Global Network
on Extremism & Technology

Décrypter la haine : emploi de l'analyse de texte expérimentale aux fins de classification des contenus à caractère terroriste

Abdullah Alrhmoun, Shiraz Maher, Charlie Winter

Le GNET est un projet spécial du Centre international d'étude de la radicalisation du King's College, à Londres.

Les auteurs de ce rapport sont Abdullah Alrhmoun, chercheur doctorant à l'Université d'Europe centrale, Vienne, Autriche ; le Dr. Shiraz Maher, directeur du Centre international d'étude de la radicalisation (ICSR) du King's College, à Londres ; et le Dr. Charlie Winter, chercheur principal à l'ICSR.

Le Global Network on Extremism and Technology (Réseau mondial sur l'extrémisme et la technologie – GNET) est une initiative de recherche universitaire bénéficiant du soutien du Forum mondial de l'Internet contre le terrorisme (GIFCT), une initiative indépendante mais financée par le secteur qui vise à mieux comprendre et lutter contre l'utilisation des technologies par les groupes terroristes. Le GNET est formé et dirigé par le Centre international d'étude de la radicalisation (ICSR), un centre de recherche universitaire basé dans les locaux du Département d'étude des guerres du King's College, à Londres. Les opinions et conclusions exprimées dans ce document sont celles des auteurs et ne doivent en aucun cas être interprétées comme représentant les opinions et conclusions, expresses ou implicites, du GIFCT, du GNET ou de l'ICSR.

Cet ouvrage a été financé par une bourse de recherche de Facebook dans le cadre de son projet de recherche « Content Policy Research on Social Media Platforms ». Les opinions et conclusions exprimées dans ce document sont celles des auteurs et ne doivent pas être interprétées comme représentant les politiques, expresses ou implicites, de Facebook.

COORDONNÉES

Pour toute question, demande d'information et demande de copies supplémentaires du présent rapport, contacter :

ICSR
King's College London
Strand
Londres WC2R 2LS
Royaume-Uni

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**

Ce rapport peut, comme toutes les autres publications du GNET, être téléchargé gratuitement à partir du site Internet du GNET : www.gnet-research.org.

Table des matières

1 Introduction

Le présent document emploie l'analyse de texte automatisée – un processus par lequel du texte non structuré est extrait, organisé et transformé en un format utile – pour développer des outils capables d'analyser la propagande de l'État islamique (EI) à l'échelle[1]. Bien que nous ayons utilisé des archives statiques de contenu publié par l'EI, le principe sous-jacent veut que ces techniques puissent être déployées, en temps réel, contre les contenus produits par un nombre indéfini de mouvements extrémistes violents. Cette étude vise par conséquent à compléter les travaux portant sur les stratégies axées sur les technologies employées par les réseaux sociaux et les plateformes d'hébergement de vidéos et de partage de fichiers pour lutter contre les propagateurs de contenu extrémiste violent[2]. Ces plateformes visent en règle générale à supprimer le contenu produit par les organisations terroristes et à caractère haineux, à moins que la diffusion dudit contenu ait lieu dans des cadres très spécifiques (par exemple, lorsqu'il est diffusé par des journalistes ou des chercheurs)[3]. Les mesures collectives déployées ces dernières années par ces plateformes sont devenus extrêmement performantes, avec la suppression de la quasi-totalité du contenu à caractère terroriste avant même qu'il ne soit signalé[4].

Tout le contenu à caractère terroriste n'a toutefois pas la même valeur[5]. Certains contenus doivent être supprimés en priorité[6]. Des problèmes d'automatisation peuvent survenir, notamment dans les cas où la technologie est employée pour repérer les contenus préjudiciables, qui sont ensuite examinés par des évaluateurs humains pour décision finale ; un tel processus peut présenter un défi majeur concernant la priorité à accorder au contenu à examiner, la méthode à employer pour le supprimer et le moment où le supprimer[7]. En d'autres termes, est-il possible de développer des technologies visant à évaluer les contenus de façon efficace et précise ? Il convient de faire la distinction entre les contenus devant être examinés de façon instantanée et les contenus pouvant être placés en file d'attente[8]. Prenons par exemple le risque relatif que représente une photographie assez inoffensive présentant une scène de socialisation terroriste par rapport à une vidéo montrant des scènes explicites de violence.

1 Justin Grimmer et Gary King, « General Purpose Computer-Assisted Clustering and Conceptualization », travaux de la National Academy of Sciences, 2011. Disponible à l'adresse : https://j.mp/2nRjqbO ; Gary King et Justin Grimmer, « Method and Apparatus for Selecting Clusterings to Classify A Predetermined Data Set », États-Unis d'Amérique 8,438,162 (7 mai), 2013. Disponible à l'adresse : https://j.mp/2ovzAuR.

2 Pour voir un exemple, consulter : James Vincent, « UK creates machine learning algorithm for small video sites to detect ISIS propaganda », The Verge, 13 février 2018. Disponible à l'adresse : https://www.theverge.com/2018/2/13/17007136/uk-government-machine-learning-algorithm-isis-propaganda.

3 Guy Rosen, « How are we doing at enforcing our community standards? » Facebook, 15 novembre 2018. Disponible à l'adresse : https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/.

4 Monica Bickert, « Hard questions: What are we doing to stay ahead of terrorists? » Facebook, 8 novembre 2018. Disponible à l'adresse : https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/.

5 Voir, par exemple : Charlie Winter, « Understanding jihadi stratcom: The case of the Islamic State », Perspectives on Terrorism vol. 13:1 (2019) : 54–62 ; Charlie Winter et Dounia Mahlouly, « A tale of two caliphates: Comparing the Islamic State's internal and external messaging priorities ». VOX-Pol, juillet 2019 ; Stephane Baele et Charlie Winter, « From music to books, from pictures to numbers: The forgotten—yet crucial—components of IS' propaganda », in Stephane Baele, Travis Coane et Katharine Boyd (dir.), The Propaganda of the Islamic State, Oxford : Oxford University Press (2019).

6 Bickert. « Hard questions ».

7 Alex Schulz et Guy Rosen, « Understanding the Facebook: Community standards enforcement report », Facebook, mai 2020. Disponible à l'adresse : https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf. p. 17.

8 Bickert, « Hard questions ».

Par ailleurs, l'automatisation a traditionnellement été déployée par le passé contre le *contexte* dans lequel les messages étaient publiés sur les réseaux sociaux et non contre le *contenu* même. Par conséquent, les enquêtes menées n'exploitent pas correctement les types d'outils les plus pratiques pour les modérateurs des sociétés technologiques.

Ce projet s'interroge sur la manière dont les sociétés technologiques peuvent faire cette distinction de manière opportune et précise. Il nuancera la façon dont les contenus préjudiciables sont classés en utilisant l'EI comme cas d'essai. Nous partons de l'hypothèse suivante : il est possible de codifier l'intention de publier des contenus préjudiciables en étudiant puis en testant la logique sous-tendant leur production[9]. En effet, s'il est possible de repérer l'intention – en d'autres termes, s'il est possible de faire une distinction claire entre le contenu tactique fondé sur l'action et le contenu stratégique fondé sur la marque[10] –, il sera alors possible de mieux définir les priorités concernant l'examen et la suppression en fonction du risque posé.

À cette fin, l'objectif de ce rapport est de synthétiser l'expertise en la matière et la science des données, en employant des techniques expérimentales de traitement de texte pour interroger et catégoriser nos archives de contenu officiel de l'EI. Notre but principal est de concevoir des méthodes automatisées capables, lorsque appliquées à des corpus de supports similaires (y compris des corpus beaucoup plus vastes), d'accélérer le processus par lequel les données peuvent être séparées et, le cas échéant, triées pour modération et/ou aiguillage. Cela permettra, à son tour, d'améliorer les politiques de modération de contenu en vigueur.

Compte tenu du volume considérable de contenu produit chaque minute, le besoin d'une telle approche est clairement pressant. Plus de 300 heures de vidéos sont téléchargées chaque minute sur la seule plateforme YouTube, dont les utilisateurs visionnent plus d'un milliard d'heures de vidéos chaque jour[11]. En moyenne, 500 millions de tweets sont produits par jour, soit environ 200 milliards par an[12]. Au premier trimestre 2020, Facebook enregistrait plus de 2,6 milliards d'utilisateurs actifs mensuels, tandis qu'Instagram, qui appartient à Facebook, publie plus de 500 millions de « stories Instagram » chaque jour[13]. Bien entendu, la très grande majorité des utilisateurs se sert de ces plateformes à des fins totalement innocentes et légitimes. Nous ne suggérons en aucun cas que ce contenu doive être censuré ou surveillé d'une quelconque façon. Toutefois, la présence d'extrémistes violents et mal intentionnés opérant au cœur de ce raz de marée prouve qu'il est nécessaire d'adopter des méthodes automatisées efficaces, capables d'identifier, d'analyser et de séparer l'ensemble de contenus qu'ils produisent.

9 Selon les termes de Berger, les récits respectifs des différentes formes d'extrémisme sont très variés, contrairement aux structures dans le cadre desquelles ils sont racontés. Des versions futures du classificateur pourraient être développées pour aider les politiques et pratiques de Facebook en matière de contenus concernant d'autres formes d'extrémisme. Voir : JM Berger, Extremism, Cambridge : The MIT Press (2018) : 51–112.

10 Pour un rapport introductif sur cette distinction, voir : Winter, « Understanding jihadi stratcom ».

11 Données collectées le 11 août 2020 sur « YouTube pour la presse ». Disponible à l'adresse : https://www.youtube.com/intl/fr/about/press/.

12 David Sayce, « The number of tweets per day in 2020 », David Sayce, mai 2020. Disponible à l'adresse : https://www.dsayce.com/social-media/tweets-day/.

13 Jessica Clement, « The number of monthly active Facebook users worldwide as of the first quarter of 2020 », Statista, 10 août 2020. Disponible à l'adresse : https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ ; Maryam Mohsin, « 10 Instagram Stats Every Marketer Should Know in 2020 », Oberlo, 6 février 2020. Disponible à l'adresse : https://www.oberlo.co.uk/blog/instagram-stats-every-marketer-should-know.

C'est pour cette raison que nous avons choisi de nous concentrer sur les contenus de l'EI, qui a mis en lumière la question du contenu extrémiste violent. Globalement, le mouvement djihadiste au sens général a exploité les technologies à des fins de propagande bien mieux que les autres mouvements[14]. Dans les années 1990, des sites Internet statiques, tels qu'Azzam.com, apportaient aux publics anglophones des informations sur les campagnes djihadistes en Tchétchénie, en Bosnie et en Afghanistan. Après l'invasion de l'Irak en 2003, les forums de discussion protégés par mot de passe, tels qu'Ansar al-Mujahideen (« partisans du moudjahidine »), Faloja (référence à la ville irakienne de Fallujah, devenue l'un des foyers d'activité des insurgés) et Shamukh (« noble », ou « personne admirable »), sont devenus les principales formes de diffusion de contenus extrémistes violents, tels que vidéos et communiqués de groupes comme al-Qaïda, al-Shabab et Boko Haram[15].

Ces forums étaient des environnements relativement statiques et insulaires. Les contenus extrémistes violents devaient par conséquent être recherchés délibérément, puisqu'ils n'existaient que dans des recoins d'Internet difficiles d'accès. Au moment des soulèvements du Printemps arabe, en 2011, les réseaux sociaux étaient devenus le principal instrument employé par les acteurs violents pour diffuser des contenus et attirer de nouvelles recrues. La montée de l'EI et du Front al-Nosra, affilié à al-Qaïda, entre 2011 et 2016, illustre cette situation de façon saisissante[16]. Le problème ne se limitait pas seulement à leur présence sur ces plateformes, mais portait aussi sur le fait que du contenu extrémiste était désormais très facilement accessible à toute personne souhaitant y accéder – et qu'un certain nombre de personnes tombaient dessus de façon purement accidentelle. La question de savoir comment lutter contre ce phénomène est ainsi devenue particulièrement épineuse pour les sociétés technologiques, les forces de l'ordre et les décideurs politiques spécialisés dans la lutte antiterroriste[17].

Notre rapport traite de cette question en réfléchissant aux façons dont l'automatisation peut aider à repérer et séparer ces contenus, permettant ainsi aux sociétés technologiques de détecter ces contenus plus facilement lorsqu'ils côtoient des publications postées par des utilisateurs légitimes de leurs plateformes.

Il convient de remarquer, avant de poursuivre, que les outils développés ici pourraient avoir des effets sur d'autres types de contenus extrémistes non djihadistes que les entreprises voudront surveiller en fonction de leurs propres besoins. Le contenu préjudiciable en ligne peut prendre différentes formes, telles qu'abus,

14 Pour une perspective technologique, voir : Brian Fishman, « Crossroads: Counter-terrorism and the Internet », Texas National Security Review, février 2019. Disponible à l'adresse : https://tnsr.org/2019/02/crossroads-counter-terrorism-and-the-internet/. Pour une perspective gouvernementale, voir « How social media is used to encourage travel to Syria and Iraq: Briefing note for schools », ministère de l'Éducation du Royaume-Uni, juillet 2015. Disponible à l'adresse : https://www.gov.uk/government/publications/the-use-of-social-media-for-online-radicalisation. Pour une perspective gouvernementale internationale, voir « Utilisation de l'Internet à des fins terroristes », Office des Nations Unies contre la drogue et le crime, septembre 2012. Disponible à l'adresse : https://www.unodc.org/documents/terrorism/Publications/The_Use_of_Internet_for_Terrorist_Purposes/Use_of_the_Internet_for_Terrorist_Purposes_French.pdf.

15 Evan Kohlmann, « A beacon for extremists », CTC Sentinel, février 2010, vol. 3, n° 2. Disponible à l'adresse : https://ctc.usma.edu/a-beacon-for-extremists-the-ansar-al-mujahideen-web-forum/ ; Manuel R. Torres-Soriano, « The Hidden Face of Jihadist Internet Forum Management: The Case of Ansar Al Mujahideen », Terrorism and Political Violence vol. 28, n° 4 (2016).

16 Gunnar J. Weimann, « Competition and Innovation in a Hostile Environment: How Jabhat Al-Nusra and Islamic State Moved to Twitter in 2013–2014 », Studies in Conflict & Terrorism, vol. 42 (2019) : 1–2, 25-42.

17 Pour une perspective technologique, voir : Fishman, « Crossroads ». Pour une perspective gouvernementale, voir « How social media is used ». Pour une perspective gouvernementale internationale, voir « Utilisation de l'Internet à des fins terroristes ».

intimidations, harcèlement, menaces, misogynie, discours haineux ou propagande terroriste ou violente, entre autres. Ces activités en ligne peuvent se concrétiser sous forme de préjudices causés dans la vie réelle, mais peut-être de façon moins impressionnante que ceux causés par la violence djihadiste[18]. Dans tous les cas, ces situations nécessitent elles aussi le type de modération nuancée et contextualisée dont nous parlons dans ce rapport.

18 Fin 2018, soit quatre mois seulement avant que les derniers bastions de l'EI en Syrie soient libérés par des forces soutenues par la coalition, l'utilisation des réseaux sociaux par cette organisation était considérée comme constituant une « menace [directe] à la stabilité du Moyen-Orient et de l'Afrique ». Antonia Ward, « ISIS's use of social media still poses a threat to stability in the Middle East and Africa », RAND Corporation, 11 décembre 2018. Disponible à l'adresse : https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html.

