schmid, M., Balaskas, K., Topol, E., Bachmann, L., Keane, P., en Denniston, A. (2019). 'A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis', The Lancet Digital Health, 1: e271-e297.

Loucks, J., Hupfer, S., Jarvis, D., & Murphy, T. (2019). *Future in the balance? How countries are pursuing an AI advantage*. Deloitte Center for Technology, Media & Telecommunications. Available at: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/public-sector/lu-global-ai-survey.pdf

Luttwak, E. (1990). From Geopolitics to Geo-Economics: Logic of conflict, grammar of commerce. *The National Interest, 20*, 17–23.

Lynch, S. (2021, May 4). *Andrew Ng: Why AI is the new electricity*. Stanford Graduate School of Business. Available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

Macaulay, T. (2020, September 8). The Guardian's GPT-3-Generated article is everything wrong with AI media Hype. *The Next Web*. Available at: https://thenextweb.com/news/the-guardians-gpt-3-generated-article-is-everything-wrong-with-ai-media-hype

Marchant, G. (2011). The Growing Gap between emerging technologies and the Law. In G. E. Marchant, B. Allenby, and J. Herkert (reds.), *The growing gap between emerging technologies and legal-ethical oversight the pacing problem*. (pp. 19–33). Springer.

Marcus, G. (2018). *Deep learning: A critical appraisal*. Available at: https://arxiv.org/pdf/1801.00631.pdf?ut

Marcus, G., & Davi, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. Vintage.

Marsh, H. (2019, January 10). Can man ever build a mind? *Financial Times*. Available at: https://www.ft.com/content/2e75c04a-0f43-11e9-acdc-4d9976f1533b

Marx, K., & Engels, F. (2010 [1932]). *Die Deutsche Ideologie* (Vol. 17). Akademie Verlag.

Mastercard. (n.d.). *Mastercard rolls out Artificial Intelligence across its Global Network*. [Press Release]. Retrieved from https://newsroom.mastercard.com/press-releases/mastercard-rolls-out-artificial-intelligence-across-its-global-network/

Mateescu, A., & Ngyun, A. (2019). *Explainer: Workplace monitoring En surveillance*. Data en Society Research Institute. Available at: https://datasociety.net/wp-content/uploads/2019/02/DS_Workplace_Monitoring_Surveillance_Explainer.pdf

May, T. (1994). The Cyphernomicon: Cypherpunks FAQ and More, Version 0.666. Available at: https://hackmd.io/@jmsjsph/TheCyphernomicon

Mayor, A. (2018). *Gods and Robots: Myths, machines, and ancient dreams of technology*. Princeton University Press.

Mazzucato, M. (2014). *The entrepreneurial state: Debunking Public vs. Private Myths*. Anthem Press.

McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115–133.

McKinsey & Company. (2020). *How nine digital front-runners can lead on AI in Europe*. McKinsey & Company. Available at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/how%20nine%20digital%20frontrunners%20can%20lead%20on%20ai%20in%20europe/how-nine-digital-frontrunners-can-lead-on-ai-in-europe.pdf

McLuhan, M. (1994 [1964]). *Understanding Media. The Extensions of Man.* MIT Press.

Meijer, A., Grimmelikhuijsen, S., & Bovens, M. (2021). De Legitimiteit Van Het Algoritmisch Bestuur. *Nederlands Juristenblad, 18*, 1470–1478.

Menting, M. C. (2016). *Industry codes of conduct in a multi-layered Dutch Private Law.* Dissertation, Tilburg University.

Ministerie van Defensie. (2020). *Defensievisie 2035*. Ministerie van Defensie.

Ministerie van Economische Zaken en Klimaat. (2018). *Nederlandse Digitaliseringsstrategie*. Ministerie van Economische Zaken en Klimaat. Available at: https://www.rijksoverheid.nl/documenten/kamerstukken/2021/04/26/nederlandse-digitaliseringsstrategie-2021

Ministerie van Economische Zaken en Klimaat, Ministerie van Justitie en Veiligheid, Ministerie van Sociale Zaken en Werkgelegenheid, Ministerie van Onderwijs, Cultuur en Wetenschap, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2019). *Strategisch Actieplan voor AI* (bijlage bij Kamerstukken 26 643 en 32 761, nr. 640). Ministerie van Economische Zaken. Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie/Rapport+SAPAI.pdf

Misuraca, G., en van Noordt, C. (2020). *AI Watch – Artificial Intelligence in public services: Overview of the use and impact of AI in Public Services in the EU*. Publications Office of the European Union.

Mitchell, M. (2021). *Why AI is harder than we think*. Available at: https://arxiv.org/pdf/2104.12871.pdf

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*, 501–507.

Moerel, L., & Prins, C. (2016a). Privacy Voor De Homo Digitalis: Proeve Van Een Nieuw Toetsingskader Voor Gegevensbescherming In Het Licht Van Big Data En Internet Of Things. In *Homo Digitalis, Preadviezen 2016 Nederlandse Juristen-Vereniging 2016* (pp. 9–124). Kluwer.

Moerel, L., & Prins, C. (2016b). *Privacy for the Homo digitalis: Proposal for a new regulatory framework for data protection in the light of big data and the Internet of Things*. Wolters Kluwer. Available at: https://doi.org/10.2139/ssrn.2784123

Moore, M., & Tambini, D. (reds.). (2018). *Digital dominance: Power of Google, Amazon, Facebook, and Apple.* Oxford University Press.

Moravec, H. (1988). *Mind Children: The future of robot and human intelligence*. Harvard University Press.

Morgus, R. (2019). The spread of Russia's digital Authoritarianism. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 89–97). Air University Press.

Morozov, E. (2011). *The Net Delusion: How to Not liberate the World*. Penguin.

Morozov, E. (2013). *To save everything, click here: The Folly of technological solutionism*. Penguin.

