HANDBOOK OF COMPUTATIONAL SOCIAL SCIENCE, VOLUME 1

Theory, Case Studies and Ethics

Edited by Uwe Engel, Anabel Quan-Haase, Sunny Xun Liu, and Lars Lyberg

First published 2022

ISBN: 978-0-367-45653-5 (hbk)
ISBN: 978-0-367-45652-8 (pbk)
ISBN: 978-1-003-02458-3 (ebk)

21

EFFECTIVE FIGHT AGAINST EXTREMIST DISCOURSE ONLINE

The case of ISIS's propaganda

Séraphin Alava and Rasha Nagem

DOI: 10.4324/9781003024583-24

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

21

EFFECTIVE FIGHT AGAINST EXTREMIST DISCOURSE ONLINE

The case of ISIS's propaganda

Séraphin Alava and Rasha Nagem

Context of ISIS's online terrorist propaganda

The Internet has become a key element of our daily life, having an impact on people's entertainment, learning and even socialization. Extremist groups are taking advantage of this dynamic and reaching out to people on the Internet and seeking to propagate their ideologies: hatred and violence. They do not hesitate to identify potential recruits and show real skill in adapting their message to the intended targets. By bringing together researchers in the fields of sociology, linguistics, artificial intelligence and machine learning, we have developed research aimed at building up new approaches for the exploration of cyberspace. Our project is based on a multidisciplinary approach targeted at analyzing extremist content on Internet for French-speaking communities. The purpose of our work is twofold. First of all, we explore extremist content online reflected by indoctrination sources as well as messages and chat exchanges in French-speaking cyberspace. This exploration is carried out at a crossing point of sociological and linguistic perspectives. The aim is to acquire knowledge about the characteristics of this type of content and its etiology and to shed light on the underlying social models and highlight its specific forms of communication. The second objective is to design innovative solutions for the characterization of content. We use models and resources created by researchers in sociology to characterize content according to lexical, discursive and semantic criteria. The project adopts supervised adaptive learning, which is a hybrid paradigm that augments learning algorithms with adaptation mechanisms, capable of acquiring new descriptors, in order to recognize and follow the evolution of concepts. The project thus implements advanced functionalities for a systematic exploration of cyberspace and aims to overcome the limitations of current approaches, which are often ill-suited to monitoring developments in the field. In terms of practical application, we want to design help for users concerned with the fight against the propagation of extremist ideologies online: researchers in sociology will be able to better understand new phenomena, and intelligence analysts and homeland security experts will be able to explore cyberspace more precisely.

It is essential to produce results that go beyond the state of the art and "technological bricks", the use of which will highlight new discoveries on societal and technological challenges at the border between real and virtual environments. Our praxeological objective is to produce

DOI: 10.4324/9781003024583-24

effective adaptive tools for the security forces by taking into account changes in indoctrination discourse and in the mobility of groups and propagators. Our scientific objective is to build bridges between the sociology of deviance, criminology specializing in terrorism and information science which explores networks and virtual worlds, as well as digital sciences, artificial intelligence and mathematics, to equip our societies so they can detect and counter terrorist and extremist speeches.

Interdisciplinary problems to be solved

Methods of extremist recruitment on the Internet

In this area, the ISIS terrorist group, like other Islamist terrorist groups, has particularly innovated in cyberspace. From 2012 to 2019, these groups built a real digital propaganda strategy targeting four dimensions that are specific to military propaganda: disseminating an ideological corpus, attacking the enemy's discourse, wreaking havoc by discrediting objective information and convincing and enlisting future recruits. Obviously, the role of the Internet for terrorist groups is not just about propaganda. The groups must also coordinate, attack and capitalize on the action of terrorist organizations. In this chapter, we focus on propaganda by examining – in the case of jihadist Islamist groups – our technical ability to identify and characterize these speeches online.