2 Analyse documentaire

Il existe déjà un vaste corpus documentaire s'intéressant au contenu extrémiste violent en ligne, en particulier à celui publié par l'EI. Après son émergence en Syrie et en Irak, l'EI s'est rapidement établi comme un pionnier à la fois de la communication stratégique extrémiste et de la diffusion sur Internet. La recherche analysant ses productions à cette fin se manifeste généralement sous trois formes : i) analyse quantitative de sa base de soutien sur les réseaux sociaux ; ii) analyse qualitative des textes ou genres de propagande individuelle ; et iii) analyse fondée sur les données de sa production médiatique agrégée.

Le premier ensemble étudie les interactions en ligne des partisans de l'EI. Leur activisme sur les plateformes grand public, comme Twitter, Facebook et YouTube, suscite beaucoup d'attention depuis 2014 en particulier. Les recherches de Carter, Maher et Neumann ont constitué l'un des premiers efforts visant à recenser ces communautés diverses, en étudiant les réseaux d'influence en ligne existant entre djihadistes anglophones. Elles ont été suivies par d'autres études de même nature menées par des collègues respectés tels que Klausen, Berger et Morgan[19]. Des recherches ultérieures sur cette même question, menées par Conway et ses collègues ou encore par Alexander, montrent que la présence djihadiste sur les plateformes grand public s'est réduite à partir de 2015, les groupes concernés se tournant davantage vers de nouveaux services assurant le respect de la vie privée pour communiquer[20]. Malgré cette migration, Winterbotham, entre autres chercheurs, a également montré que les plateformes grand public, telles que Twitter et Facebook, continuent d'avoir de l'importance pour le mouvement, même si ses activités y ont diminué[21]. Un certain nombre d'études suivent par ailleurs les dynamiques idiomatiques au sein des communautés extrémistes sur les réseaux sociaux en s'appuyant sur des modèles linguistiques visant à détecter la présence de

19 Joseph A. Carter, Shiraz Maher et Peter R. Neumann, « #Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks », Centre international d'étude de la radicalisation, avril 2014. Disponible à l'adresse : https://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Influence-in-Syrian-Foreign-Fighter-Networks.pdf ; Jytte Klausen, « Tweeting the jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq », Studies in Conflict & Terrorism, vol. 38:1 : 1–22 ; J. M. Berger et Jonathon Morgan, « The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter », The Brookings Project on U.S. Relations with the Islamic World, n° 20, mars 2015. Disponible à l'adresse : https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

20 Maura Conway *et al.*, « Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts », Studies in Conflict & Terrorism, vol. 42, 2019. Disponible à l'adresse : http://doras.dcu.ie/21961/1/Disrupting_DAESH_FINAL_WEB_VERSION.pdf ; Audrey Alexander, « Digital Decay: Tracing Change Over Time Among English-Language Islamic State Sympathizers on Twitter », programme sur l'extrémisme, Université George Washington, octobre 2017. Disponible à l'adresse : https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/DigitalDecayFinal_0.pdf ; Miron Lakomy, « Mapping the online presence and activities of the Islamic State's unofficial propaganda cell: Ahlut-Tawhid Publications », Security Journal, disponible en ligne uniquement.

21 Ugur Kursuncu *et al.*, « Modeling Islamist Extremist Communications on Social Media Using Contextual Dimensions: Religion, Ideology and Hate », Procès-verbal de l'ACM sur les interactions personne-machine, 3:1, août 2019 ; Moustafa Ayad, « 'The Baghdadi Net': How a Network of ISIL-Supporting Accounts Spread across Twitter », Institut pour le dialogue stratégique, novembre 2019. Disponible à l'adresse : https://www.voxpol.eu/download/report/E28098The-Baghdadi-NetE28099-How-A-Network-of-ISIL-Supporting-Accounts-Spread-Across-Twitter.pdf ; Leevia Dillon *et al.*, « A comparison of ISIS foreign fighters and supporters social media posts: an exploratory mixed-method content analysis », Behavioural Sciences of Terrorism and Political Aggression, disponible en ligne uniquement ; Airbus Defence and Space, « Mapping Extremist Communities: A Social Network Analysis Approach », Centre d'excellence de l'OTAN pour les communications stratégiques, janvier 2020. Disponible à l'adresse : https://www.voxpol.eu/download/report/web_stratcom_coe_mapping_extremist_strategies_31.03.2020_v2.pdf.

discours radicalisés voire, parfois, à prédire les comportements[22]. Ces études sont largement expérimentales à l'heure actuelle, et ne portent pas sur des groupes spécifiques.

Le deuxième ensemble de travaux consiste en des interrogations qualitatives de produits et genres de propagande individuelle. D'innombrables études des magazines de l'EI en langues étrangères ont été réalisées ces dernières années, dont certaines portant sur son journal officiel en langue arabe, *al-Naba*[23]. D'autres chercheurs, comme Winkler et ses collègues ou Adelman, ont préféré se concentrer sur les centaines d'infographies publiées par l'EI depuis 2015, tandis que d'autres, tels que Nanninga, Dauber et Robinson, ont axé leurs efforts sur leur production vidéo[24]. El Damanhoury et Milton font partie des rares chercheurs à avoir étudié les vastes archives d'images fixes de l'EI, desquelles on peut – et devrait – parler davantage[25]. Malgré la diversité de leur sujet, ces études fondées sur des genres parviennent à des conclusions similaires concernant la présence dominante de motifs visuels occidentaux dans la propagande de l'EI.

Le dernier ensemble se caractérise par des travaux menés par des chercheurs comme Zelin, Milton et Winter, dont les efforts respectifs portent sur l'étude d'archives de production médiatique officielle de l'EI[26]. D'une manière générale, leurs conclusions sont cohérentes entre elles ; elles identifient toutes une diminution nette de la quantité de propagande produite par le groupe, en corrélation approximative avec ses contractions territoriales depuis 2015, situation également

22 Tom De Smedt *et al.*, « Automatic Detection of Online Jihadist Hate Speech », Computation and Language vol. 7 :1–31, 2018 ; Adam Bermingham *et al.*, « Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation », Conférence internationale 2009 sur les avancées en matière d'analyse et d'exploitation des réseaux sociaux, 2009 : 231–6 ; Edna Reid *et al.*, « Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites », Conférence internationale sur l'informatique du renseignement et de la sécurité, 2005 : 402–11 ; Enghin Omer, « Using machine learning to identify jihadist messages on Twitter », Université d'Uppsala, 2015. Disponible à l'adresse : http://www.diva-portal.org/smash/get/diva2:846343/FULLTEXT01.pdf.

23 Haroro J. Ingram, « An analysis of Islamic State's Dabiq Magazine », Australian Journal of Political Science, vol. 51:3, 2016 : 458–577 ; Julian Droogan et Shane Peattie, « Mapping the thematic landscape of Dabiq magazine », Australian Journal of Political Science, vol. 71:6, 2017 : 591–620 ; Haroro J. Ingram, « An analysis of Inspire and Dabiq: Lessons from AQAP and Islamic State's propaganda war », Studies in Conflict & Terrorism, vol. 40:5, 2017 : 357–75 ; Nuria Lorenzo-Dus *et al.*, « Representing the West and 'non-believers' in the online jihadist magazines Dabiq and Inspire », Critical Studies on Terrorism, vol. 11:3, 2018 ; Carol K. Winkler *et al.*, « The medium is terrorism: Transformation of the about to die trope in Dabiq », Terrorism and Political Violence, vol. 31:2, 2019 ; Peter Wignell *et al.*, « Under the shade of AK47s: a multimodal approach to violent extremist recruitment strategies for foreign fighters », Critical Studies on Terrorism, vol. 10:3, 2017 : 429–52 ; Logan Macnair et Richard Frank, « Changes and stabilities in the language of Islamic state magazines: a sentiment analysis », Dynamics of Asymmetric Conflict, vol. 11:2, 2018 : 109–20 ; Orla Lehane *et al.*, « Brides, black widows and baby-makers; or not: an analysis of the portrayal of women in English-language jihadi magazine image content », Critical Studies on Terrorism, vol. 11:3, 2018 ; Dounia Mahlouly et Charlie Winter, « A Tale of Two Caliphates: Comparing the Islamic State's Internal and External Messaging Priorities », VOX-Pol, 2018. Disponible à l'adresse : https://www.voxpol.eu/download/vox-pol_publication/A-Tale-of-Two-Caliphates-Mahlouly-and-Winter.pdf ; Miron Lakomy, « Towards the 'olive trees of Rome': Exploitation of propaganda devices in the Islamic State's flagship magazine 'Rumiyah' », Small Wars & Insurgencies vol. 31:3, 2020 : 540–68 ; Michael Zekulin, « From Inspire to Rumiyah: does instructional content in online jihadist magazines lead to attacks? », Behavioural Sciences of Terrorism and Political Aggression, disponible en ligne uniquement.

24 Pieter Nanninga, « Meanings of savagery », in Lewis, J. (dir.), The Cambridge Companion to Religion and Terrorism, Cambridge : Cambridge University Press, 2017 : 172–90 ; Cori E. Dauber et Mark Robinson, « ISIS and the Hollywood visual style », Jihadology, 6 juillet 2015. Disponible à l'adresse : https://jihadology.net/2015/07/06/guest-post-isis-and-the-hollywood-visual-style/ ; Cori E. Dauber *et al.*, « Call of Duty: Jihad – How the Video Game Motif has Migrated Downstream from Islamic State Propaganda Videos », Perspectives on Terrorism, vol. 13:3, 2019 ; Pieter Nanninga, « Branding a Caliphate in Decline: The Islamic State's Video Output (2015–2018) », International Centre for Counter-terrorism – La Haye, avril 2019. Disponible à l'adresse : https://icct.nl/publication/branding-a-caliphate-in-decline-the-islamic-states-video-output-2015-2018/.

25 Kareem El Damanhoury *et al.*, « Examining the military-media nexus in ISIS's provincial photography campaign », Dynamics of Asymmetric Conflict, vol. 11:2, 2018 : 89–108 ; Daniel Milton, « Fatal attraction: Explaining variation in the attractiveness of Islamic State propaganda », Conflict Management and Peace Science vol. 37:4, 2018 ; Carol Winkler *et al.*, « Intersections of ISIS media leader loss and media campaign strategy: A visual framing analysis », Media, War & Conflict, 2019, disponible en ligne uniquement.

26 Aaron Y. Zelin, « Picture Or It Didn't Happen: A Snapshot of the Islamic State's Media Output », Perspectives on Terrorism vol. 9:4, 2015 ; Daniel Milton, « Communication Breakdown: Unraveling the Islamic State's Media Efforts », Combating Terrorism Center at West Point, 2016. Disponible à l'adresse : https://ctc.usma.edu/communication-breakdown-unraveling-the-islamic-states-media-efforts/ ; Daniel Milton, « Down, but Not Out: An Updated Assessment of the Islamic State's Visual Propaganda », Combating Terrorism Center at West Point, 2018. Disponible à l'adresse : https://ctc.usma.edu/down-but-not-out-an-updated-examination-of-the-islamic-states-visual-propaganda/ ; voir également : Charlie Winter, « Apocalypse, later: A longitudinal study of the Islamic State brand », Critical Studies in Media Communication, vol. 35:1, 2018 : 103–21.

observée par Nanninga. Il convient de noter, toutefois, que cette baisse n'est pas nécessairement due à la perte de territoire, même si les deux phénomènes semblent reliés. Si une série de conclusions intuitives peuvent être tirées de cette situation, il n'y a pas encore de consensus définitif sur les raisons de ce ralentissement et, sans surprise, l'EI n'a jamais abordé la question[27].

L'étude actuelle s'appuie sur les premier et troisième ensembles de travaux. En ce qui concerne le premier, elle propose une nouvelle forme de traitement de texte expérimental, axée non pas sur les publications sur les réseaux sociaux, mais sur leur contenu même. Sa contribution au troisième groupe est assez évidente, compte tenu de la nature archivistique et multimédias des données. Nous espérons apporter des nuances supplémentaires au débat portant sur l'évolution, ces dernières années, des activités de sensibilisation de l'EI et les raisons de cette évolution.

27 Il convient de noter que le consensus n'est pas tout à fait complet, Fisher affirmant qu'il n'existe pas de baisse de la productivité. Ali Fisher, « ISIS: Sunset on the "decline narrative" », Online Jihad, 2018. Disponible à l'adresse : https://onlinejihad.net/2018/06/01/isis-sunset-on-the-decline-narrative/.

3 Méthodologie

L'aspect le plus important de notre méthodologie porte sur la façon dont l'outil d'analyse automatisée de texte a été développé et déployé. C'est sur cette hypothèse que nous avons fondé notre modèle d'analyse automatique de contenu à l'échelle dans le but de trier les contenus susceptibles d'être préjudiciables.

L'ensemble de données que nous avons utilisé a été tiré d'un site Internet statique administré par un ou plusieurs partisan(s) inconnu(s) de l'EI[28]. Facilement accessible sur le web surfacique, ce site est en circulation parmi les djihadistes depuis plusieurs années, avec des liens apparaissant sur les forums de discussion publics, les plateformes de réseaux sociaux grand public et d'autres plateformes plus tournées vers l'intérieur comme Telegram. Le site Internet a été entièrement téléchargé, archivé et stocké en février 2020. Il comprenait au total 6 290 éléments individuels – allant de reportages photos et vidéos à des déclarations des dirigeants, en passant par des bulletins radio et des magazines.

Élaboration de notre cadre thématique

L'algorithme que nous avons développé (expliqué ci-dessous) est fondé sur un cadre analytique conçu pour séparer le contenu de l'EI en fonction de différents thèmes, développés par l'un des auteurs, le Dr Winter, pour un projet antérieur[29]. Ce cadre se compose de deux catégories – i) guerre et ii) vie civile, contenant 22 catégorisations thématiques différentes[30] listées ci-dessous, dont neuf portant sur la guerre et treize sur la vie civile.

I) Thématiques guerrières

1. *Opérations* : offensives militaires. Celles-ci varient en fonction du contexte, de l'objectif déclaré et de la/des tactique(s) employée(s). Les trois éléments les plus fréquents décrivent des attaques au sol, des opérations à base d'engins explosifs artisanaux et des attentats-suicides. Parmi les autres tactiques présentées, citons, entre autres, les frappes de drones ou encore les embuscades nocturnes.

2. *Résumé* : rapports d'actualité agrégés provenant de tous les territoires du califat. Ils prennent la forme de bulletins journaliers, hebdomadaires et mensuels, souvent appuyés par des statistiques.

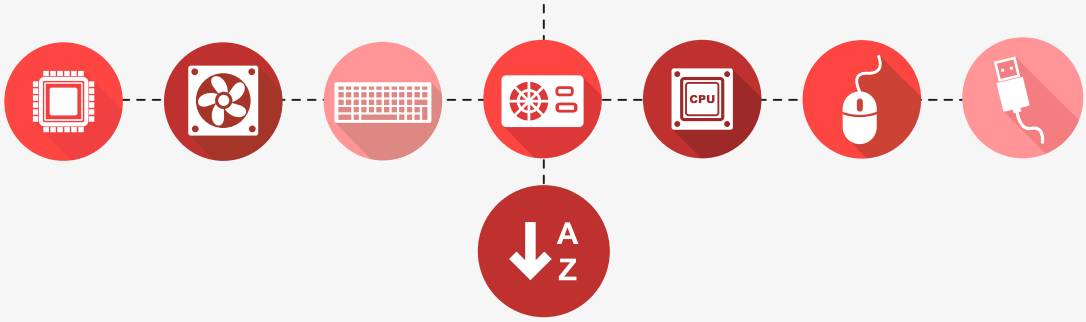
28 Nous avons choisi de ne pas nommer l'archive en question, puisqu'elle demeure facilement accessible en ligne et que cela lui apporterait de la publicité inutile. Les lecteurs sont invités à contacter les auteurs directement pour toute question.