Mozur, P. (2019, April 14). One Month, 500,000 Face Scans: How China is using A.I. to profile a minority. *The New York Times*. Available at: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

National Security Commission on Artificial Intelligence. (2021). *Final Report*. NSCAI.

Nationale Proeftuin Precisie Landbouw. (n.d.). *Precisielandbouw Voor Alle Telers*. Nationale Proeftuin Precisie Landbouw. Available at: https://www.proeftuinprecisielandbouw.nl/

NATO. (2019). *Performance Audit Reports International Board of Auditors for NATO (IBAN)*. Available at: www.nato.int/cps/en/natolive/topics_111783.htm

Nemitz, P. (2018). Constitutional Democracy and Technology in the Age of Artificial Intelligence. *Philosophical Transactions of the Royal Society A, 376*(2133), 20180089. Available at: https://doi.org/10.1098/rsta.2018.0089

Nemitz, P., & Pfeffer, M. (2020). *Prinzip mensch. Macht, Freiheit und Democratie im Zeitalter der Künstlichen Intelligenz*. Dietz.

Nilsson, N. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce Racism*. New York University Press.

NOS. (2016, June 29). *Stiekem experiment Op Facebook*. NOS. Available at: https://nos.nl/artikel/668173-stiekem-experiment-op-facebook

NOS. (2018, October 1). *AI-Special: Het Spanningsvlak Tussen Mens En Machine*. NOS. Available at: https://nos.nl/nieuwsuur/artikel/2252834-ai-special-het-spanningsvlak-tussen-mens-en-machine

Nuvolari, A. (2019). Understanding successive industrial revolutions: A "Development Block" approach. *Environmental Innovation and Societal Transitions, 32*, 33–44.

O'Neil, C. (2016). *Weapons of Math destruction: How Big Data increases inequality and threatens democracy*. Penguin.

OESO. (2018). *Private Equity Investment in Artificial Intelligence* (OECD Going Digital Policy Note). OECD Publishing. Available at: www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf

OESO. (2019). *Artificial Intelligence in society*. OECD Publishing. Available at: https://doi.org/10.1787/eedfee77-en

OESO. (n.d.). *What are the OECD principles on AI?* Available at: https://www.oecd.org/going-digital/ai/principles/

Olsthoorn, P. (2015). *25 Jaar Internet In Nederland*. Fast Moving Targets.

Onderwijsraad. (2017). *Doordacht Digitaal: Onderwijs In Het Digitale Tijdperk*. Onderwijsraad. Available at: https://www.onderwijsraad.nl/upload/documents/publicaties/volledig/Doordachtdigitaal-a.pdf

OneThird. (n.d.). *Our food loss and waste solutions*. Available at: https://onethird.io/food-waste-solutions/

Overheidsbreed Beleidsoverleg Digitale Overheid. (2018). *NL Digibeter: Agenda Digitale Overheid*. Digitale Overheid. Available at: https://www.digitaleoverheid.nl/wp-content/uploads/sites/8/2018/07/nl-digibeter-agenda-digitale-overheid.pdf

OvV. (2019). *Wie Stuurt? Verkeersveiligheid En Automatisering In Het Wegverkeer*. Onderzoeksraad voor Veiligheid. Available at: https://www.onderzoeksraad.nl/nl/media/attachment/2019/11/28/wie_stuurt_verkeersveiligheid_en_automatisering_in_het_wegverkeer.pdf

Pall, M. (2019, June 8). *How the telecommunications industry 5G strategy will use artificial intelligence to replace human intelligence: The end of mankind as we know it*. Available at: www.salzburg.gv.at/gesundheit_/Documents/5G-AI%20(002).pdf

Palmer, A. (2021, February 12). Amazon uses an App called mentor to track and discipline delivery drivers. *CNBC*. Retrieved from: cnbc.com/2021/02/12/amazon-mentor-app-tracks-and-disciplines-delivery-drivers.html

Pasquale, F. (2015). *Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

Pasquale, F. (2020). *New Laws of Robotics: Defending human expertise in the age of AI*. Harvard University Press.

Passchier, R. (2020). *Artificiële Intelligentie En De Rechtsstaat. Over Verschuivende Overheidsmacht, Big Tech En De Noodzaak Van Constitutioneel Onderhoud*. Boom.

PBL. (2017). *Mobiliteit En Elektriciteit In Het Digitale Tijdperk. Publieke Waarden Onder Spanning*. Planbureau voor de Leefomgeving. Available at: https://www.pbl.nl/sites/default/files/downloads/pbl-2017-mobiliteit-en-elektriciteit-in-het-digitale-tijdperk-1874.pdf

Pearl, J. (2019). The limitations of opaque learning machines. In J. Brockman (red.) *Possible minds: Twenty-five ways of looking at AI* (pp. 13–19). Penguin.

Perez, C. (2003). *Technological revolutions and financial capital: The dynamics of bubbles and golden ages*. Edward Elgar Publishing.

Perez, C. (2016). Capitalism, technology and a green global golden age: The role of history in helping to shape the future. In M. Jacobs and M. Mazzucato (reds.), *Rethinking Capitalism: Economics and policy for sustainable and inclusive growth* (pp. 191–217). Wiley.

Perez, C. (2017–2020). *Second Machine Age or Fifth Technological Revolution?*, blog. Available at: http://beyondthetechrevolution.com/blog/

Perez, C. C. (2019). *Invisible Women: Exposing data Bias in a world designed for Men*. Vintage.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. (2019). *The AI Index 2019 Annual Report*. Stanford University, Human-Centered AI Institute. Available at: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

Pethokoukis, J. (2019, November 25). How AI is like that other general purpose technology, electricity, blog, *AEI*. Available at: https://www.aei.org/economics/how-ai-is-like-that-other-general-purpose-technology-electricity/

Politie. (2018, May 23). *Nieuwe technologie in oude politiezaken*. Available at: https://www.politie.nl/nieuws/2018/mei/23/00-nieuwe-technologie-in-oude-politiezaken.html

Polyakova, A., & Meserole, C. (2019). *Exporting digital authoritarianism: The Russian and Chinese models* (Policy Brief, Democracy and Disorder Series). Brookings.