Propaganda is a concept designating a set of persuasion techniques used to propagate with all available means an idea, an opinion, an ideology or a doctrine and stimulate the adoption of behaviors within a target audience. In the case of ISIS or Al-Qaeda, we can easily speak of structuring a global and strategic communication network.

ISIS has developed a global and strategic communication policy using all communication techniques. Their aim is to broadcast in information vacuums where public discourse is redundant and not critical. Conspiracy or disruptive speeches are "media projectiles" which reach the most varied audiences with the same degree of precision and leave an informational vacuum in their wake. Twitter and Snapchat become launch points for rumors and disinformation which will then be passed on in social media and finally spread out on encrypted networks. This marketing technique, starting with an appealing product and transforming into loyalty, is all the easier as the contents of the messages combine cultural, linguistic, religious and often economic discourse in order to cover all the areas of widening rifts between youth and society.

Another pressure technique used to cover the informational field is directly related to the meaning of words. Extremist communication firstly seeks to saturate – by disinformation, misuse, repetitions of meaning – the sense of certain words or concepts (family, community, fraternity, violence) in order to break the emotional barrier included in these words and then cause a change of direction in the meaning and gradually impose a shift in the meaning. This shift in the meaning makes the reading of reality impossible outside the framework of extremist assignments of meaning (Jewish, unbelievers, caricature, the West, etc.). Finally, terrorist groups also use their communications by taking into account the media coverage of their actions. This is the terrorist second breath that Regis Debré refers to. The blast of an attack provokes a media frenzy. Media from all over the world will relay the information supplied by ISIS. The target to be recruited is constantly mobbed by articles with catchy titles which, slowly but most likely,

undermine the resilience and cohesion of our societies. We have grown accustomed to information warfare in which we've become actors unintentionally.

Digital radicality, process of radicalization?

The individualization of radicalization is achieved through terrorism, and it implicitly betrays the prevalence of a subjectivist and psychologizing approach (Guibet Lafaye & Rapin, 2017). Radicalization and terrorism are linked, the radicalization becomes a "breeding ground for terrorism". In radicalization, it is no longer about hardening an ideology but about describing an individual phenomenon, a turnaround that begins in dialogue and radical conversation.

It is no longer a collective and social movement which is radicalized, but the target of radicalization is an individual (often considered lost) who is in a need of psychological support and deradicalization. Khosrokhavar (2014) proposed a definition of radicalization where violence was inevitably inherent. Indeed, he sees radicalization as a process in which an individual or a group takes a violent form of action, which is directly linked to an extremist ideology with a political, social or religious content, contesting the established order at the political, social or cultural level.

In fact, regarding radicalization as an individual and violent phenomenon only brings about a biased analysis of the process, and thus generates ineffective counter-discourse. Thus, conversion, allegiance and radicalization should not be confused (Salazar, 2015; Filiu, 2015; Haddad, 2015). Therefore, radicalization would be defined as a process of desocialization in relation to a given ideology and of re-socialization in the form of adherence to a radical political line (Guibet Lafaye & Rapin, 2017).

What is a discourse?

In order to recognize and gather elements of ultra-right discourse, it is important to characterize these groups but also to be familiar with the elements that reveal the discourse's structure. In everyday language, the term "discourse" refers to an oral production performed in front of an assembly or intended for an audience. In linguistics, the term encompasses all of a person's written or oral linguistic productions. The word "discourse" becomes, then, very broad and seems close to the definition of the dictionary of the French language, "littré", in which the discourse encompasses all the words used by a person.

In our case, these definitions are not effective because it is not a question of knowing and characterizing all the productions of an ultra-right group but of identifying the types of discourse that have an objective or an impact in the radicalization of an individual. In this sense, we follow the analysis of Vincent D. (2005), for whom language productions form a coherent whole, interpretable only by the superposition of multiple layers of analysis; a laminate made by using production methods that are repetitive and unique at the same time, each interaction being seen as a structured and structuring social activity.