29 Charlie Winter, « The Terrorist Image: A Mixed Methods Analysis of Islamic State Photo-Propaganda », thèse de doctorat non publiée, King's College London, juillet 2020.

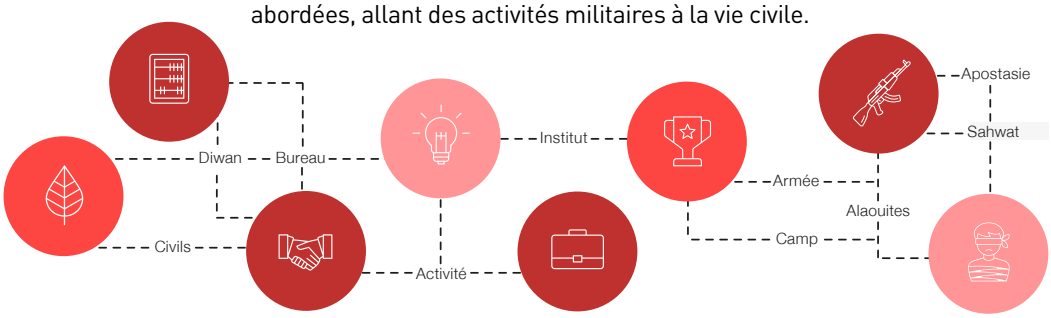
30 Voir également : Milton, « Communication Breakdown » ; Milton, « Down, but Not Out » ; Zelin, « Picture Or It Didn't Happen » ; Winter, « Apocalypse, later ».

L'archive a été téléchargée et enregistrée.

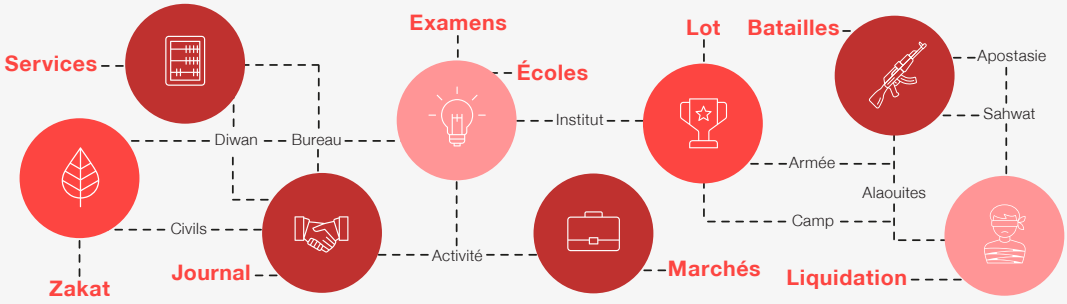
Les 6 290 éléments qu'elle contenait ont été triés par type de média et date de publication. Les 803 mots qui apparaissaient le plus fréquemment dans les titres de chacun de ces éléments ont été identifiés.



Ces mots ont ensuite été marqués d'un à trois tag en fonction des thématiques abordées, allant des activités militaires à la vie civile.



Les mots portant exclusivement sur une thématique ont été nommés « super-tags ».



L'algorithme a appliqué ce système de tags et de super-tags à l'ensemble des données extraites.

3. *Guerre indirecte* : attaques contre les positions ennemies à l'aide de roquettes, de missiles et de mortiers. Ces supports portent généralement sur le lancement de projectiles et montrent rarement les conséquences des attaques.

4. *Martyrologie* : Ce contenu glorifie les hommes et les garçons qui meurent en combattant au nom de l'EI. Presque tous les individus encensés dans ces images sont photographiés de leur vivant, soit dans un environnement bucolique ou urbain idyllique, soit à bord du véhicule à bombe embarquée dans lequel ils se préparent à mourir. Les personnes ainsi commémorées vont des agents commentant des attentats-suicides aux dirigeants de faible rang ou de rang intermédiaire en passant par les responsables de la propagande.

5. *Service de garnison* : supports amplifiant la vie sur la ligne de front de l'EI. Ces supports donnent un aperçu de la vie quotidienne dans les garnisons, allant des moments de prière à la préparation des repas, du nettoyage des armes à l'exercice physique.

6. *Exécutions* : supports documentant l'exécution des « espions » ou des prisonniers de guerre capturés lors d'enlèvements ou d'attaques, y compris des membres de groupes extrémistes violents rivaux. La propagande liée aux exécutions ayant lieu dan un contexte ouvertement martial ne doit pas être confondue avec la propagande portant sur les exécutions menées dans un contexte civil.

7. *Opérations de défense* : activités militaires à caractère défensif. La plupart de ces supports portent sur des offensives ennemies déjouées et des contre-attaques « réussies ». D'autres relaient des informations sur les défenses antiaériennes, les mesures de préparation militaire, la construction de fortifications et l'entretien des systèmes d'armement.

8. *Conséquences* : les conséquences d'une attaque. Cette catégorie englobe quatre groupes distincts : butins de guerre ; prisonniers ennemis ; cadavres ennemis ; et véhicules au sol et drones endommagés.

9. *Entraînement* : ces supports décrivent les camps d'entraînement, et montrent des activités comme les exercices de maniement des armes, les exercices physiques et le combat à mains nues.

II) Thématiques de la vie civile

10. *Ordre public* : administration de l'ordre public dans les territoires de l'EI. Ces supports présentent trois aspects principaux : images de la police religieuse (*hisbah*) ; images de procédures pénales ; et images de déploiements de la police.

11. *Victimisation* : ces supports illustrent les suites des attaques contre les territoires de l'EI, montrant généralement des enfants morts ou blessés et des infrastructures publiques détruites. Ils servent à justifier le régime de l'EI, à obtenir des soutiens et à provoquer des représailles violentes.

12. *Sensibilisation* : la plupart de ces supports suivent le travail des représentants des médias, principalement leur diffusion de contenu et l'installation d'infrastructures médiatiques. D'autres illustrent des activités telles que des réunions de réconciliation entre tribus rivales et des rassemblements réunissant des civils et des dignitaires.

13. *Visites guidées* : ces supports présentent la vie « quotidienne » dans les bastions de l'EI. Le contenu de cette catégorie prend souvent la forme de visites organisées dans des villages, villes ou quartiers spécifiques, et recense généralement tous les aspects de la vie quotidienne allant des activités religieuses à la vie des oiseaux en passant par les activités commerciales et les loisirs.

14. *Vie religieuse* : supports principalement axés sur les activités religieuses, montrant des civils dans un ensemble de contextes « islamiques », comme les fêtes de l'Eid et du Ramadan, les prières du vendredi et les concours de mémorisation du Coran.

15. *Vie commerciale* : supports montrant les aspects commerciaux de la vie au sein du califat. La plupart portent sur les échanges et le commerce, tandis que d'autres représentent des visites de marchés, magasins et salles d'exposition.

16. *Services municipaux* : la prestation de services dans les territoires de l'EI. Très variés, ces supports montrent les bureaux de services exerçant toutes sortes d'activités telles que réparations de pylônes et entretien des égouts.

17. *Protection sociale* : ces supports montrent le calcul, la préparation et la distribution de la protection sociale (financière et alimentaire) au sein de la population civile des territoires détenus par l'EI. La plupart s'accompagnent de vignettes de type « une journée dans la vie de… » donnant un aperçu général des activités du bureau de la protection sociale.

18. *Vie industrielle* : supports portant sur les activités industrielles du califat, des usines de fabrication de tuyaux et ateliers de fabrication d'air conditionné aux usines de production de fromage et de séchage des graines de tournesol.

19. *Vie agricole* : ces supports couvrent l'activité agricole, de l'ensemencement au transport de fruits sur les marchés en passant par la récolte.

20. *Éducation* : ces supports tournent généralement autour des activités de séminaire organisées par les autorités officielles chargées du prosélytisme. D'autres montrent des visites dans les écoles, allant d'adolescents passant leurs examens de mi-trimestre et de fin d'année à des enfants en bas âge jouant pendant la récréation.

21. *Santé* : ces supports montrent les activités médicales réalisées sous l'autorité de l'EI. Ils présentent des visites guidées d'hôpitaux et de cabinets dentaires ou des visites à domicile de médecins généralistes et des campagnes de vaccination pour les enfants.

22. *Paysages et nature* : ce contenu contient presque essentiellement des photographies de sites de beauté naturelle ou monumentale.

Développement de l'algorithme

Notre algorithme a été codé pour rechercher des marqueurs linguistiques dans le seul titre de chacun des 6 290 éléments individuels identifiés dans notre base de données. De nombreux mouvements millénaristes et réactionnaires emploient leurs propres marqueurs linguistiques lexicaux et syntaxiques, qui définissent l'intragroupe. L'un des exemples les plus notables de marqueurs est fourni par les triples parenthèses ajoutées autour du (((nom))) d'éminents Juifs par les membres de certaines communautés néo-nazies ou de la droite alternative. L'idée d'origine sous-jacente consiste à identifier les Juifs ou les personnes d'origine juive[31]. Les individus associés à la droite alternative et/ou au mouvement néo-nazi savent alors que la personne concernée doit être considérée sous le prisme de la conspiration ou de l'intrigue antisémite.

La communauté incel (« célibataires involontaires », ou « involuntary celibate » en anglais), qui regroupe des hommes sexuellement inactifs hostiles ou méfiants vis-à-vis des femmes et du féminisme, offre un autre exemple de langage spécifique à certaines sous-cultures présentes sur le web[32]. Le terme « Chad », par exemple, est utilisé par la communauté incel pour parler de façon désobligeante des hommes attirants et populaires, présumés avoir une vie sexuelle active avec les femmes. De même, le terme « Stacy » est utilisé pour décrire les femmes attirantes et séduisantes uniquement intéressées par les « Chads »[33].

Il en va de même pour le titrage du contenu publié par l'EI. L'organisation utilise des marqueurs linguistiques pour traiter de thématiques ou de questions spécifiques. Les personnes maîtrisant le langage requis pourront ainsi porter des jugements sur la nature thématique du contenu à partir de son seul titre. Cette approche n'est évidemment pas infaillible, mais elle suffit à nous permettre de trouver des moyens de repérer, classer et trier le contenu extrémiste en vue d'une analyse humaine ultérieure. Nous avons par conséquent orienté et enrichi nos outils de traitement de texte sur cette base.

Nous devions cependant vérifier que notre cadre était valide avant de développer l'algorithme. Nous avons donc créé un livre-code et l'avons testé pour nous assurer de la fiabilité intercodeurs[34]. Pour cela, nous avons sélectionné un échantillon aléatoire de 286 éléments (soit 5 % de l'ensemble des supports) et demandé à trois chercheurs de l'ICSR, tous arabophones et bien au fait des contenus publiés par l'EI, de coder ces supports de façon indépendante. Dans tous les cas sauf 41 (soit 14 % de l'échantillon global), les éléments ont été codés de façon uniforme, les éléments restants faisant l'objet de discussions et étant ultérieurement réconciliés par la cohorte de codage. La plupart de ces incohérences provenaient de désaccords traductionnels concernant l'application syntaxique ou contextuelle précise d'un terme arabe spécifique.

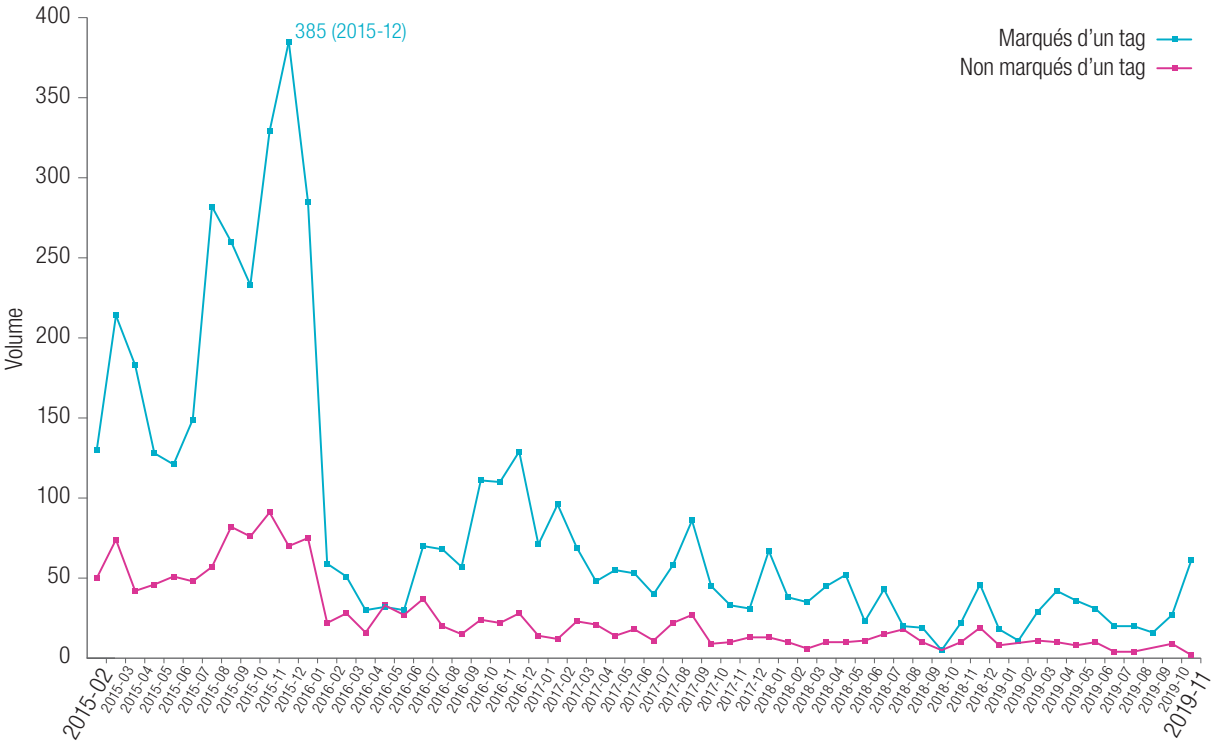
31 Matthew Yglesias, « The (((echo))) explained », Vox, 6 juin 2016. Disponible à l'adresse : https://www.vox.com/2016/6/6/11860796/echo-explained-parentheses-twitter.

32 Rebecca Jennings, « Incels categorize women by personal style and attractiveness », Vox, 28 avril 2018. Disponible à l'adresse : https://www.vox.com/2018/4/28/17290256/incel-chad-stacy-becky.

33 « A parent's guide to the secret language of internet extremists », CBS News, 16 mars 2020. Disponible à l'adresse : https://www.cbsnews.com/news/incels-radicalization-glossary-parents-cbsn-orginals-extremists-next-door/.

34 Moin Syed et Sarah Nelson, « Guidelines for Establishing Reliability When Coding Narrative Data », Emerging Adulthood vol. 3:6, 2015 ; Paul J. Lavrakas, Encyclopedia of Survey Research Methods, Thousand Oaks, CA : Sage Publications, 2008.