Poon, M. (2016). Corporate capitalism and the growing power of big data: Review essay. *Science, Technology, & Human Values, 41*, 1088–1108.

Prins, C. (2017). Politiek Profileren. *Nederlands Juristenblad, 92*(38), 2799.

Prins, C. (2018). Urgenda en Digitalisering. *Nederlands Juristenblad*, 2018/1098, 22.

Prufer, J., & Schottmüller, C. (2017). *Competing with Big Data* (TILEC Discussion Paper Nr. 2017-006, Center Discussion Paper Nr. 2017-007). Available at: https://ssrn.com/abstract=2918726 or https://doi.org/10.2139/ssrn.2918726

Pruis, M. (2017, October 11). Kennen of gekend worden. *De Groene Amsterdammer*. Available at: https://www.groene.nl/artikel/kennen-of-gekend-worden

Purtova, N. (2018). The Law of everything. Broad Concept of Personal and future of EU Data Protection Law. *Law, Innovation and Technology, 1*, 40–81.

QuTech. (2019). *Creating the Quantum Future*. *QuTech Annual Report 2019*. Available at: https://qutech.h5mag.com/annual_report_2019/

Raad van Europa. (n.d.). *CAHAI – Ad Hoc committee on Artificial Intelligence*. Raad van Europa. Available at: https://www.coe.int/en/web/artificial-intelligence/cahai

Raad van State. (2018). *Ongevraagd Advies Over De Effecten Van De Digitalisering Voor De Rechtsstatelijke Verhoudingen*. Raad van State. W04.18.0230/I. Available at: https://www.raadvanstate.nl/@112661/w04-18-0230/

Raad van State. (2020a). *Jaarverslag 2020*. Raad van State.

Raad van State. (2020b). *Ongevraagd advies over ministeriële verantwoordelijkheid*. Raad van State. Available at: https://www.raadvanstate.nl/@121396/advies-ministeriele-verantwoordelijkheid/

Raad van State (2021). *Digitalisering. Wetgeving en bestuursrechtspraak*. Raad van State. Available at: https://www.raadvanstate.nl/publish/library/13/digitalisering_wetgeving_en_bestuursrechtspraak.pdf

Ram, A. (2019, May 7). Tencent trials AI diagnosis program for Parkinson's in London. *Financial Times*. Available at: https://www.ft.com/content/183c412a-6766-11e9-9adc-98bf1d35a056

Rao, A., & Verweij, G. (2017). *Sizing the Prize: What's the real value of AI for your business and how can you capitalise?* PricewaterhouseCoopers. Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Rathenau Instituut. (2021a). *Waardevol Gebruik Van Menselijke DNA-Data. Onderzoek Naar Het Borgen Van Publieke Waarden In De Waardeketen Van DNA-Data*. Rathenau Instituut. Available at: https://www.rathenau.nl/sites/default/files/2021-05/Waardevol_gebruik_van_menselijke_DNA%20data_Rathenau_Instituut.pdf

Rathenau Instituut. (2021b, June 7). *International mobility of AI scientists*, Factsheet. Available at: https://www.rathenau.nl/en/science-figures/international-mobility-ai-scientists

Rechtbank Amsterdam. (2019). ECLI:NL:RBAMS:2019:4799, ruling 4 July 2019. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2019:4799

Rechtbank Amsterdam. (2020). ECLI:NL:RBAMS:2020:2917 – Rechtbank Amsterdam, 11-06-2020 / C/13/684665/KG ZA 20-481, ruling 11 June 2020. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2020:2917

Rechtbank Amsterdam. (2021). ECLI:NL:RBAMS:2021:1018, ruling 11 March 2021. *Rechtspraak.nl.* Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2021:1018

Rechtbank Arnhem. (2008). ECLI:NL:RBARN:2008:bd7578, ruling 18 July 2008. Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBARN:2008:BD7578

Rechtstreeks. (2019). *Algoritmes in de rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak*, Rechtstreeks 2019, nr. 2, Den Haag: Sdu. Available at: https://www.rechtspraak.nl/SiteCollectionDocuments/rechtstreeks-2019-02.pdf

Reclaim Your Face. (2021, April 16). *61 Meps Urge the EU to Ban Biometric mass surveillance!*. Available at: https://reclaimyourface.eu/61-meps-urge-eu-ban-biometric-mass-surveillance/

Reed, C. (2018). How should we regulate Artificial Intelligence? *Philosophical Transactions of the Royal Society A*, *376,* 20170360. Available at: https://doi.org/10.1098/rsta.2017.0360

Reinhold, F. (2021, April 22). Algorithmwatch's response to the European Commission's proposed regulation on Artificial Intelligence – A major step with major gaps. *AlgorithmWatch*. Available at: https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021/

Reuters. (2020, April 24). False claim: Covid-19 stands For Certification Of Vaccination Identification By Artificial Intelligence. *Reuters*. Available at: https://www.reuters.com/article/uk-factcheck-covid-name-abbreviation/false-claim-covid-19-stands-for-certification-of-vaccination-identification-by-artificial-intelligence-idUSKCN2262AS?edition-redirect=in

Rid, T. (2016). *Rise of the machines: A cybernetic history*. WW Norton & Company.