Our mission isn't to recognize the discourse but the discourses of the extreme right, not only by distinguishing them by their explicit or implicit contents but also by their strategic wills, their rhetorical qualities and their own typologies.

Therefore, identifying, collecting and characterizing far-right discourse within cyberspace requires a triple identification (content, typologies, rhetoric).

Building a base of expertise: web-crawlers and characterization of extremist texts

One of the main concerns linked to the misuse of social media is the dissemination of content related to extremist attitudes and ideologies, such as online propaganda, hate speech or radicalizing content, disseminated on social media platforms to supposedly pollute these platforms with content intended to influence public opinion, harm other people or recruit (Lorenzi & Moïse, 2018).

We know today that the goals of the discursive presence of extremist ideologies are complementary. Spreading the ideology and providing arguments that can be used by followers is not the only goal. The propaganda of these ideologies is reinforced by the dissemination of false information, relying on a strategic use of conspiracy theories. All these typologies of discourse also aim at the isolation of the individual to be recruited in order to produce an intimate space for deploying a radical conversation that we have already analyzed (Alava, Najjar, & Hussein, 2019). Online propaganda campaigns have attracted the attention of the research community (Chatfield, Reddick, & Brajawidagda, 2015) (Scanlon & Gerber, 2014), (Allendorfer & Herring, 2015), and numerous independent studies suggest that the social media played a central role in the rise of ISIS (Ferrara, Wang, Varol, Flammini, & Galstyan, 2016) or at least that it benefited from the use of social media for propaganda and recruitment purposes (Berger & Morgan, 2015).

To establish our baselines, which will serve as a basis for indexing and then as a learning source for artificial intelligence, we have chosen to focus on a single jihadist group (ISIS) and to take extracts from texts not in relation to their contents but to three levels of discourse. Indeed, there are many variations between the generic discourse given by a politician, a political scientist and an intellectual; the discourse propagated by a journalist, an intellectual and a media man; and the integrated discourse carried by a grassroots activist and a follower of the movements, and it is very useful to understand these discourses. We have therefore defined three levels of discourse:

1 *Generic discourses:* We identify on the Internet, the founding discourses of ISIS ideologies in order to know how to distinguish the differences between the discourse that founds a concept or an analysis and will then be appropriated or propagated. For this, we have taken into account the speeches of Abou Bakr al-Baghdadi and written references of ISIS's ideologists in their online journals.
2 *Propagated discourses:* We then identify the discourses coming from the media relays, the followers of ideas and the associated intellectuals. These discourses have a desire for impact, adherence and greater statement, and studying these variations is useful to better characterize ideas and rhetoric. To do this, we extracted all the text from recruitment videos by Niçois Omar Diaby, alias Omar Omsen. We used seven videos from the 19HH channel.
3 *Appropriated discourses:* We refer here to statements made by Internet users, which are the targets aimed at by this propaganda. Interviews were conducted with 32 French radicalized youngsters who agreed to discuss their ideologies and to explain these ideas to us.

Tagging extremist sentences and artificial intelligence learning

Overall, our research project takes into account five technological stages. The characterization of jihadist discourses is taken into account by sociologists. The setting up of the base and its analysis requires cooperation between computer sciences and language sciences. Indexation