Figure 1 : Éléments codés et non codés



Après avoir calibré le cadre à l'aide du processus intercodeurs, nous avons programmé notre algorithme pour qu'il catégorise automatiquement l'ensemble du cache de supports archivés. L'algorithme a appris à procéder ainsi en associant des termes spécifiques à des thématiques précises, processus décrit dans l'infographie de la page 12. Concernant notre ensemble de données, nous avons dans un premier temps identifié les 803 mots qui revenaient le plus fréquemment, afin d'assurer une « couverture » complète du corpus – c'est-à-dire pour veiller à ce que l'ensemble des 6 290 éléments soient pris en compte dans cette phase du processus. Nous sommes parvenus à ce nombre en procédant par approximations successives de façon à trouver la méthode la plus efficace pour obtenir une couverture globale du cache archivé. Ces 803 mots, qui excluaient les noms de lieux et les noms propres, ont ensuite été associés à un ensemble d'une à trois thématique(s) parmi les 22 listées dans notre cadre.

Les mots apparaissant dans le contexte de plusieurs thématiques étaient taggués comme tels. Par exemple, le mot *mahud* (« institution ») a été associé aux thématiques « Éducation », « Entraînement » et « Sensibilisation » parce qu'il est utilisé à la fois pour décrire les écoles, camps militaires et séminaires religieux de l'EI. De même, le terme *riddah* (« apostasie ») a été associé aux thématiques « Opérations », « Exécutions » et « Ordre public » parce qu'il est utilisé dans ces trois contextes. En marquant les 803 mots de cette façon, nous avons pu associer l'ensemble des 6 290 éléments du corpus à une, deux ou trois thématiques.

Le fait de savoir qu'un élément se rapporte à un ensemble d'une à trois thématique(s) différente(s) n'est pas particulièrement utile ; nous avons donc ensuite superposé une série de « super-tags » à ce processus initial. Les super-tags se définissent comme des mots uniques utilisés exclusivement dans le cadre d'une thématique spécifique. Par exemple, un super-tag correspondant à la thématique « Vie commerciale » a été attribué au terme *aswaq* (« marchés ») puisqu'il n'apparaît que dans le contexte des activités commerciales, et nulle par ailleurs. Si, par exemple, les termes « institution » et « marchés » apparaissaient ensemble dans un même titre, la présence du super-tag « marchés » signifierait que le support en question serait automatiquement classé dans la catégorie « Vie commerciale », quels que soient les autres termes présents. Au total, 232 super-tags ont été créés.

Une fois superposé à l'analyse de texte initiale, elle-même fondée sur 803 identifiants linguistiques, le système de super-tags a permis d'assigner une thématique unique à 4 848 éléments (soit 79 % des supports archivés). Nous avons ensuite comparé les résultats produits par l'algorithme à ceux produits par nos chercheurs humains. Une concordance entre l'algorithme et le codage pleinement réconcilié des équipes humaines a été constatée dans 91 % des cas. Bien qu'encore imparfaite, nous avons jugé cette marge d'erreur suffisante pour les recherches actuelles, exploratoires par nature.

L'algorithme n'a pu coder le reste des archives, à savoir 1 432 éléments. Ceux-ci étaient principalement constitués de vidéos et de déclarations audios, dont les titres ne sont généralement pas descriptifs par nature et citent généralement les écritures islamiques ou d'autres sources plus obscures. Cela signifie qu'ils ne contenaient aucun des 803 identifiants linguistiques ni aucun des 232 super-tags, et donc qu'aucune thématique n'a pu leur être assignée.

L'algorithme s'est également heurté à des difficultés lors du traitement de titres composés de mots liés à plusieurs thématiques mais ne contenant aucun super-tag. Par exemple, un reportage photo intitulé « La production des navires de pêche » contient un marqueur « pêche » lié à la thématique « Vie agricole » et un autre, « production », lui-même relié à la thématique « Vie industrielle ». L'algorithme n'a pu appliquer de thématique unique à l'élément en question, en raison de la présence de ces deux marqueurs et de l'absence de super-tag.

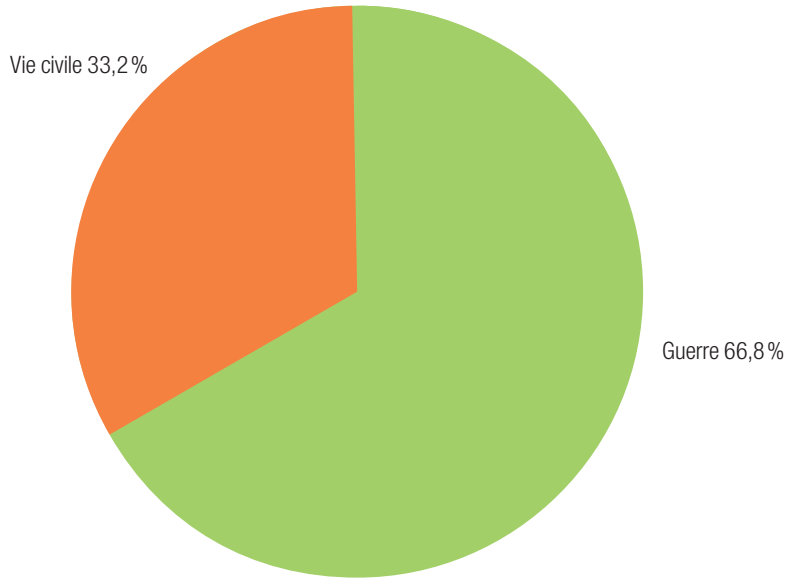
La Figure 1 présente les éléments qui ont pu être codés par l'algorithme et ceux qui ne l'ont pas été. La corrélation entre les deux lignes montre la validité de l'algorithme pour toute la période en question. Notre étude indique que près d'un cinquième du corpus n'a pas été catégorisé, ce qui a représenté une limite importante de notre approche. La résolution de ce problème, qui impliquerait de recourir à un ensemble complètement nouveau d'outils méthodologiques, suppose d'étudier d'autres facteurs que l'analyse linguistique, ce qui dépasse le champ d'application de cette étude. Nous nous pencherons sur cette question plus tard.

4 Conclusions

Triage des données

Une fois terminée l'analyse des 6 290 éléments de notre ensemble de données par notre algorithme, nous avons découvert que ce dernier pouvait aussi contribuer au repérage et au triage des supports. Comme le montre la Figure 2, 66,8 % de l'archive portaient sur des thématiques guerrières, contre 33,2 % consacrées à des thématiques civiles. Sur le plan pratique, ceci a des répercussions claires pour les sociétés technologiques cherchant à analyser le contenu référencé même en se fondant sur les distinctions initiales les plus basiques : supports relatifs à la guerre contre supports relatifs à la vie civile.

Figure 2 : Thématiques générales du contenu marqué d'un tag



Ce point devient plus intéressant lorsque nous cherchons à savoir quels types de supports relatifs à la guerre apparaissent le plus souvent. La Figure 3 montre que les supports les plus fréquents portaient sur la catégorie « Opérations » (30,3 % de notre ensemble de données), suivie des catégories « Résumé », à 10,4 %, « Guerre indirecte », 7,2 % et « Exécutions », 3,2 %. Ainsi, nous constatons que si les exécutions ne constituent pas le type de support violent le plus publié, une entreprise voudra tout de même donner la priorité à l'examen de ces types de contenus compte tenu de leur nature presque toujours explicite et préjudiciable. Les autres catégories, bien que plus fréquentes et liées à la guerre, sont plus susceptibles de contenir des images moins explicites, telles que du matériel militaire, des armements et d'autres objets connexes.

Nous ne cherchons certainement pas à suggérer que les sociétés technologiques emploient des instruments bruts pour détecter le contenu extrémiste violent sur leurs plateformes. Ce n'est clairement pas le cas,

et un vaste ensemble d'instruments complexes est à l'œuvre pour repérer les contenus préjudiciables. Cet outil algorithmique vise à compléter les pratiques existantes en employant un processus d'identification par marqueurs linguistiques pour perfectionner le processus de triage.

Nos résultats, illustrés dans la Figure 4, montrent aussi que les thématiques les plus importantes relevant de la propagande relative

Figure 3 : Thématiques figurant dans la propagande guerrière

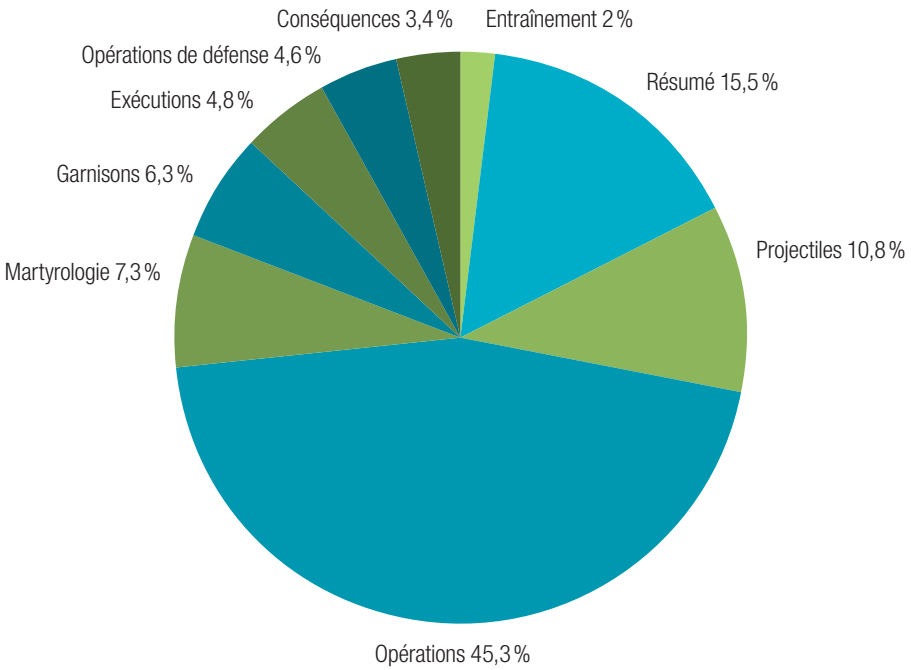
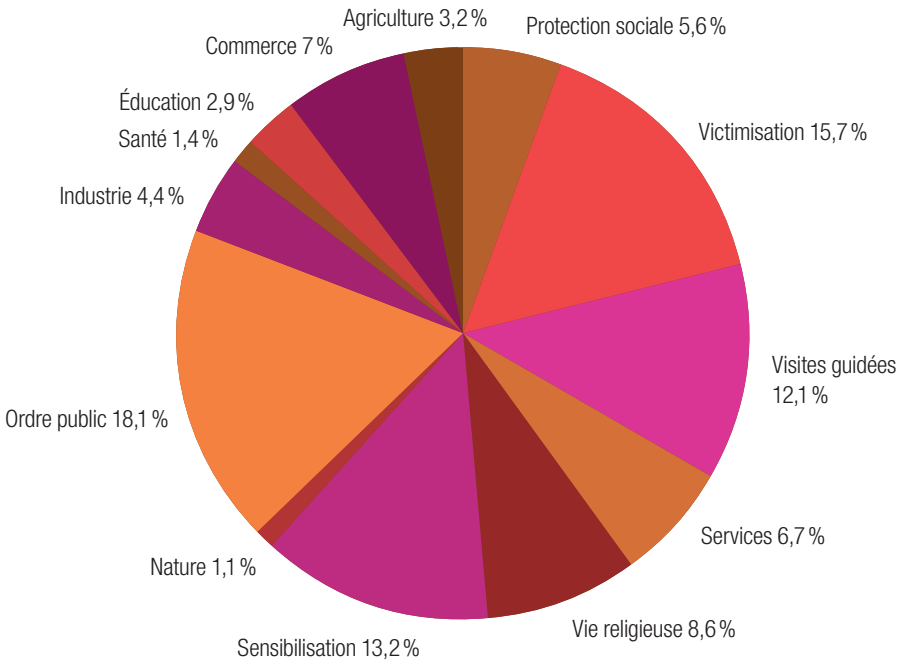


Figure 4 : Thématiques figurant dans la propagande relative à la vie civile



à la vie civile sont : « Ordre public » (6 %) ; « Victimisation » (5,2 %) ; « Sensibilisation » (4,4 %) et « Visites guidées » (4 %). Puisque le contenu consacré à l'ordre public montre à la fois fréquemment des scènes inoffensives de patrouilles de police et des scènes violentes d'exécution et/ou de mutilation de personnes accusées de crimes, il convient de les prioriser pour examen humain urgent immédiatement après la thématique guerrière consacrée aux « Exécutions ». L'examen immédiat est quelque peu moins urgent pour les autres contenus liés à la vie civile puisque, à l'exception de la propagande relative à la victimisation, qui concerne les victimes civiles, ces contenus sont rarement, voire jamais, violents.

L'algorithme nous permet par ailleurs de détecter et visualiser les changements d'emphase ou de priorisation thématique au fil du temps, comme le montre la Figure 5. Cette dernière témoigne d'un éloignement progressif des contenus liés à la vie civile, qui représentaient près de 49 % des éléments recensés en 2015 mais seulement 7 % de ceux listés en 2019. Cela reflète les changements ayant eu lieu sur le terrain, à mesure que l'EI perdait des territoires et laissait de côté ses communications sur l'abondance apparente de sa supposée utopie pour privilégier les messages plus agressifs et guerriers, faisant l'écho des tropes djihadistes traditionnels mentionnant les « pouvoirs croisés » hostiles menant la « guerre à l'Islam ». Sa propagande a donc peu à peu cessé de suggérer aux partisans de migrer vers la Syrie et l'Irak pour soutenir leur « califat » et commencé à les encourager de plus en plus à rester dans leurs pays pour y mener des attaques terroristes.