Rijksoverheid. (2019). *Minister Ollongren Komt Met Maatregelen Voor Vernieuwing Van De Democratie*. nieuwsbericht 26 June 2019. Available at: https://www.rijksoverheid.nl/actueel/nieuws/2019/06/26/minister-ollongren-komt-met-maatregelen-voor-vernieuwing-van-de-democratie

Rijksoverheid. (2021a, January 12). *Minister Dekker Maakt Kennis Met De Digitale Assistent Julo Op De Website Van Het Juridisch Loket*, video. Available at: https://www.rijksoverheid.nl/documenten/videos/2021/01/12/minister-dekker-maakt-kennis-met-de-digitale-assistent-julo-op-de-website-van-het-juridisch-loket

Rijksoverheid. (2021b). *Duitsland, Frankrijk, Nederland: 'Alle Fusies En Overnames Digitale Poortwachters Beoordelen*. nieuwsbericht 27 May 2021. Available at: https://www.rijksoverheid.nl/actueel/nieuws/2021/05/27/duitsland-frankrijk-nederland-alle-fusies-en-overnames-digitale-poortwachters-beoordelen

Rli. (2021). *Digitaal Duurzaam*. Raad voor de Leefomgeving en Infrastructuur. Available at: https://www.rli.nl/sites/default/files/rli_2021-02_digitaal_duurzaam_-_defintiief_advies.pdf

ROB. (2021). *Sturen of Gestuurd Worden? Over De Legitimiteit Van Sturen Met Data*. Raad voor het Openbaar Bestuur. Available at: https://www.raadopenbaarbestuur.nl/binaries/raad-openbaar-bestuur/documenten/publicaties/2021/05/25/advies-sturen-of-gestuurd-worden/Sturen_of_gestuurd_worden_Adviesrapport_2021_05.pdf

Robotique et Mathématiques. (2017, July 24). 'Kiva Robots: Amazon', *YouTube*. Available at: https://www.youtube.com/watch?v=ULswQgd73Tc

Rühlig, T. (2020). *Technical Standardisation, China and the future international order: A European Perspective*. Heinrich Böll Stiftung.

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A modern approach* (4th ed.). Pearson.

Rutkin, A. (2014, June 25). Even online, emotions can be contagious. *New Scientist*. Available at: https://www.newscientist.com/article/mg22229754-900-even-online-emotions-can-be-contagious/?ignored=irrelevant#.U7EHbI21au8

RVS. (2019). *Waarde(N)Volle Zorgtechnologie. Een Verkennend Advies Over De Kansen En Risico's Van Kunstmatige Intelligentie In De Zorg*. Raad voor Volksgezondheid en Samenleving.

Salian, I. (2019, April 4). AI in the sky aids feet on the ground spotting human rights violations. blog, *NVIDIA*. Available at: https://blogs.nvidia.com/blog/2019/04/04/human-rights-watch-ai-gtc/#:~:text=AI%20in%20the%20Sky%20Aids%20Feet%20on%20the%20Ground%20Spotting%20Human%20Rights%20Violations&text=In%20a%20traditional%20human%20rights,collect%20hospital%20or%20autopsy%20records

Sample, I. (2019, October 24). Human Compatible By Stuart Russell Review – AI and our future. *The Guardian*. Available at: https://www.theguardian.com/books/2019/oct/24/human-compatible-ai-problem-control-stuart-russell-review

Scassa, T. (2021, June 8). *Privacy in the Precision Economy: The Rise of AI-Enabled Workplace Surveillance during the Pandemic*. CIGI. Retrieved from: https://www.cigionline.org/articles/privacy-in-the-precision-economy-the-rise-of-ai-enabled-workplace-surveillance-during-the-pandemic/

Scharre, P. (2018). *Army of None: Autonomous weapons and the future of war*. WW Norton & Company.

Schick, N. (2020). *Deep fakes and the Infocalypse: What you urgently need to know*. Octopus Publishing Group.

Schiphol. (2019, February 18). Schiphol launches pilot for boarding by means of facial recognition. *Schiphol Nieuws*. Available at: https://news.schiphol.com/schiphol-launches-pilot-for-boarding-by-means-of-facial-recognition/

Scholvin, S., & Wigell, M. (2018). *Geo-economics as a concept and practice in international relations: Surveying the state of the art* (Working Paper nr. 102). Finnish Institute of International Affairs.

Schothorst, Y., & Verhue, D. (2018). *Nederlanders over Artificiële Intelligentie. Onderzoek naar de kennis en houding van burgers en ondernemers over Artificiële Intelligentie*, Kantar Public.

Schubert, C. (2013). How to evaluate creative destruction: Reconstructing Schumpeter's approach. *Cambridge Journal of Economics, 37*(2), 227–250.

Schulz, W., & van Hoboken, J. (2016). *Human Rights and Encryption*. UNESCO Publishing.

Schuyt, K. (2006). *Steunberen van de samenleving*. Amsterdam University Press.

Schwab, K. (2016). *The Fourth Industrial Revolution*. Random House.

Scott, C. (2007). Rethinking regulatory governance for the age of biotechnology. In H. Somsen (red.) *The regulatory challenge of biotechnology: Human genetics, food and patents* (pp. 19–35). Edward Elgar Publishing.

Select Committee on Artificial Intelligence. (2018). *AI in the uk: Ready, willing and able?* (Report of Session 2017-19, hl Paper 100). Authority of the House of Lords. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Semaan, N. (2020). Die Demokratisierung von Deepfakes: Wie technologische Entwicklung unseren gesellschaftlichen Konsens beeinflussen kann. *International Reports, 1*, 60–68.

Seo, S. (2019). *Policing The Open Road: How cars Transformed American Freedom*. Harvard University Press.

SER. (2016). *Mens En Technologie, Samen Aan Het Werk*. Sociaal-Economische Raad. Available at: https://www.ser.nl/-/media/ser/downloads/adviezen/2016/mens-technologie-publieksversie.pdf

SER. (2018). *Technologische ontwikkelingen en rol ondernemingsraad. Handreiking voor ondernemingsraden*. Sociaal-Economische Raad.