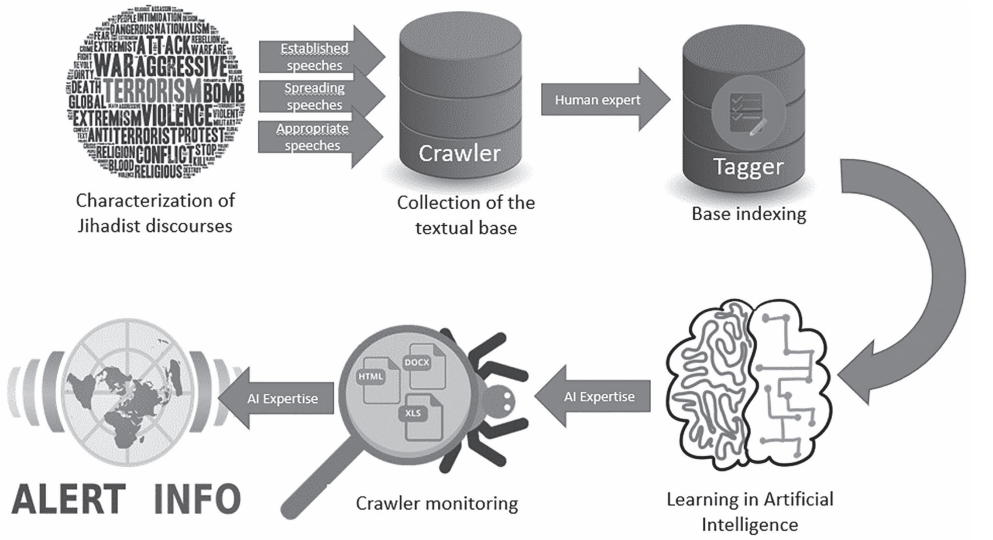

Figure 21.1 General technological graphic

of the database requires the help of human experts, sociologists or Islamologists. AI learning with the indexed database is performed by AI specialists. The adaptation of the crawler and its interaction with the semantic engine requires cooperation between computer engineers and AI specialists. It is obvious that characterization of alerts requires human expertise. A human feedback to the AI analysis engine is something to consider in order to improve the system.

Methodology for setting up the starting base

The main objective of our technological cooperation was to respond to several fundamental technological problems allowing us to offer the security service relevant and effective tools in the fight against online extremism. Our first problem to be addressed was the identification of the different registers of extremist discourse by relying on identifiable contents in this discourse and at the same time on the methods of dissemination and the linguistic components of these discourses. This issue required a substantial dialogue among sociologists capable of recognizing the extremist groups, identifying the involved media and the names of the most important propagators. The work carried out within the framework of this study yielded a base of 19,612 sentences indexed with three criteria (radical, non-radical, out of context).

To achieve this, we worked with two tools: an internet robot (crawler) able to search a database of 24,000 sentences using provided samples (lists of words, profiles to follow, list of hashtags) and an indexing tool that allows human experts to characterize these sentences.

The Database Tagger is a computer application used to automate and facilitate the process of assigning tags to a dataset. In the context of our research, the content of these datasets may be texts or images, which are tagged with the desired tags associated with the analysis tools. The Database Tagger allows to create the datasets necessary for the development of the text analysis module. Furthermore, it will help to increase these datasets in the future by adding and tagging more texts coming from "Alert" results, thus improving the performance and reliability of the developed analysis tools.

Table 21.1 Breakdown of the number of items retained for the database

	Total	*Out of context*	*Radical*	*Non radical*
Magazines	6697	1155	3116	2426
Social media	8119	4273	2804	1042
Interviews	2201	270	1191	740
Videos	2078	432	1074	572
Various texts	517	59	226	232
	19,612	**6189**	**8411**	**5012**

Network monitor client (proxy) methodology

A network monitoring client was developed as part of the project with the aim of collecting real-time information on network traffic in explicitly monitored locations. The proxy captures traffic from a monitored network to send it to the crawler for further analysis. Along with the captured traffic, additional metadata such as the proxy ID or an ID of the monitored machine is also sent.

The network monitor client has the following capabilities:

- Real-time network analysis, request by request;
- Ability to read HTTP connections if the used browser has been correctly configured with the root certificate;
- As the communication is asynchronous with the crawler, the surveillance activity therefore has a minimum latency penalty that could potentially raise suspicion about the person being monitored.

The network monitor client consists of two modules [Squid proxy and Internet Content Adaptation Protocol (ICAP) server], which interact with each other via the ICAP protocol. The proxy performs the following actions in sequential order:

- Intercepting the request from the monitored client(s) and passing the same request to the server, acting as the original client;
- Reading the response from the server and sending it to the ICAP server;
- Re-encrypting the server's response with a custom certificate and sending it to the client.