Figure 5 : Priorités thématiques des contenus marqués d'un tag (par mois)

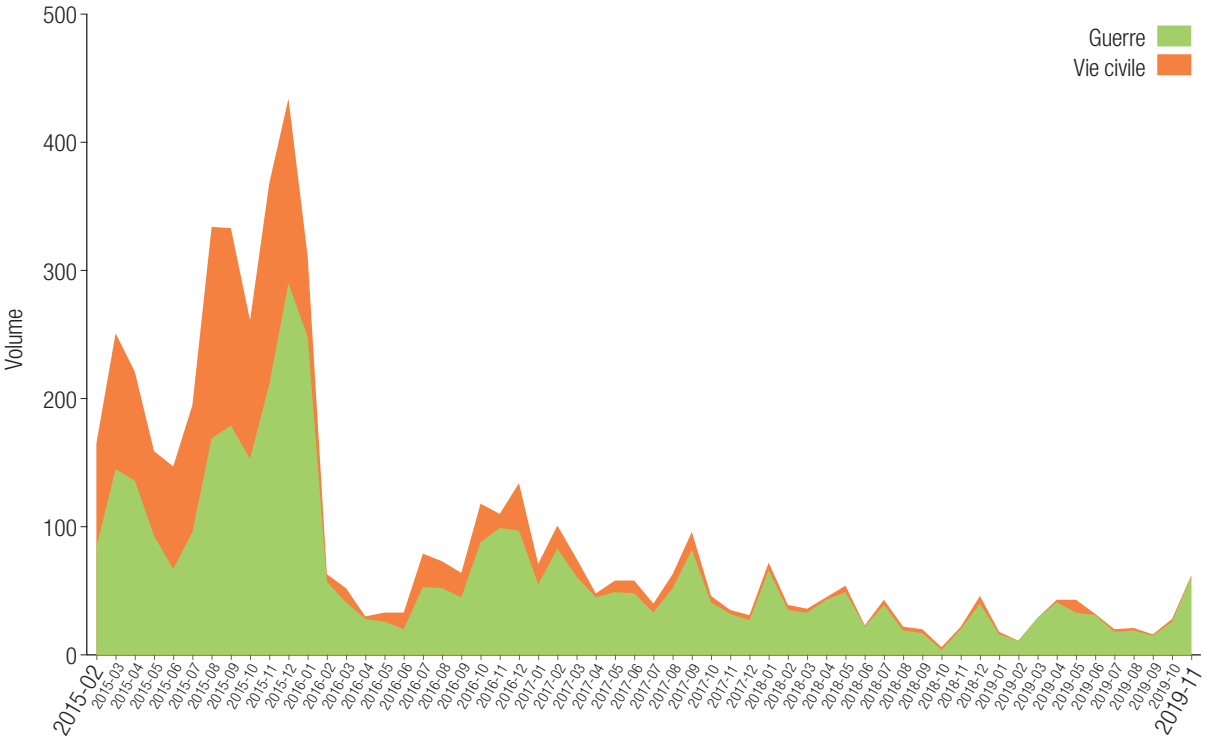


Figure 6 : Contenu relatif à la guerre et à la vie civile en 2015

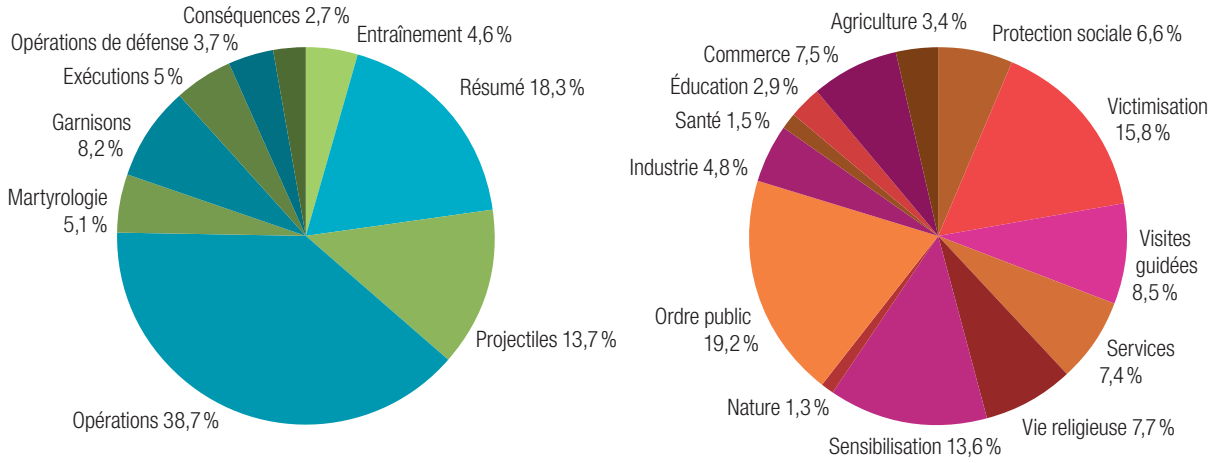
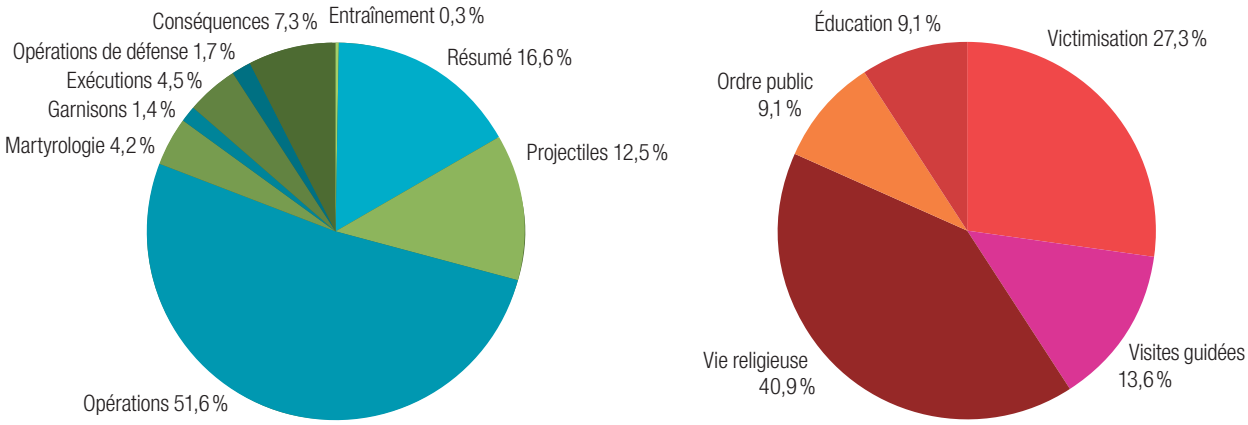


Figure 7 : Contenu relatif à la guerre et à la vie civile en 2019



Parallèlement à cela, le discours général de l'EI s'est peu à peu simplifié. En effet, entre 2016 et 2019, la variabilité thématique de ses communications baissait parallèlement à sa productivité, comme en témoignent les données. Selon les Figures 6 et 7, sa production en 2015 était constituée de l'ensemble des 22 catégories décrites ci-dessus. Dès 2019, toutefois, elle ne traitait plus que de 14 de ces thématiques – dont neuf relatives à la guerre.

Cette transformation et cette simplification simultanées du récit de l'EI ne se limitent en aucun cas à l'archive étudiée. Il s'agit plutôt d'une dynamique qui représente l'évolution des activités officielles de sensibilisation de l'EI tenant compte des nouvelles exigences de la situation sur le terrain. En bref, à mesure que ses territoires se réduisaient et que la nature de la guerre changeait de cap, l'EI est entré dans un nouveau paradigme de guerre d'insurrection. Au lieu de combattre pour des gains matériels et territoriaux, l'organisation est retournée vers des opérations principalement axées sur une consolidation clandestine progressive. À cette fin,

la teneur de sa propagande a changé, portant davantage sur la volonté de signaler sa détermination et une présence permanente que sur la volonté de former de nouveaux cadres ou de provoquer l'indignation mondiale. Ainsi, comme le montrent les Figures 5, 6 et 7, les contenus publiés traitaient moins de la vie civile dans le proto-État de l'EI et plus de ses efforts de guerre.

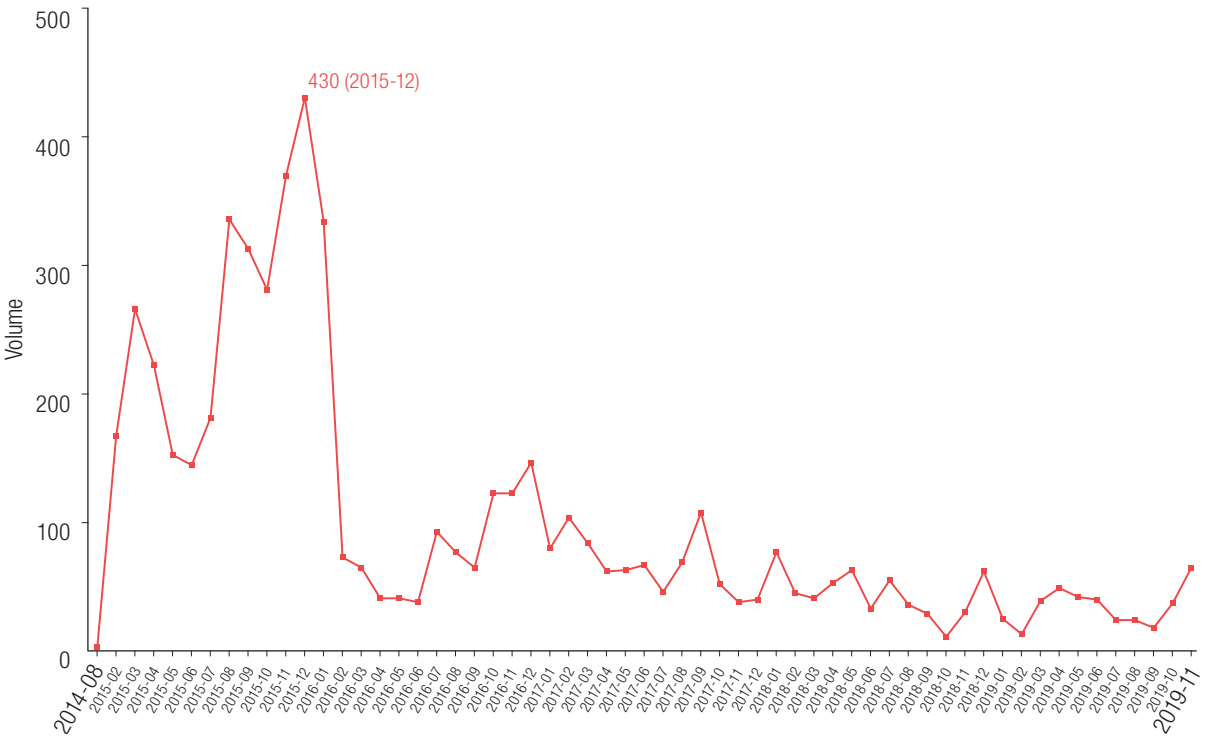
Cet aspect visuel de notre algorithme peut ainsi permettre aux sociétés technologiques de suivre les changements d'orientation et de priorité au fil du temps et de s'adapter lorsqu'elles ont la possibilité d'anticiper la publication d'un type de support par rapport à un autre.

Utilité de l'identification des caractéristiques temporelles

La Figure 8 montre la répartition temporelle de tous les éléments composant l'archive. Chacun des 6 290 éléments a été trié en fonction de sa date de publication et classé en conséquence. Cela nous a permis de mieux comprendre et visualiser le volume de contenu publié au fil du temps pour cet ensemble de données. Dans un cadre réel, cette visualisation serait continue et évoluerait en temps réel, mais est fournie ici à des fins de démonstration.

Le graphique montre que la plupart du contenu a été publiée en 2015, après quoi survient une chute brutale qui se poursuit globalement sur toute la période 2016-2019. Cela nous renseigne probablement davantage sur la nature partielle de l'archive

Figure 8 : Volume total de contenu (par mois)



qu'autre chose. Comme nous le savons, la production de propagande par l'EI s'est sensiblement réduite au cours de la période 2017-2019. Par conséquent, nous pouvons être sûrs que l'archive est plus complète pour ces années. En revanche, pour les années 2015-2016, même si l'archive recense un volume important de propagande, elle ne constitue qu'une petite part de ce qui a été produit au cours de cette période. De plus, elle ne contient qu'un seul élément pour 2014, année durant laquelle l'EI a publié des milliers de contenus. Ce décalage s'explique probablement par le fait que la personne derrière l'archive a commencé à la créer après la suppression de la plupart des contenus officiels de l'EI du web surfacique. Si c'est effectivement le cas, il s'ensuit qu'elle n'a été en mesure d'accéder qu'à une petite partie des supports officiels publiés pendant les années où elle était intéressée.

Quoi qu'il en soit, la baisse d'année en année illustrée dans la Figure 8 s'aligne globalement avec l'évolution des capacités de production médiatique de l'EI après 2015. Comme l'ont déjà montré un certain nombre de chercheurs, 2015 représente l'une des années les plus productives de l'EI en termes de propagande[35].

Cette baisse confirme que le contrôle territorial est largement corrélé aux capacités de production de contenu par l'EI. Lorsque le groupe régnait sur plusieurs millions de personnes et menait plusieurs formes de gouvernance, ses responsables de la propagande étaient non seulement capables de jouir d'une plus grande liberté de mouvement et d'un plus grand accès à des ressources monétaires et humaines, mais se noyaient aussi sous les sujets à traiter[36]. De plus, lorsque l'EI était à son apogée territorialement parlant, il combattait simultanément sur plus d'une dizaine de fronts par des moyens principalement conventionnels[37]. Cela se prête davantage à une couverture propagandiste que les opérations secrètes, ces dernières gagnant en importance lorsque l'EI a été forcé à la défensive.

Les modifications de la composition interne de l'EI ont également joué un rôle dans son déclin médiatique. Naturellement, à mesure que les formes de son insurrection évoluaient, il en a été de même pour ses priorités de sensibilisation, puisque l'organisation cherchait moins à recruter et plus à retenir sa base de soutien locale[38].

Ce point est pertinent pour les objectifs poursuivis dans le cadre du présent rapport parce que notre algorithme peut potentiellement aider les sociétés à détecter l'efficacité de leurs efforts de suppression de contenu. Clairement, certains contenus sont plus difficiles à supprimer ou peuvent être plus facilement dissimulés. De nombreuses études ont prouvé combien l'EI et ses partisans avaient conscience du besoin de se déguiser et de camoufler leur contenu lorsqu'ils opéraient sur les plateformes grand public[39].

35 Voir, par exemple : Milton, « Communication Breakdown » ; Milton, « Down, but Not Out » ; Winter, « Apocalypse, later » ; Nanninga, « Branding a Caliphate in Decline ».
36 Aaron Y. Zelin, « The Islamic State's Territorial Methodology », The Washington Institute for Near East Policy, 2016. Disponible à l'adresse : https://www.washingtoninstitute.org/policy-analysis/view/the-islamic-states-territorial-methodology.
37 Pour comprendre son évolution, voir Ahmed S. Hashim, « The Islamic State's Way of War in Iraq and Syria: From Its Origins to the Post Caliphate Era », Perspectives on Terrorism, vol. 13:1, 2019 : 23–32.
38 Ce phénomène est illustré par le fait que l'EI n'a pas publié de nouvelles vidéos ou d'articles de magazine à travers le centre de presse al-Hayat, sa fondation de propagande en langue étrangère la plus tournée vers l'extérieur, depuis janvier 2019.
39 Voir Berger et Jonathon Morgan, « The ISIS Twitter census ».

En grande majorité, ces efforts n'aboutissent à rien. Néanmoins, notre outil peut aider à repérer les marqueurs linguistiques résilients capables de perdurer, soulignant ainsi la nécessité pour l'analyse de se doter d'une telle dimension temporelle.

Caractéristiques géographiques

Compte tenu de la capacité des outils algorithmiques pour détecter les insignes, logos et autres formes de marquage, nous avons également recherché des caractéristiques géographiques au sein de notre ensemble de données. Encore une fois, la réalisation de cette opération en temps réel permettrait aux sociétés technologiques d'apprendre à leurs outils existants à rechercher un type de contenus plutôt qu'un autre. Cela serait particulièrement utile dans un contexte d'effort de mobilisation de combattants étrangers comme celui déployé par l'EI entre 2013 et 2015 : si le contenu provenant de la destination de ladite mobilisation (dans le cas qui nous intéresse, la Syrie) pouvait être supprimé en priorité, les modérateurs pourraient porter un véritable coup de grâce aux efforts visant à faire de la publicité pour les avantages présumés de l'enrôlement.

Pour cela, nous avons catalogué les supports en fonction de l'endroit d'où ils provenaient de manière ostensible. Nous nous sommes fondés pour cela sur l'unité médiatique de l'EI responsable de la production du contenu concerné. Pour les années couvertes par notre archive, l'EI a opéré un système de production médiatique à trois niveaux, composé de bureaux centraux tels que la Fondation al-Furqan et le centre de presse al-Hayat, d'agences auxiliaires comme l'agence de presse Amaq et le centre de presse Furat, et d'unités médiatiques provinciales comme les centres de presse des provinces du Levant et d'Irak[40]. Avant l'été 2018, le réseau médiatique de l'EI en Syrie et en Irak était subdivisé en 23 bureaux de presse régionaux, l'un pour la province de Raqqa, l'autre pour la province d'Alep, etc.[41].