Serket. (n.d.). [Webpage]. Retrieved from https://www.serket-tech.com/

Sheikh, H. (2021). Aanbevelingen. *ESB, 106*(4801), 407–410. Available at: https://esb.nu/esb/20066595/aanbevelingen-voor-een-geo-economische-wereld

Sheikh, H., & Timmers, P. (2020, December 3). Na Trump is Het Tijd Voor 'Make Europe Great Again'. *NRC*. Available at: https://www.nrc.nl/nieuws/2020/12/03/na-trump-is-het-tijd-voor-make-europe-great-again-a4022477

Simonite, T. (2019, September 3). Behind the rise of China's Facial-Recognition Giants. *Wired*. Available at: https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/

Singer, P., & Brooking, E. (2018). *LikeWar: The weaponization of Social Media*. Mariner Books.

SkinVision. (n.d.). *Ontdek De Slimme Huidcheck*. Available at: https://www.skinvision.com/nl/

Smith, C. (1999). International collaboration in science and technology: Lessons from CERN. *European Review, 7*(1), 77–92.

Smith, C. (2013, September 18). Facebook users are uploading 350 million new photos each day. *Business Insider*. Available at: https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?IR=T

Smits, J. (2015). Wetgeving En Andere Normenstelsels: Zes Aanwijzingen Aan De Nederlandse Wetgever. *RegelMaat, 30*(5), 357–359.

Smuha, N. (2019). From A 'Race To AI To A 'Race to AI regulation': Regulatory competition for Artificial Intelligence. *Law Innovation and Technology, 13*(1), 57–84.

Smuha, N., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). *How The Eu can achieve legally trustworthy AI: A response to the European Commission's proposal for an Artificial Intelligence Act*. Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3899991_code3594902.pdf?abstractid=3899991&mirid=1

Soldatov, A., & Borogan, I. (2015). *The Red Web: The struggle between Russia's digital dictators and the new online revolutionaries*. Public Affairs.

Solove, D. (2011). *Nothing to hide. The False tradeoff between privacy and security*. Yale University Press.

Šopova, J. (2018). Audrey Azoulay: Making the most of Artificial Intelligence. *The UNESCO Courier, 3*, 36–41. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000265211

Staatscourant. (2009, January 19). Convenant tussen de Sociale Inlichtingen- en Opsporingsdienst en de Stichting Inlichtingenbureau. *Staatscourant van het Koninkrijk der Nederlanden 2009–11*. Available at: https://zoek.officielebekendmakingen.nl/stcrt-2009-791.html

Staatscourant-2017-69426. (2017). *Besluit Van De Minister-President, Minister Van Algemene Zaken, Van 22 December 2017, Nr. 3215945, Houdende Vaststelling Van De Tiende Wijziging Van De Aanwijzingen Voor De Regelgeving*. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Steijns, M. (2021). *"Van Repliek Gediend?" Een Verkenning Van Tegenmacht Vanuit Maatschappelijke Organisaties* (WRR Working Paper nr. 50). Wetenschappelijke Raad voor het Regeringsbeleid.

Stikker, M. (2019). *Het Internet is Stuk*. De Geus.

SURF. (2021, February). SURF Start Met Bouw Nieuwe Nationale Supercomputer. *SURF*. Available at: https://www.surf.nl/nieuws/surf-start-met-bouw-nieuwe-nationale-supercomputer

Svantesson, D. (2020). Is International Law ready for the (Already Ongoing) digital age? Perspectives from Private and Public International Law. In M. Busstra, W. Theeuwen, Y. Buruma, and D. Svantesson (reds.), *International Law for a Digitalised World* (Royal Netherlands Society of International Law, Collected Papers nr. 147) (pp. 113–155). T.M.C. Asser Press.

Sykes, K., & Macnaghtan, P. (2013). Responsible innovation: Opening up dialog and debate. In R. Owen, J. Bessant, and M. Heintz (reds.), *Responsible Innovation: Managing the responsible emergence of science and innovation in society* (pp. 85–107). Wiley.

Taplin, J. (2017). *Move fast and break things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us*. Pan Macmillan.

Taylor, L., Floridi, L., & van der Sloot, B. (reds.). (2016). *Group Privacy: New challenges of data technologies*. Springer.

Tegmark, M. (2017). *Life 3.0: Being Human in the age of Artificial Intelligence*. Penguin.

Tenner, E. (1997). *Why things Bite Back: Technology and the revenge of unintended consequences*. Vintage.

Thiel, P. (2009, April 13). The education of a Libertarian. *Cato Unbound.* Available at: https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian

Thomas, D. (2021, February 9). Is this Beverly Hills Cop playing Sublime's 'Santeria' to avoid being live-streamed?. *VICE*. Available at: https://www.vice.com/en/article/bvxb94/is-this-beverly-hills-cop-playing-sublimes-santeria-to-avoid-being-livestreamed

Tian, H., Wang, T., Yadong, L., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation – A review. *Information Processing in Agriculture, 7*(1), 1–19.

Tielbeke, J. (2018, May 16). Lessen van de Luddieten. *De Groene Amsterdammer*. Available at: https://www.groene.nl/artikel/lessen-van-de-luddieten

Tijdelijke Commissie Digitale Toekomst. (2020). *Update Vereist: Naar Meer Parlementaire Grip Op Digitalisering* (eindrapport). Tweede Kamer der Staten-Generaal. Available at: https://www.tweedekamer.nl/sites/default/files/atoms/files/eindrapport_tijdelijke_commissie_digitale_toekomst_tweede_kamer_der_staten-generaal.pdf

Tijdschrift voor Toezicht. (2020). Aflevering 1. *Boom Juridisch Tijdschriften*. Available at: https://www.bjutijdschriften.nl/tijdschrift/tijdschrifttoezicht/2020/1

Tilburg University. (n.d.). *Zero Hunger Lab: Met Wiskunde Minder Honger In De Winter*. Available at: https://www.tilburguniversity.edu/nl/onderzoek/impact/creating-value-data/zero-hunger-lab

Tilley, A. (2016, March 24). Alphabet's 'Moonshots' Head Astro Teller: Fear of AI and robots is wildly Overblown. *Forbes*. Available at: www.forbes.com/sites/aarontilley/2016/03/24/alphabets-moonshots-head-astro-teller-fear-of-ai-and-robots-is-wildly-overblown/?sh=7246137973bb

Timmers, P. (2019). Challenged by "Digital Soevereignty". *Journal of Internet Law, 23*(6), 12–21.