The ICAP server performs the following actions:

- Receiving responses from the proxy via the ICAP protocol;
- Adding some additional metadata and grouping the data into a JSON object;
- Asynchronous sending of the message via HTTP to the crawler.

Crawler assistant methodology

The objective of the crawler assistant is to generate likely user interactions based on previous context. The crawler assistant works with a model that is built with data extracted from user interactions. Those interactions must be provided during a training phase.

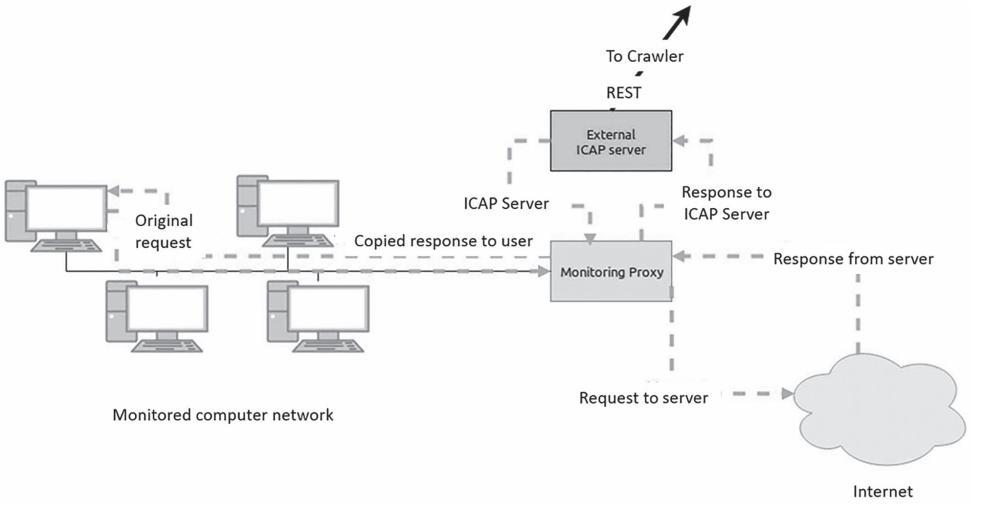

Figure 21.2 Network monitoring customer overview (Practicies report 7, 2019)

Figure 21.3 shows the architecture of the crawler assistant. The crawler assistant has a set-up phase and a normal deployment phase. During the set-up phase, the data preparation module takes the training corpus and stores it in a Knowledge database. This step is only executed if the configuration system does not point to an existing knowledge base (e.g., typically the first time the system is run). After this step, the system is ready to receive queries (normal deployment phase). The response generation takes the input query of the user and browses efficiently the Knowledge database in order to find a suitable response. This is performed every time the user queries the system using the REST API.

The Knowledge database is read-only. Therefore, it cannot be updated with new samples and a new database must be built from scratch instead.

The provided context must be a sentence (although it will depend slightly on the training data, the algorithm employed works much better with short texts or sentences). The output has no limitations in size.

To build the training dataset, we used the Dark Web Forums dataset2. We took all the post entries of the forums to build the Knowledge databases in English and French separately. The original post entries are split in sentences. After that, they are clustered by their shallow similarity. The most frequent sentences (biggest clusters) are selected, and the precedent sentences of those that are part of the clusters are employed as the training corpus of the knowledge database.