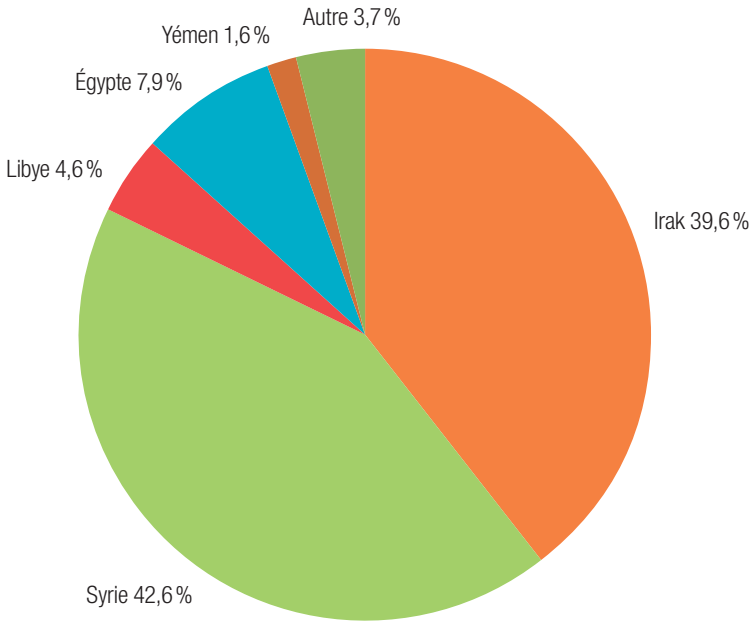
Dans notre archive, 3 831 éléments ont pu être classés comme provenant des bureaux de presse provinciaux. Par ailleurs, 554 autres éléments ont été identifiés comme ayant été préparés par une unité médiatique centrale, principalement la Fondation al-Furqan, le centre de presse al-Hayat, la Fondation al-Itisam, la radio al-Bayan ou al-Naba. Les 1 905 éléments restants, dont la plupart étaient publiés par l'agence de presse Amaq, n'ont pu être classés géographiquement, aucun bureau de presse spécifique à un lieu n'ayant été identifié dans l'index.

La Figure 9 montre que 42,6 % des éléments associés à un bureau de presse spécifique à un lieu provenaient de Syrie, contre 39,6 % produits par les bureaux de presse de l'EI en Irak. Par ailleurs, 7,9 % d'entre eux avaient été préparés par le bureau de presse de la province du Sinaï en Égypte, et les autres provenaient, en ordre décroissant, de Libye, du Yémen, d'Afghanistan, du Nigéria,

40 Abu Abdullah al-Masri, « The Isis papers: A masterplan for consolidating power », The Guardian, 7 décembre 2015. Disponible à l'adresse : https://www.theguardian.com/world/2015/dec/07/islamic-state-document-masterplan-for-power.

41 BBC Monitoring, « Analysis: The Islamic State restructures its 'provinces' a year on from 2017 defeats », 17 octobre 2018. Disponible à l'adresse : https://monitoring.bbc.co.uk/product/c200bdcn.

Figure 9 : Contenu par État



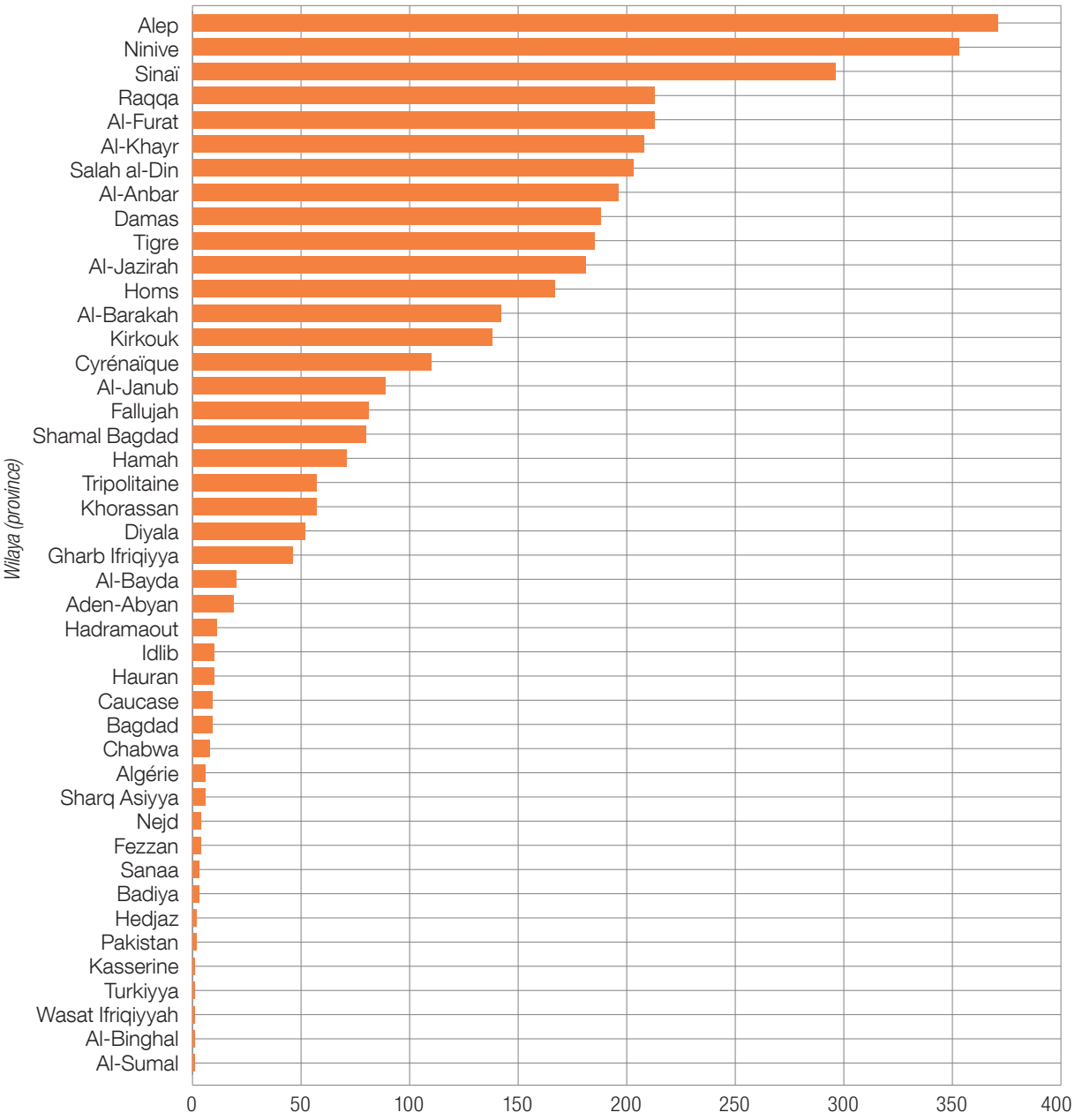
de Russie, d'Algérie, d'Indonésie, du Pakistan, d'Arabie saoudite, de Turquie, de Tunisie, de République démocratique du Congo, du Bangladesh et de Somalie.

La Figure 10 illustre la provenance des éléments par *wilaya*, et non par État. Elle montre que la plupart du contenu de l'archive a été produite par trois seulement des unités médiatiques de l'EI : le bureau de presse de la province d'Alep en Syrie, le bureau de presse de la province de Ninive en Irak, et le bureau de presse de la province du Sinaï en Égypte.

La prévalence du contenu provenant de Syrie et d'Irak s'aligne sur la présence géographique de l'EI pendant les années en question[42]. Toutefois, les supports produits par le bureau de presse de la province du Sinaï sont surreprésentés dans l'archive, ce qui laisse penser que son créateur était plus intéressé ou plus à même d'accéder à des supports liés aux activités de l'EI en Égypte. Cela nous informe davantage sur les origine du créateur que sur la provenance du contenu lui-même.

42 Voir Winter, « Apocalypse, later ».

Figure 10 : Contenu par bureau de presse des *wilayas*



5 Conclusion

Cet article de recherche s'est appuyé sur des techniques de traitement de texte automatisé pour développer une série d'outils analytiques capables d'analyser et de classer la propagande de l'EI à l'échelle. Bien que nous n'ayons étudié que les supports de l'EI, les idées présentées ici sont, en principe, applicables à toute forme d'extrémisme violent. En nous servant d'une archive de 6 290 éléments comme échantillon, nous avons tenté d'automatiser la catégorisation temporelle, géographique et thématique de contenus extrémistes violents. Notre outil a réussi à séparer l'ensemble de données, permettant ainsi de repérer les contenus susceptibles d'être les plus préjudiciables et de leur accorder la priorité par rapport à des contenus moins urgents, bien que très problématiques. L'application généralisée de notre outil a également permis de repérer un certain nombre d'autres dynamiques largement confirmées, liées notamment à la baisse de production, aux caractéristiques géographiques ou à la simplification narrative.

Notre but principal était de concevoir des méthodes capables, lorsque appliquées à des corpus de supports similaires (y compris des corpus beaucoup plus vastes), d'accélérer le processus par lequel les données peuvent être séparées et, le cas échéant, triées pour modération et/ou aiguillage.

D'emblée, nous avons cherché à produire un outil qui pouvait séparer le contenu afin d'aider au triage des supports en vue d'une analyse humaine. Il va de soi que cet outil n'a pas vocation à agir seul. Nous reconnaissons que les sociétés technologiques ont déjà mis en place des systèmes sophistiqués pour détecter et éliminer certains contenus. Dans la grande majorité des cas, ce processus est entièrement automatisé. Pourtant, les analystes humains continueront de jouer un rôle important lorsqu'il conviendra de statuer sur des supports plus litigieux ; c'est dans ce cadre que notre outil a la capacité de simplifier les processus existants.

Nous estimons que les outils tels que celui présenté ici gagneront en importance compte tenu de la diversification des défis rencontrés par les sociétés technologiques. Citons parmi ceux-ci les campagnes de désinformation soutenues par les États, les comportements trompeurs coordonnés ou encore la prolifération de théories du complot. Compte tenu du degré élevé de redondance avec lequel les organisations extrémistes violentes présentent leurs produits – elles utilisent le même ensemble limité de logos, adoptent les mêmes séquences introductives et superposent les mêmes pistes audios –, ces supports constituent des cibles relativement faciles pour les outils de détection automatique. Toutefois, quand il s'agit de limiter les préjudices hors ligne tout en maintenant la liberté de parole, la seule capacité de détection devient alors insuffisante. Ces supports doivent aussi être traités et compris, à la fois dans un contexte plus large et au niveau de l'intention sous-tendant leur production, ce que, en l'état actuel des choses, seuls les modérateurs humains sont capables de faire.

Les outils développés dans le cadre de ces recherches permettraient dans une certaine mesure de simplifier ce processus de triage en vue d'un examen humain ultérieur.

Contexte politique

Cette section a été rédigée par Armida van Rij et Vivienne Moxham-Hall, toutes deux adjointes de recherche au Policy Institute du King's College, à Londres. Elle fournit un aperçu du contexte politique pertinent dans lequel s'inscrit ce rapport.

Introduction

L'utilisation abusive d'Internet par les terroristes et les personnes ayant des opinions extrémistes est devenue de plus en plus problématique ces 10 dernières années. Les réseaux sociaux et autres plateformes de sociétés technologiques servent à communiquer de la propagande terroriste, à attirer de nouvelles recrues dans les organisations terroristes ou à inciter à la violence. La classification de ces contenus préjudiciables ou, plus simplement, les décisions concernant ce qui est illicite ou devrait être retiré des plateformes, se sont avérées poser de véritables problèmes tant pour les sociétés technologiques que pour les décideurs politiques. Ce processus mène inévitablement à une prise de décisions concernant ce qui doit être considéré comme « extrémiste » ou « terroriste », les ressources et les capacités requises pour modérer les heures de nouveaux contenus téléchargés toutes les heures, ou encore la méthode à adopter pour bloquer les contenus préjudiciables tout en assurant la liberté d'expression, de pensée et de débat. Cet ensemble de facteurs a fait de la lutte contre les contenus terroristes en ligne et de la classification de ce qui est permis et de ce qui ne l'est pas un défi très difficile à relever, mais avec de véritables conséquences. La diffusion en direct sur Facebook d'une attaque terroriste perpétrée sur deux mosquées à Christchurch, en Nouvelle-Zélande, a été visionnée au moins 4 000 fois avant d'être retirée, ce qui n'a pas empêché le retéléchargement et la diffusion de la vidéo sur les réseaux sociaux, y compris Facebook[43].

Afin de mieux comprendre comment les pays luttent contre le problème des contenus terroristes en ligne, comment les classer au mieux et quelles mesures les États ont prises pour les supprimer, nous avons examiné le paysage politique relatif aux contenus terroristes en ligne de neuf législateurs. Ceux-ci n'ont pas été choisis au hasard, mais sont membres du Comité consultatif indépendant (IAC) du Forum mondial de l'Internet contre le terrorisme (GIFCT). L'IAC lui-même se compose de 21 représentants, issus de sept gouvernements, de deux organisations internationales et de 12 organisations de la société civile (OSC), représentant un éventail de compétences. Les législateurs en question sont les suivants :

- Le gouvernement du Canada
- Le gouvernement de la France
- Le gouvernement du Ghana

43 « Facebook: New Zealand attack video viewed 4,000 times », BBC News, 19 mars 2019. Disponible à l'adresse : https://www.bbc.co.uk/news/business-47620519.

- Le gouvernement du Japon
- Le gouvernement de la Nouvelle-Zélande
- Le gouvernement du Royaume-Uni
- Le gouvernement des États-Unis
- L'Union européenne
- La Direction exécutive du Comité contre le terrorisme (DECT) du Conseil de sécurité des Nations Unies

Pour chacun de ces législateurs, ce rapport répondra aux questions suivantes : i) Qui sont les principales parties prenantes ? ; ii) À quelles difficultés doivent-elles faire face ? ; iii) Quels sont les développements politiques et principales lois en vigueur dans chaque juridiction ? et iv) Que prévoient les principales parties prenantes pour l'avenir ?

Concernant le premier sujet de recherche, il existe bien entendu de nombreuses parties prenantes au sein de chaque entité, des ministères gouvernementaux et fournisseurs de services Internet aux entreprises de médias sociaux, organisations de la société civile, en passant par le grand public. Nous ne chercherons pas à recenser toutes les parties prenantes liées à un État ou à une organisation donné(e) dans ce rapport. Nous axerons plutôt nos efforts sur les principales parties prenantes ayant des responsabilités capitales en matière de lutte contre la publication de contenu extrémiste en ligne.

Avant d'effectuer une évaluation pays par pays, il est possible de décrire plusieurs difficultés auxquelles chacun est confronté. Une difficulté à laquelle font face notamment les pays occidentaux est la nécessité d'équilibrer le droit à la liberté d'expression et la protection des populations. Les défenseurs de la liberté d'expression ont, par le passé, critiqué les gouvernements pour avoir légiféré en faveur de la suppression de vidéos extrémistes en ligne, en avisant que cela pouvait porter atteinte à la liberté d'expression et équivaloir, au bout du compte, à de la censure.

De même, les plateformes plus modestes sont de plus en plus utilisées pour héberger des contenus extrémistes, mais n'ont pas les capacités de suivre, analyser ou supprimer les contenus illicites. Si les plus grandes entreprises de réseaux sociaux et informatiques ont plus de ressources et sont donc mieux à même de faire face à ces difficultés, cela s'est avéré plus compliqué pour les organisations plus modestes.