Timmers, P., & Dezeure, F. (2021). *Nederlandse Strategische Autonomie En Cybersecurity* (onderzoek in opdracht van de Cyber Security Raad). Cyber Security Raad. Available at: https://www.cybersecurityraad.nl/binaries/cybersecurityraad/documenten/rapporten/2021/02/18/onderzoeksrapport-digitale-autonomie/Onderzoeksrapport+%27Nederlandse+strategische+autonomie+en+cybersecurity%27.pdf

TNO. (2021a). *Het Technologische Ecosysteem Van AI In Nederland* (WRR Working Paper nr. 47). Wetenschappelijke Raad voor het Regeringsbeleid.

TNO. (2021b, October 4). *Veiligere Europese Wegen Dankzij Doorbraak In Truck Platooning*. Retrieved from: https://www.tno.nl/nl/over-tno/nieuws/2021/10/veiligere-europese-wegen-dankzij-doorbraak-in-truck-platooning/

Tonin, M. (2019). Artificial Intelligence: Implications for NATO's Armed Forces. *149 stctts 19 E rev. 1 fin.*

TOP500. (2021, November 16). *TOP500 list*. Retrieved from: https://www.top500.org/lists/top500/2021/11/

Topol, E. (2019). High-performance medicine: The convergence of human and Artificial Intelligence. *Nature medicine, 25*(1), 44–56.

Trajtenberg, M. (2018). *AI as the next GPT: A political-economy perspective* (NBER Working Paper Series, nr. 24245). National Bureau of Economic Research. Available at: https://www.nber.org/system/files/working_papers/w24245/w24245.pdf

Trommel, S. (2021, April 20). Europese datastrategie vereist nationale strategie. *iBestuur online*. Available at: https://ibestuur.nl/magazine/europese-datastrategie-vereist-nationale-regie

Trouw. (2016, October 17). RDW Vindt 'Autopilot' Van Tesla Misleidende Term. *Trouw*. Available at: www.trouw.nl/nieuws/rdw-vindt-autopilot-van-tesla-misleidende-term~b191a2e3/

TU Delft. (n.d.). *De Computer Als Herdershond*. Available at: https://www.tudelft.nl/stories/articles/de-computer-als-herdershond

TUC. (2020). *Technology managing people. The worker experience. [Report]*. Trade Union Congress. Retrieved from: https://www.tuc.org.uk/sites/default/files/2020-11/Technology_Managing_People_Report_2020_AW_Optimised.pdf

Turing, A. (2009 [1950]). Computing machinery and Intelligence. In R. Epstein, G. Roberts, and G. Beber (reds.), *Parsing the turing test*. Springer.

Turner, F. (2006). *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital Utopianism*. University of Chicago Press.

Tweede Kamer. (n.d.). *Vaste commissie digitale zaken*. Tweede Kamer der Staten-Generaal. Available at: https://www.tweedekamer.nl/kamerleden_en_commissies/commissies/diza

UNESCO. (n.d.). *Elaboration of a recommendation on the ethics of Artificial Intelligence*. Available at: https://en.unesco.org/artificial-intelligence/ethics

van Asselt, M., Voss, E., & Fox, T. (2010). Regulating technologioes and the uncertainty Paradox. In M. Goodwin, E. Koops, and R. Leenes (reds.), *Dimensions of technology regulation* (pp. 261–286). Wolf Legal Publishers.

van Boom, W., & Weber, F. (2017). Collectief Procederen – Ontwikkelingen In Nederland En Duitsland. *Weekblad voor Privaatrecht, Notariaat en Registratie*, wpnr 2017/7145: 291–299.

van Buchem, M., Boosman, H., Bauer, M., Kant, I., Cammel, S., & Steyerberg, E. (2021). The digital scribe in clinical practice: A scoping review and research Agenda. *NPJ Digital Medicine, 4*(57), 1–8.

Hoven, J. van den (2013). Value sensitive design and responsible innovation. In R. Owen, J. Bessant, and M. Heintz (reds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 85–107). Wiley.

van der Sloot, B., & van Schendel, S. (2020). Tien Voorstellen Voor Aanpassingen Aan Het Nederlands Procesrecht In Het Licht Van Big Data. *Computerrecht, 1*, 4–13.

van der Sloot, B., Keymolen, E., Noorman, M., Pechenizkiy, M., Weerts, H., Wagensveld, Y., Visser, B., & i.s.m. het College voor de Rechten van de Mens. (2021). *Non-Discriminatie By Design*. Tilburg University. Available at: www.tilburguniversity.edu/sites/default/files/download/01%20handreiking%20non-discriminatie%20by%20design%28NL%29.pdf

van der Vleuten, E., Oldenziel, R., & Davids, M. (2017). *Engineering the future, understanding the past: A social history of technology*. Amsterdam University Press.