Final monitoring tool architecture

This section focuses on the technical details of the final monitoring tool, the architecture of which is described in Figure 21.1. This architecture is divided in four main parts. The first one is the ingest part, where the data from the crawler is obtained and transformed to be processed across the system. The second main part, which could be called the manager part, is where the adapted data are received, processed and sent to the cores to be analyzed. The third part is the combination of the core modules, which analyses the input data, generates the results and sends

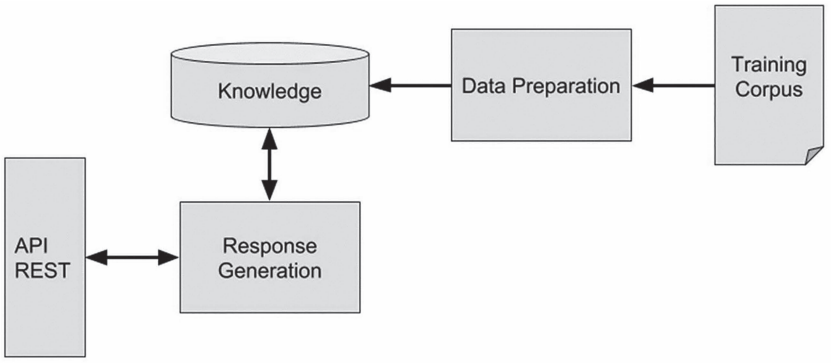

Figure 21.3 Architecture of the crawler assistant (Practicies report 7, 2019)

them back to the manager sector of the platform in order to update the stored data. And, finally, the last part comprises the modules responsible for recovering the data objects with their respective results and presenting them to the graphical user interface.

Evaluation of machine and human efficiency: methodological reflections

The more advanced digital technology gets, the more terrorism will adapt accordingly. In the security battle fought currently by various countries with terrorist groups, the terrorists are always one step ahead. The Visible Web, the Dark Web, the Web of Things – terrorist groups are gradually adapting to technologies. Censorship, de-listing, financial fines against Internet companies – all these actions will have reduced effectiveness in the long term. There is an urgent need for understanding how extremist recruiting works and describing the modalities of targeting young people sensitive to extremist ideas. There is an urgent need for characterizing the content, topics and terrorist arguments, because even if we thought at one point that radicalized young people were victims, it is clear that a part of the jihadist changeover is a commitment that starts at the communicational level, then psychological and finally relational with the propagators of radical ideas.

Our research confirms American work on the diversity, quality and quantity of ISIS's propaganda. These speeches, videos, chats, tweets and posts are as many beacons for the propagators of extremist ideas and as many decoys to attract young people. Between these texts and the radical conversation that will lead the young person to engage, there are structured steps that the ISIS organization deploys with a wide variety of tailor-made online contributions to bring their targets into their bosom (Berger & Morgan, 2015).

Sometimes referred to as "touching up", these contributions are carried out by small teams of prolific social media users who grab the attention of potential recruits in order to shape their worldview and encourage direct action in favor of the Islamic State. We need to understand these contributions. With the help of digital engineering and robotics, thanks to the promising tools of interrelation graphs on the Net and with artificial intelligence, we can catch up with terrorist groups and identify sensitive cognitive phases to alert the machine to detect violent extremism in its present form. Traditional approaches to content analysis prove limited in highly heterogeneous and dynamic environments such as online media. Hybrid mechanisms, using

supervised learning cores capable of adapting to changing domains through the integration of asynchronously inferred knowledge, are better suited for the detection of extremist content on the Internet. Our goals for ongoing projects are to develop artificial intelligence methods to analyze extremist content, messages and conversations on the Internet. The analyzed content will be retrieved from sources of indoctrination but also from exchanges on digital platforms. It is important to develop methods of in-depth characterization of online content in order to produce a description that is rich at the lexical, terminological and semantic level. The characterization will also highlight domain-specific concepts (radicalization, extreme right, violence, threats) and subjective engagement (support, rejection, preference, disagreement) expressed by users (Battistelli, Bruneau, & Dragos, 2020). Within the FLYER research project, we want to build new approaches for adaptive supervised learning. It is a hybrid paradigm of artificial intelligence, augmenting learning algorithms with adaptation mechanisms that make them capable of acquiring new descriptors in order to cope with the evolution of the field (Dragos, Battistelli, & Kellodjoue, 2020).