Canada

Le ministère de la Sécurité publique et de la Protection civile (Sécurité publique Canada) a appuyé le développement de la Plateforme d'analyse des contenus à caractère terroriste (TCAP). Statistique Canada, le bureau national de statistique du Canada, traque le terrorisme dans le pays. Le Centre canadien d'engagement communautaire et de prévention de la violence dirige les efforts du Canada pour lutter contre la radicalisation en travaillant avec le gouvernement, la société civile, les forces de l'ordre et des organisations internationales.

En 2017, le Centre canadien a lancé une Stratégie nationale sur la lutte contre la radicalisation menant à la violence, qui porte sur la prévention précoce, la prévention auprès des personnes à risque et le désengagement des idéologies violentes[44]. Le Canada a déjà recours à la catégorisation pour traquer le terrorisme. Plus spécifiquement, pour traquer le terrorisme, Statistique Canada se fonde sur 13 codes d'infractions de la Déclaration uniforme de la criminalité (DUC)[45]. À l'heure actuelle, ces codes ne définissent pas les activités en ligne comme une catégorie spécifique. Pourtant, selon une enquête menée en 2016 auprès de 13 services de police municipale canadiens, 40 % d'entre eux n'avaient pas connaissance des codes DUC, et près de la moitié d'entre eux avaient du mal à déterminer quel code utiliser dans certains cas[46]. Cela met en lumière les difficultés plus vastes auxquelles font face les États et forces de l'ordre pour définir les activités terroristes.

Justin Trudeau, le Premier ministre canadien, s'est joint à l'Appel de Christchurch en 2019, un engagement mondial visant à supprimer les contenus terroristes et extrémistes violents en ligne. Dans le cadre du respect de ces engagements, le Canada a engagé Tech Against Terrorism pour développer la Plateforme d'analyse des contenus à caractère terroriste (TCAP), qui vise à permettre aux sociétés technologiques plus modestes de lutter plus efficacement contre ce type de contenus[47]. Il s'agit d'une plateforme centralisée de contenu vérifié à caractère terroriste conçue pour aider les sociétés technologiques modestes à repérer les contenus à caractère terroriste et à prendre des décisions en matière de modération de contenu[48]. Elle automatise la détection et l'analyse des contenus vérifiés à caractère terroriste sur les plateformes et constitue le premier ensemble de données sur les contenus vérifiés à caractère terroriste, et également le plus étendu. Son deuxième rôle est de permettre la réalisation de recherches universitaires en toute sécurité, en aidant à mieux comprendre la menace posée par le terrorisme et les contenus extrémistes. Cela permettra de concrétiser l'engagement pris dans le cadre de l'Appel de Christchurch visant à aider les plateformes en ligne plus modestes à renforcer leurs capacités de retrait des contenus extrémistes en ligne[49].

44 Voir « Canada: Extremism & counter-extremism', Counter Extremism Project, 23 juin 2020. Disponible à l'adresse : https://www.counterextremism.com/countries/canada.

45 Patrick McCaffery *et al.*, « Classification and Collection of Terrorism Incident Data in Canada », Perspectives on Terrorism, vol. 10:5, 2016 : 43. Disponible à l'adresse : https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2016/issue-5/505-classification-and-collection-of-terrorism-incident-data-in-canada-by-patrick-mccaffery-lindsy-richardson-jocelyn-j.-belanger.pdf.

46 Ibid.

47 « Press release : Tech Against Terrorism award grant by the Government of Canada to build Terrorist Content Analytics Platform », Tech Against Terrorism, 27 juin 2019. Disponible à l'adresse : https://www.techagainstterrorism.org/2019/06/27/press-release-tech-against-terrorism-awarded-grant-by-the-government-of-canada-to-build-terrorist-content-analytics-platform/.

48 « Update : Initial version of the Terrorist Content Analytics Platform to include far-right terrorist content », Tech Against Terrorism, 2 juillet 2020. Disponible à l'adresse : https://www.techagainstterrorism.org/2020/07/02/update-initial-version-of-the-terrorist-content-analytics-platform-to-include-far-right-terrorist-content/.

49 Gouvernement du Canada, Sécurité publique Canada, « Le gouvernement du Canada annonce des initiatives pour contrer l'extrémisme violent et le contenu terroriste en ligne », communiqué de presse, 26 juin 2019. Disponible à l'adresse : https://www.canada.ca/fr/securite-publique-canada/nouvelles/2019/06/le-gouvernement-du-canada-annonce-des-initiatives-pour-contrer-lextremisme-violent-etlecontenu-terroriste-en-ligne.html.

Union européenne

L'UE a mis en place un certain nombre de législations applicables au contenu extrémiste en ligne. La Directive 2017/541 relative à la lutte contre le terrorisme vise à harmoniser les législations des États membres sur l'incrimination des infractions terroristes. Plus spécifiquement, son article 21 impose aux États membres de prendre des mesures garantissant la suppression rapide des contenus en ligne, en leur laissant le choix des mesures à prendre[50]. Cette disposition couvre le matériel incitant au terrorisme, mais aussi celui utilisé à des fins de recrutement ou de formation, et tient compte d'autres infractions liées aux activités terroristes. Certains États membres ont mis en place des procédures de notification et d'action applicables aux plateformes en ligne dans le cadre de leur législation nationale[51]. Parmi eux, citons la France, l'Allemagne et l'Espagne. En vertu de l'article 14.3 de la Directive sur le commerce électronique, les États membres doivent mettre en place des procédures régissant le retrait des informations ou les actions visant à en rendre l'accès impossible[52]. Ils conservent toutefois le pouvoir d'interpréter cet article comme ils l'entendent, ce qui entraîne des différences dans les champs d'application au sein de l'UE. En 2016, la Commission européenne s'est également mise d'accord avec Microsoft, Twitter, Facebook et YouTube pour adopter un code de conduite volontaire visant à combattre les discours de haine illicites en ligne[53].

L'UE est confrontée à de nombreux défis en matière de classification et de lutte effective contre les contenus à caractère terroriste en ligne. Le premier, qui est également le plus fondamental, est la nécessité de se mettre d'accord et de mettre en œuvre des normes et procédures communes dans l'ensemble des 27 États membres de l'Union après la période de transition du Brexit. Ce défi au cœur des travaux de l'UE a conduit jusqu'ici à un paysage politique fragmenté.

Par exemple, il n'existe aucune règle à l'échelle européenne concernant les procédures de notification et d'action pour les contenus illicites hébergés sur les plateformes en ligne. Seuls quelques États membres ont introduit des cadres réglementaires relatifs auxdites procédures. Parmi eux, certains l'ont fait dans l'esprit de la Directive sur le commerce électronique, comme la France, le Royaume-Uni ou la Hongrie, tandis que d'autres, comme l'Espagne, sont intervenus dans le cadre d'un instrument juridique autonome[54]. Dans certains pays, les interventions visant à mesurer, bloquer, filtrer et retirer le contenu en ligne ne sont pas conformes à l'article 10 de la Convention européenne des droits de l'homme, selon lequel les restrictions en matière de liberté d'expression doivent être « légales, légitimes et nécessaires »[55].

50 Commission européenne, « Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne », 2018/0331 (COD), 2018(a) : 3. Disponible à l'adresse : https://eur-lex.europa.eu/resource.html?uri=cellar:dc0b5b0f-b65f-11e8-99ee-01aa75ed71a1.0002.02/DOC_1&format=PDF.

51 Commission européenne, « Impact Assessment accompanying the document Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online », 2018(b) : 122. Disponible à l'adresse : https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.

52 Ibid.: 10

53 Conseil de sécurité des Nations Unies, Direction exécutive du Comité contre le terrorisme (DECT), « More support needed for smaller technology platforms to counter terrorist content », CTED trends alert, novembre 2018 : 4. Disponible à l'adresse : https://www.un.org/sc/ctc/wp-content/uploads/2019/01/CTED-Trends-Alert-November-2018.pdf.

54 Commission européenne, 2018(b) : 122.

55 Conseil de l'Europe, « Étude comparative sur le blocage, le filtrage et le retrait de contenus illégaux sur Internet », 20 décembre 2015. Disponible à l'adresse : https://edoc.coe.int/fr/medias/7286-pdf-etude-comparative-sur-le-blocage-le-filtrage-et-le-retrait-de-contenus-illegaux-sur-internet.html.

Figure 11 : Initiatives existantes au sein des États membres de l'UE relatives aux procédures de notification et d'action[56]

EM	Acte juridique	Législation en cours d'élaboration	Contenu illicite concerné
BE	Aucun	Notification et action (N&A)	
DE	Loi sur l'exercice des droits sur les réseaux sociaux (NetzDG)		Discours haineux
DK	Aucun		
ES	Décret royal 1889/2011 régissant le fonctionnement de la Commission de protection des droits de propriété intellectuelle – modifié par la loi n° 21/2014		Violation des droits d'auteur
FI	Loi n° 2002/458 relative à la prestation de services de la société de l'information		Violation des droits d'auteur
FR	Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique		Uniquement applicable aux contenus manifestement illicites
HU	Loi CVIII de 2001 relative à certains aspects du commerce électronique et aux services de la société de l'information		Violation des droits de propriété intellectuelle
IT	Décision de l'AGCOM concernant l'application des droits d'auteur en ligne, 680/13/CONS, 12 décembre 2013		Violation des droits d'auteur
LT	Règlement sur le déni d'accès à des informations acquises, créées, modifiées ou utilisées illégalement, approuvé par la Résolution gouvernementale n° 881 du 22 août 2007		Horizontal
PL	Aucun	Potentiellement en train de travailler sur une initiative de N&A	
PT	Décret-loi n° 7/2004 du 7 janvier, Lei do Comércio Electrónico, 7 janvier 2004		Règlement préliminaire extrajudiciaire des litiges
SE	Loi sur la responsabilité des panneaux d'affichage électroniques		Violation des droits d'auteur, contenu à caractère raciste
UK	Règlements S.I. 2002/2013 sur le commerce électronique		Horizontal – fixe les conditions de la notification

56 Figure extraite de Commission européenne, « Impact Assessment accompanying the document Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online », 2018(b) : 123. Disponible à l'adresse : https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.

En 2018, la Commission européenne a proposé une législation visant à assurer le retrait des contenus extrémistes dans l'heure suivant leur mise en ligne. Cette législation aurait également fait reposer sur les plateformes la responsabilité du devoir de diligence vis-à-vis de leurs utilisateurs[57]. Cette proposition de règlement aurait également obligé les États membres à s'assurer que leurs autorités et forces de l'ordre aient les capacités requises pour combattre les contenus à caractère terroriste en ligne[58]. Cette mesure a toutefois été mal accueillie par certains États membres et députés européens. Sa définition du terrorisme était trop vaste et l'exigence de retrait dans l'heure a été considérée comme trop limitative et comme instaurant potentiellement une culture de la censure[59]. Elle a été amendée par le Parlement européen afin de tenir compte des principales préoccupations de ses détracteurs[60]. Les négociations relatives à la réglementation européenne des contenus à caractère terroriste en ligne entre le Conseil, la Commission et le Parlement ont été suspendues par la pandémie de coronavirus. En juillet 2020, la Commission européenne a annoncé la mise en place de lignes directrices non contraignantes relevant de la Directive « Services de médias audiovisuels » de 2018. Ces lignes directrices demandent aux plateformes en ligne de s'assurer que leurs utilisateurs sont protégés contre les discours haineux et que les mineurs sont protégés contre les contenus préjudiciables[61]. L'UE travaille actuellement sur une loi relative aux services numériques, qui visera à « réguler l'écosystème en ligne dans toute une série de domaines, y compris … les contenus choquants »[62].

France

Les officiers de la police nationale responsables de la lutte contre les infractions numériques sont chargés de faire appliquer les lois. De même, l'Office central de lutte contre la criminalité liée aux technologies de l'information et de la communication est chargé de vérifier si les plateformes antérieurement bloquées pour avoir hébergé des contenus à caractère terroriste continuent d'afficher lesdits contenus[63]. Ce service, essentiellement chargé de lutter contre la cybercriminalité, est également celui auquel les contenus illicites sont signalés.

En France, l'article 6 de la loi n° 2004-575 du 21 juin 2004 détermine les responsabilités des plateformes d'hébergement. La loi dispose que si les entreprises ne sauraient être responsables des activités ou informations postées sur leur plateforme à moins d'avoir eu effectivement connaissance de leur contenu manifestement illicite et de ne pas l'avoir retiré, elles sont toutefois chargées d'intervenir suite au signalement d'un contenu illicite. Les plateformes sont tenues d'avoir un mécanisme de signalement permettant à tout

57 Commission européenne, 2018(b) : 3.
58 Ibid. : 4.
59 Faiza Patel, « EU 'Terrorist Content' Proposal Sets Dire Example for Free Speech Online », Just Security, 5 mars 2019. Disponible à l'adresse : https://www.justsecurity.org/62857/eu-terrorist-content-proposal-sets-dire-free-speech-online/.
60 Al Cuddy, « EU struggles over law to tackle spread of terror online », BBC News, 17 avril 2019. Disponible à l'adresse : https://www.bbc.co.uk/news/world-europe-47962394.
61 « Facebook, YouTube, Twitter to face same EU rules on hateful content as broadcasters », EURACTIF, 3 juillet 2020. Disponible à l'adresse : https://www.euractiv.com/section/digital/news/facebook-youtube-twitter-to-face-same-eu-rules-on-hateful-content-as-broadcasters/.
62 Samuel Stolton, « Platform clamp down on hate speech in run up to Digital Services Act », EURACTIF, 23 juin 2020. Disponible à l'adresse : https://www.euractiv.com/section/digital/news/platforms-clamp-down-on-hate-speech-in-run-up-to-digital-services-act/.
63 Commission européenne, 2018(b) : 117.

individu de signaler du contenu. S'il est signalé, le contenu illicite doit être retiré dans les 24 heures, ou les autorités peuvent notifier aux plateformes les adresses électroniques des services de communication au public en ligne contrevenant à la loi, plateformes qui doivent empêcher sans délai l'accès à ces adresses[64].
En février 2015, au lendemain des attentats terroristes perpétrés contre *Charlie Hebdo*, la France a également introduit une loi octroyant à la police nationale le pouvoir de retirer, sans mandat judiciaire, tout site Internet contenant des contenus illicites[65].
Le décret 2015-253 adopté par la suite, en mars 2015, s'applique spécifiquement aux actes de provocation et/ou d'incitation au terrorisme ainsi qu'aux actes de glorification du terrorisme[66].

Plus récemment, la France a adopté une loi obligeant les sociétés technologiques à retirer les contenus extrémistes dans l'heure suivant la réception d'un ordre de la police, au risque de se voir imposer une amende allant jusqu'à 4 % de leurs revenus totaux. L'autorité française de régulation de l'audiovisuel, le Conseil supérieur de l'audiovisuel (CSA), aura le pouvoir d'imposer des amendes aux sociétés technologiques contrevenantes. L'un des principaux défis pour certaines plateformes est le retéléchargement de contenus précédemment détectés et supprimés. Pourtant, en France, la tendance amorcée par la Cour de Cassation semble être l'absence d'obligation pour les plateformes de prévenir la réapparition de contenu précédemment retiré – impliquant ainsi une portée limitée du devoir de diligence des entreprises de médias sociaux[67].
Les procédures de notification et d'action se limitent au contenu illicite, qui peut inclure les contenus extrémistes sans pour autant en englober la totalité.