Vorst, T. van der, Jelicic, N., de Vries, M. en Albers, J. (2019). *De (On)Mogelijkheden Van Kunstmatige Intelligentie In Het Onderwijs, Nr. 2018.068.1828*. Dialogic. Available at: https://www.dialogic.nl/wp-content/uploads/2019/04/Dialogic-De-onmogelijkheden-van-kunstmatige-intelligentie-in-het-onderwijs-v1.0.116.pdf

van Dijck, G. (2020a). Algoritmische Risicotaxatie Van Recidive. Over De Oxford Risk of Recidivism Tool (*OXREC*), Ongelijke Behandeling En Discriminatie In Strafzaken. *Nederlands Juristenblad, 25*, 1784–1790.

van Dijck, J. (2020b). Seeing the Forest for the trees: Visualizing platformization and its governance. *New Media & Society, 00*(0), 1–19.

van Eck, M., Zouridis, S., & Bovens, M. (2018). Algoritmische Rechtstoepassing In De Democratische Rechtsstaat. *Nederlands Juristenblad, 40*, 3008–3017.

van Ettekoven, B. J., & Prins, C. (2018). Data analysis, Artificial Intelligence and The Judiciary System. In V. Mak, E. Tjong Tjin Tai, and A. Berlee (reds.), *Research handbook in data science and law* (pp. 425–447). Edward Elgar Publishing.

van Ginkel, J., & Strijp, P. (2020). Van Beleidscyclus Naar Datacyclus. *iBestuur*. Available at: https://ibestuur.nl/podium/van-beleids-naar-datacyclus

van Gool, E., de Bruyne, J., & Fierens, M. (2021). De Regulering Van Artificiële Intelligentie (Deel 2). Een Analyse Van Buitencontractuele Aansprakelijkheid. *Rechtskundig Weekblad, 84*(26), 1003–1024.

van Heukelom-Verhage, S. (2020). Maatwerk Bieden In Een Gedigitaliseerde En Datagedreven Samenleving #Hoedan?. In L. van den Berge, M. Vermaat, M. Lurks, N. van Renssen en S. van Heukelom-Verhage (reds.), *Maatwerk in het bestuursrecht.* Boom.

van Noort, W. (2018, August 27). *Nederland Kampt Met Braindrain In Artificiële Intelligentie*. NRC. Available at: https://www.nrc.nl/nieuws/2018/08/27/nederland-kampt-met-ai-braindrain-a1614393

van Roy, V., Rossetti, F., Perset, K., en Galindo-Romero L. (2021). *AI watch – National strategies on Artificial Intelligence: A European Perspective* (EUR 30745): Publications Office of the European Union.

van Veenstra, A. F., Djafari, S., Grommé, F., Kotterink, B., & Baartmans, R. (2019). *Quick scan AI in de Publieke Dienstverlening*. TNO. Available at: www.rijksoverheid.nl/documenten/rapporten/2019/04/08/quick-scan-in-de-publiekedienstverlening

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*, 1–6.

Veale, M. (2020). A critical take on the policy recommendations of the EU high-level expert group on Artificial Intelligence. *European Journal of Risk Regulation, 11*, 1–10.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International, 22*(4), 97–112.

Verbeek, P. P. (2014). *Op De Vleugels Van Icarus: Hoe Techniek En Moraal Met Elkaar Meebewegen*. Lemniscaat.

Verhagen, L. (2021, April 21). Europa Komt Als Eerste Ter Wereld Met Regels Voor Kunstmatige Intelligentie. 'Dit Wordt Een Feest Voor Juristen'. *de Volkskrant*. Available at: https://www.volkskrant.nl/nieuws-achtergrond/europa-komt-als-eerste-ter-wereld-met-regels-voor-kunstmatige-intelligentie-dit-wordt-een-feest-voor-juristen~b2509f56/

Verhey, L. & Verheij, N. (2005). De Macht Van De Marktmeesters: Markttoezicht In Constitutioneel Perspectief. In A. A. Rossum, L. F. M. Verhey, and N. Verheij (red.) *Toezicht* (Vol. 135, Handelingen der Nederlandse Juristen-Vereniging) (pp. 135–332), Kluwer.

Vermaas, P., Nas, D., Vandersypen, L., & Elkouss Coronas, D. (2019). *Quantum Internet: The Internet's next big step*. TU Delft.

Vetzo, M., Gerards, J., & Nehmelman, R. (2018). *Algoritmes en Grondrechten*. Boom Juridisch.

Villani, C. (2018). *For a meaningful Artificial Intelligence: Towards a French and European strategy*. Available at: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. In *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace* (NASA Conference Publication 10129) (pp. 11–22). NASA Lewis Research Center. Available at: https://www.researchgate.net/profile/Carol-Stoker-2/publication/234229828_Telepresence_in_the_human_exploration_of_Mars_Field_studies_in_analog_environments/links/554bb5600cf29752ee7e78f8/Telepresence-in-the-human-exploration-of-Mars-Field-studies-in-analog-environments.pdf#page=23

von Neumann, J. (2012 [1958]). *The Computer and the Brain.* Yale University Press.

Voorzieningenrechter Arnhem. (2008, July 18). Radboud Universiteit Mag Artikel MIFARE Classic Chip Publiceren. *persbericht*. Rechtbank Arnhem. Available at: https://www.sos.cs.ru.nl/applications/rfid/pressrelease-courtdecision.nl.html

VSNU. (2021). *Advies Publieke Waarden Voor Het Onderwijs*. VSNU. Available at: https://vsnu.nl/files/documenten/Domeinen/Onderwijs/Advies_werkgroep_publieke_waarden_onderwijs.pdf

Waag. (2020). *Algoritme: De Mens In De Machine – Casusonderzoek Toepasbaarheid Van Conceptrichtlijnen Voor Algoritmen*. Waag.

Waardenburg, L., Sergeeva, A., & Huysman, M. (2020). Predictive policing Ontcijferd: Een Etnografie Van Het 'Criminaliteits Anticipatie Systeem' in De Praktijk. In J. Janssens, W. Broer, M. Crispel, and R. Salet (reds) *Informatiegestuurde politie* (Cahiers Politiestudies 54) (pp. 69–88). Gompel & Svacina.

Waarlo, N., & Verhagen, L. (2020, March 27). De Stand Van Gezichtsherkenning In Nederland. *De Volkskrant*. Available at: https://www.volkskrant.nl/kijkverder/v/2020/de-stand-van-gezichtsherkenning-in-nederland%7Ev91028/?referrer=https%3A%2F%2Fwww.google.com%2F

Wade, R. (2018). *Governing the market*. Princeton University Press.