The scope of the project is large, but the stakes are high. The Internet war has started, and terrorist groups are ahead of us. Therefore, we must describe, explain, understand and above all anticipate.

Useful research projects:

- PRACTICIES Europa H2020: Partnership Against Violent Radicalisation in Cities, https://practicies.org/home-en-gb/
- SAFFRON Europa FSI: Semantic Analysis against Foreign Fighters Recruitment On-line Network, www.saffron-project.eu/en/home/
- RISTRACK EUOPA Justice Tracking tool based on social media for risk assessment on radicalisation, www.risk-track.eu/en/
- FLYER ANR DEFENSE France Artificial intelligence for extremist content analysis, https://anr.fr/en/funded-projects-and-impact/funded-projects/project/funded/project/b2d9d3668f92a3b9fbbf7866072501ef-ab458e3cd0/?tx_anrprojects_funded%5Bcontroller%5D=Funded&cHash=c1b49f4f0de62f426c2961b334b57110

References

Alava, S., Najjar, N., & Hussein, H. (2017). *Study of radicalization processes within social networks: Role of conspiracy arguments and rupture discourses*, *Quaderni* [En ligne], 94, posted online October 5, 2019. Retrieved December 18, 2020, from http://journals.openedition.org/quaderni/1106

Allendorfer, W. H., & Herring, S. C. (2015). ISIS vs. the US government: A war of online video propaganda. *First Monday*, *20*(12).

Battistelli D, Bruneau C, Dragos V, (2020). Building a formal model for hate detection in French corpora: 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. *Procedia Computer Science*, *176*(2020), 2358–2365.

Berger, J. M., & Morgan, J. (2015). The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter. *The Brookings Project on US Relations with the Islamic World*, *3*(20), 4–1.

Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015, May). Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided Twitter networks. In *Proceedings of the 16th annual international conference on digital government research* (pp. 239–249). ACM.

Dragos, V., Battistelli, D., & Kellodjoue, E. (2020). A formal representation of appraisal categories for social data analysis: 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. *Procedia Computer Science*, *176*(2020), 928–937.

Ferrara, E., Wang, W. Q., Varol, O., Flammini, A., & Galstyan, A. (2016, November). Predicting online extremism, content adopters, and interaction reciprocity. In *International conference on social informatics* (pp. 22–39). Cham: Springer.

Filiu, J. (2015). Barbarie jihadiste et terreur médiatique. *Cités*, *61*, 27–38.
Guibet Lafaye, C., & Rapin, A. (2017). La "radicalization": Individualisation et dépolitisation d'une notion. *Politiques de communication*, *8*, 127–154.
Haddad, G. (2015). *Dans la main droite de Dieu. Psychanalyse du fanatisme*. 1ER PARALLELE.
Haddad, G. (2015). *Dans la main droite de Dieu. Psychanalyse du fanatisme*. Armand Colin.
Khosrokhavar, F. (2014). *Radicalisation*. Éditions de la Maison des sciences de l'homme.
Lorenzi, N., & Moïse, C. (2018). *Radicality, radicalization, hate: Discourse and critical reflections: Practicies project, European Project H2020* [Research report]. Université Grenoble Alpes, 47 p.
Practicies report 7, Fernández, V., Dago, P., Martín, M. I., Nieto, I., Cerezo, H., Abalde, A., Lago, J., et al. (2019). D7.2– D7.6 – Final version of the PRACTICIES platform with the different components integrated Europa.
Salazar, P. (2015). *Paroles armées. Comprendre et combattre la propagande terroriste*. Lemieux Éditeur.
Salon, J. R., & Gerber, M. S. (2014). Automatic detection of cyber-recruitment by violent extremists. *Security Informatics*, *3*(1), 5.
Vincent, D. (2005). Analyse conversationnelle, analyse du discours et interprétation des discours sociaux: le cas de la trash radio. *Marges linguistiques*, numéro 9, mai 2005.