Walch, K. (2020, February 9). Why the race for AI dominance is more global than you think. *Forbes*. Available at: https://www.forbes.com/sites/cognitiveworld/2020/02/09/why-the-race-for-ai-dominance-is-more-global-than-you-think/?sh=3a34ad2b121f

Waldrop, M. (2019). News feature: What are the limits of deep learning? *Proceedings of the National Academy of Sciences, 116*(4), 1074–1077.

Wallace, R. (2021). 'The names have changed, but the game's the same': Artificial Intelligence and racial policy in the USA. *AI and Ethics, 389*, 1–6.

Wallach, W. (2015). *A dangerous master. How to keep technology from Slipping beyond our control*. Basic Books.

Warzel, C. (2018, February 11). Believable: The terrifying future of fake news. *Buzzfeed News*. Available at: buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news

Weber, V. (2019). Understanding the global ramifications of China's information-control model. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 76–80). Air University Press.

Weinberger, S. (2019). *The imagineers of war: The untold history of DARPA, The Pentagon agency that changed the world*. Vintage.

Weiser, M. (1991). The computer for the 21st century. *IEEE Pervasive Computer, 1*(1), 19–25.

Went, R., Kremer, M., en Knottnerus, A. (red.) (2015). *De Robot De Baas. De Toekomst Van Werk In Het Tweede Machinetijdperk* (WRR-Verkenning nr. 31). Amsterdam University Press.

Whitelaw, S., Mamas, M., Topol, E., & Van Spallm, H. (2021). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health, 2*(8), e435–e440.

Wiener, N. (1964). *God and Golem, Inc.: A comment on certain points where cybernetics impinges on religion*. MIT Press.

Wiener (2019 [1965]). Cybernetics: Or control and communication in the animal and the machine, : MIT Press.

Wilson, H., Daugherty, P., & Davenport, C. (2019, January 14). The future of AI will be about less data, not more. *Harvard Business Review*. Available at: https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more

Winner, I. (1983). Techne and Politeia: The technical constitution of society. In P. Durbin en F. Rapp (reds.), *Philosophy and Technology* (Boston Studies in the Philosophy of Science) (pp. 97–111). Springer.

WIPO. (2019). *WIPO Technology Trends 2019: Artificial Intelligence*. World Intellectual Property Organization.

Wittgenstein, L. (1984). *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Suhrkamp.

Wojciki, S. (2020, February 14). YouTube at 15: My personal journey and the Road Ahead, *blog*. Available at: https://blog.youtube/news-and-events/youtube-at-15-my-personal-journey

Wolswinkel, J. (2019, October). Het Algoritme Van De Afdeling: De Realiteit Van Complex Bestuursrecht. *Ars Aequi*, 776–785. Available at: https://pure.uvt.nl/ws/portalfiles/portal/31738610/AA20190776.pdf

Wolswinkel, J. (2020). *Willekeur of Algoritme? Laveren Tussen Analoog En Digitaal Bestuursrecht*. Tilburg University.

Wouda, F., & Hutink, H. (2019). *Artificial Intelligence In De Zorg: Begrippen, Praktijkvoorbeelden En Vraagstukken*. Nictiz. Available at: https://www.nictiz.nl/wp-content/uploads/Rapport_artificial_intelligence_in_de_zorg.pdf

Wright, G. (2000). Review: 'General purpose technologies and economic growth' (Helpman, 1998). *Journal of Economic Literature, 38*(1), 161–162.

Wright, N. (2019a). Global Competition. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 35–41). Air University Press.

Wright, N. (2019b). Artificial intelligence and domestic regimes: Digital authoritarian, digital hybrid, and digital democracy. In N. Wright (red.) *Artificial Intelligence, China, Russia and the global order* (pp. 21–34). Air University Press.

WRR. (2008). *Onzekere Veiligheid. Verantwoordelijkheid Rond Fysieke Veiligheid*. Amsterdam University Press.

WRR. (2011). *iOverheid*. Amsterdam University Press.

WRR. (2013). *Naar Een Lerende Economie*. Wetenschappelijke Raad voor het Regeringsbeleid.

WRR. (2015). *De Publieke Kern Van Het Internet. Naar Een Buitenlands Internetbeleid*. Amsterdam University Press.

WRR. (2016). *Big data In Een Vrije En Veilige Samenleving*. Amsterdam University Press.

WRR. (2017). *Veiligheid In Een Wereld Van Verbindingen. Een Strategische Visie Op Het Defensiebeleid*. Wetenschappelijke Raad voor het Regeringsbeleid.

WRR. (2019). *Voorbereiden Op Digitale Ontwrichting*. Wetenschappelijke Raad voor het Regeringsbeleid.

WRR. (2020). *Het Betere Werk. De Nieuwe Maatschappelijke Opdracht*. Wetenschappelijke Raad voor het Regeringsbeleid.

Wu, T. (2020). *The Curse of bigness. How corporate giants came to rule the world*. Atlantic Books.

Yeung, K., & Lodge, M. (reds.). (2019). *Algorithmic regulation*. Oxford University Press.

Yu, K., Beam, A., & Kohane, I. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering, 2*(10), 719–731.

Zarkadakis, G. (2015). *In our own image: Will artificial intelligence save or destroy us?* Ebury Publishing.

Zhang, W. W. (2012). *The China wave: Rise of a Civilizational State*. World Century Publishing Cooperation.

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford, Center for the Governance of AI, Future of Humanity Institute. Available at: https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *The AI index 2021 annual report*. Stanford University, Human-Centered AI Institute. Available at: https://arxiv.org/ftp/arxiv/papers/2103/2103.06312.pdf

Zielonka, J. (2008). Europe as a global actor: Empire by example? *International Affairs, 84*(3), 471–484.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Fataigh, R., Irion, K., Dobber, T., Bodo, B., & De Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review, 14*(1), 82–96.