En particulier, les conséquences du retrait de contenu semblent inquiéter les groupes de défense de la liberté de parole. Deux séries de problématiques ont été identifiées : tout d'abord, cette exigence est difficilement mise en œuvre par les sociétés technologiques plus modestes, qui n'ont pas les ressources pour contrôler des volumes importants de contenu à toute heure du jour et de la nuit[68].
Pour respecter cette législation et échapper aux amendes, elles seront peut-être amenées à recourir à la censure. Ensuite, cette loi risque d'être utilisée pour censurer le militantisme politique. Cela illustre notamment la difficulté que pose la qualification du contenu extrémiste, les limites étant souvent très fluides[69].

64 « Online terrorist propaganda: France and UK put internet giants in the cross-hairs », Jones Day, juillet 2017. Disponible à l'adresse : https://www.jonesday.com/en/insights/2017/07/online-terrorist-propaganda-france-and-uk-put-internet-giants-in-the-cross-hairs.
65 Commission européenne, 2018(b) : 117.
66 Ibid.
67 Ibid.: 10.
68 « France gives online firms one hour to pull 'terrorist' content », BBC News, 14 mai 2020. Disponible à l'adresse : https://www.bbc.co.uk/news/technology-52664609.
69 Benedict Wilkinson et Armida van Rij, « An analysis of the Commission for Countering Extremism's call for evidence: Report 1 – Public understanding of extremism », Policy Institute, King's College London, 9 décembre 2019.

Ghana

Le Ghana a créé une branche au sein des forces publiques chargée de la cybercriminalité qui bénéficie des contributions d'Europol, d'Interpol et des fournisseurs de services Internet[70]. Les informations disponibles publiquement ne permettent pas toutefois de définir clairement la forme que prend la lutte menée par le pays contre le terrorisme. On ne sait pas exactement non plus si des efforts sont déployés pour lutter contre les contenus à caractère terroriste en ligne ni, si c'est le cas, quels défis et discussions en sont ressortis.

Japon

Le Centre national de préparation aux incidents et de stratégie en matière de cybersécurité, créé en 2015, oblige les entreprises d'infrastructures, telles que les services publics (gaz, eau, électricité), les réseaux de transport et les institutions financières, à renforcer leurs mesures de cybersécurité de façon proactive.

En 2017, le Japon a adopté un nouveau projet de loi controversé ciblant les ententes en vue de commettre des actes de terrorisme et d'autres infractions graves, qui liste 277 infractions, y compris la reproduction de musique et la cueillette de champignons dans les forêts protégées[71]. Ces nouvelles lois ont été critiquées pour leur atteinte aux libertés civiles et leur application vague. La déclaration des dirigeants du sommet du G20 d'Osaka a également instamment demandé aux sociétés technologiques de ne pas autoriser l'utilisation abusive de leurs plateformes à des fins terroristes[72].

Nouvelle-Zélande

La réponse globale de la Nouvelle-Zélande à la lutte contre le terrorisme implique une coordination entre différents ministères du gouvernement, les communautés et des organisations du secteur privé. La gouvernance de haut niveau est assurée par le Comité du Cabinet chargé des relations extérieures et de la sécurité et par le Conseil de la sécurité et du renseignement. La stratégie globale de la Nouvelle-Zélande est présentée dans leur plan de Stratégie de lutte contre le terrorisme publié en février 2020[73].

Le Groupe de la sécurité numérique au sein du ministère de l'Intérieur est chargé de réguler les contenus publiés en ligne, tels que films, vidéos et autres publications pouvant être classées comme visant à « causer un préjudice ». Le préjudice se définit comme tout contenu en ligne qui « décrit, représente, exprime ou

70 Kristina Cole *et al.*, « Cybersecurity in Africa: An Assessment ». 2008. Disponible à l'adresse : https://www.researchgate.net/profile/Seymour_Goodman/publication/267971678_Cybersecurity_in_Africa_An_Assessment/links/54e93dca0cf25ba91c7ef580/Cybersecurity-in-Africa-An-Assessment.pdf.

71 « Japan passes controversial anti-terror conspiracy law », BBC News, 15 juin 2017. Disponible à l'adresse : https://www.bbc.co.uk/news/world-asia-40283730 ; Robin Harding, « Japan passes pre-emptive anti-terrorism law », The Financial Times, 15 juin 2017. Disponible à l'adresse : https://www.ft.com/content/75130598-5181-11e7-bfb8-997009366969.

72 Gouvernement du Japon, « G20 Osaka Leaders' statement on preventing exploitation of the internet for terrorism and violent extremism conducive to terrorism (VECT) ». Disponible à l'adresse : https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/en/documents/final_g20_statement_on_preventing_terrorist_and_vect.html.

73 Gouvernement de Nouvelle-Zélande, Comité des fonctionnaires chargés de coordonner la sécurité intérieure et extérieure, Comité de coordination de la lutte contre le terrorisme, « Countering terrorism and violent extremism national strategy overview », février 2020. Disponible à l'adresse : https://dpmc.govt.nz/sites/default/files/2020-02/2019-20%20CT%20Strategy-all-final.pdf.

traite de toute autre façon des sujets comme le sexe, l'horreur, le crime, la cruauté ou la violence d'une façon telle que sa disponibilité est susceptible de porter préjudice à l'intérêt public »[74]. L'identification de ces supports semble largement dépendre de systèmes de signalement externes ou de systèmes internes de classification de contenu préexistants sur les plateformes de médias sociaux ; le gouvernement lui-même ne semble pas diriger son propre processus de filtrage. La loi régissant la définition du contenu répréhensible identifié en Nouvelle-Zélande est la loi de 1993 sur les films, vidéos, publications et classifications, mise à jour par un amendement de 2019 faisant suite à l'attentat terroriste de Christchurch, qui a été filmé en direct et mis à la disposition des internautes. La vidéo de l'attentat a été volontairement supprimée par les fournisseurs de services Internet et de réseaux sociaux immédiatement après que le Groupe de la sécurité numérique leur a signalé la présence d'images répréhensibles. Le non-respect de la loi peut entraîner une peine de prison pouvant aller jusqu'à 14 ans et des amendes allant jusqu'à 200 000 dollars néozélandais.

La diffusion en direct sans précédent de l'assassinat de 50 personnes à Christchurch en 2019 a donné naissance à l'Appel de Christchurch, cofondé par la Nouvelle-Zélande et la France et visant à « supprimer les contenus terroristes et extrémistes violents en ligne »[75]. Cet appel a rassemblé 48 pays, dont 31 nouveaux, autour de la refonte du Forum mondial de l'Internet contre le terrorisme (GIFCT). Le GIFCT a élaboré un protocole commun de réponse à la crise en 2019, qui a été testé par Google en Nouvelle-Zélande pour assurer une gestion coordonnée des impacts en ligne de tout attentat extrémiste[76].

Royaume-Uni

Au Royaume-Uni, le ministère de l'Intérieur est responsable des lois et politiques de lutte contre le terrorisme. Avec ce mandat, il travaille en étroite collaboration avec le Centre de la sécurité nationale, créé en 2016 et faisant partie intégrante du quartier-général des communications du gouvernement. Le ministère du Numérique, de la Culture, des Médias et des Sports a la responsabilité de maintenir un Internet sûr et ouvert[77]. Le ministère de l'Intérieur travaille également en étroite collaboration avec des tiers pour développer des technologies spécifiques contribuant à la prévention et à la lutte contre les contenus extrémistes violents en ligne. Parmi ces parties prenantes se trouvent les sociétés technologiques et centres d'intelligence artificielle qui développent des outils de lutte contre les contenus extrémistes, comme Faculty (anciennement ASI Data Science). Un autre groupe est constitué de plateformes utilisées abusivement pour télécharger et partager des contenus illicites, comme Facebook, Twitter et Microsoft. Il existe d'autres organes indépendants, tels que la Commission de lutte contre l'extrémisme qui fait partie du ministère de l'Intérieur, ou encore le Conseil du

74 Gouvernement de Nouvelle-Zélande, ministère de l'Intérieur, « Objectionable and restricted material ». Disponible à l'adresse : https://www.dia.govt.nz/Censorship-Objectionable-and-Restricted-Material.
75 Voir https://www.appeldechristchurch.com/.
76 GIFCT, Joint Tech Innovation. Disponible à l'adresse : https://www.gifct.org/joint-tech-innovation/.
77 Chambres du Parlement, Clare Lally et Rowena Bermingham, « Online extremism », UK Parliament POST, 6 mai 2020 : 3. Disponible à l'adresse : https://post.parliament.uk/research-briefings/post-pn-0622/.

Royaume-Uni pour la sécurité d'Internet, qui vise à lutter contre les préjudices causés en ligne, y compris l'extrémisme et les infractions motivées par la haine[78].

Le ministère de l'Intérieur a mis en place un Service d'orientation de la police chargé de la lutte contre le terrorisme sur Internet (CTIRU), auquel toute personne peut signaler des contenus suspects. Le CTIRU a assuré la suppression de plus de 300 000 contenus terroristes en ligne[79]. En février 2018, le gouvernement britannique a annoncé le développement de nouvelles technologies visant à analyser avec précision les supports vidéos afin de déterminer s'il pouvait s'agir de propagande de l'EI[80]. Cette technologie, fondée sur l'apprentissage automatique, a été conçue spécialement pour les sociétés technologiques plus modestes qui, contrairement aux plus grandes entreprises comme Facebook et YouTube, n'ont pas la capacité de développer de tels outils elles-mêmes. Le classificateur de contenu est installé dans le flux de téléchargement d'une plateforme, ce qui signifie qu'une vidéo est rejetée avant même d'atteindre ladite plateforme. Il s'agit d'une approche préventive, qui permet d'agir avant que la vidéo potentiellement préjudiciable ne soit mise en ligne. La Secrétaire d'État à l'Intérieur de l'époque, Amber Rudd, n'a pas exclu l'adoption d'une future législation obligeant les entreprises dénuées de ressources de contrôle effectives à utiliser cet outil. Le Royaume-Uni a par ailleurs mis en place une loi pour faire face au contenu illicite en ligne au travers d'un cadre de notification et d'action[81].

Dans un Livre blanc sur les préjudices en ligne publié conjointement par les ministères de l'Intérieur et du Numérique, de la Culture, des Médias et des Sports l'année dernière, le gouvernement a proposé de créer un organisme de réglementation indépendant ayant pour mission de relever le défi de la détermination des responsabilités relatives à la réglementation des contenus en ligne[82]. Ce projet de loi servirait aussi d'outil pour exiger des comptes des plateformes en cas de diffusion de contenus préjudiciables sur leurs sites. Toutefois, il a été annoncé que le projet de loi établissant un tel organisme ne serait pas présenté au Parlement avant 2023 ou 2024[83]. En attendant, l'Ofcom, régulateur des médias et télécommunications, s'est vu octroyer plus de pouvoirs pour rendre les sociétés technologiques responsables de la protection des internautes contre les contenus préjudiciables, y compris extrémistes[84]. En outre, en septembre 2019, le Royaume-Uni a annoncé son intention de financer des recherches dans le développement de technologies visant à détecter automatiquement les vidéos modifiées dans le but de contourner les méthodes de détection existantes[85]. L'ambition est de mettre gratuitement cet outil à la disposition de toutes les sociétés technologiques.

78 Gouvernement britannique, ministère de l'Intérieur et ministère du Numérique, de la Culture, des Médias et des Sports. « Online harms – White Paper », avril 2019 : 36. Disponible à l'adresse : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

79 Gouvernement britannique, ministère de l'Intérieur, « The United Kingdom's strategy for countering terrorism », juin 2018 : 35. Disponible à l'adresse : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/716907/140618_CCS207_CCS0218929798-1_CONTEST_3.0_WEB.pdf.

80 Gouvernement britannique, ministère de l'Intérieur, « New technology revealed to help fight terrorist content online », communiqué de presse, 13 février 2018. Disponible à l'adresse : https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online.

81 Commission européenne, 2018(b) : 20.

82 Chambres du Parlement, 2020 ; ministère de l'Intérieur et ministère du Numérique, de la Culture, des Médias et des Sports, 2019.

83 « Online harms bills: warning over 'unacceptable' delay », BBC News, 29 juin 2019. Disponible à l'adresse : https://www.bbc.co.uk/news/technology-53222665.

84 Ibid.

85 Gouvernement britannique, ministère de l'Intérieur, « UK to help develop new tech to stop sharing of terrorist content », communiqué de presse, 24 septembre 2019. Disponible à l'adresse : https://www.gov.uk/government/news/uk-to-help-develop-new-tech-to-stop-sharing-of-terrorist-content.

États-Unis

Aux États-Unis, le Bureau de lutte contre le terrorisme au sein du Département d'État, dirigé par le Coordonnateur de la lutte contre la terrorisme, a la responsabilité de « développer des stratégies et approches coordonnées pour lutter contre le terrorisme à l'étranger et [assurer] la coopération des partenaires internationaux en matière de lutte contre le terrorisme »[86]. Ce bureau travaille aussi avec les sociétés technologiques pour améliorer le partage d'informations[87]. Le Département de la Sécurité intérieure travaille en étroite collaboration avec les alliés américains ainsi qu'avec des organisations comme Tech Against Terrorism pour lutter spécifiquement contre les contenus extrémistes en ligne. Les États-Unis collaborent par ailleurs avec le Forum mondial de lutte contre le terrorisme, un forum multilatéral qui met l'accent sur l'élaboration d'une approche à long terme pour lutter contre la menace que représente le terrorisme.

Selon la stratégie américaine de lutte contre le terrorisme de 2018, la « lutte contre l'influence des terroristes en ligne » est un domaine prioritaire et les États-Unis chercheront à combattre l'utilisation d'Internet par les terroristes pour recruter, lever des fonds et radicaliser des individus tout en travaillant avec leurs partenaires[88]. Les États-Unis se sont fixé trois priorités en matière de lutte contre l'influence exercée par les terroristes en ligne : 1) collaborer avec le secteur privé ; 2) soutenir les efforts de contre-discours déployés par les sociétés technologiques et les organisations de la société civile ; et 3) protéger les droits garantis par le Premier amendement[89].

Pour les décideurs politiques américains, le Premier amendement représente un défi pour la réglementation des contenus extrémistes en ligne. Le contexte américain est quelque peu différent du contexte européen, compte tenu de la portée du Premier amendement. Certains contenus considérés illicites en France ou au Royaume-Uni seront ainsi licites aux États-Unis[90]. Cette distinction entre les contenus « licites mais préjudiciables » ou « licites mais choquants » et les contenus simplement illicites rend la tâche de supprimer les contenus choquants difficile pour les fournisseurs de services Internet. L'article 230 de la loi de 1996 sur la décence dans le domaine des communications permet aux sociétés technologiques de modérer les contenus choquants mais licites[91]. Au fil des décennies, les décisions de la Cour suprême ont introduit de plus grandes nuances à cette juxtaposition de la liberté d'expression et de la facilitation de la violence[92].

86 Gouvernement des États-Unis, Département d'État, « Bureau de la lutte contre le terrorisme ». Disponible à l'adresse : https://www.state.gov/bureaus-offices/under-secretary-for-civilian-security-democracy-and-human-rights/bureau-of-counterterrorism/.

87 Gouvernement des États-Unis, Département d'État, « Country reports on terrorism 2019 », Bureau de la lutte contre le terrorisme, juin 2020. Disponible à l'adresse : https://www.state.gov/wp-content/uploads/2020/06/Country-Reports-on-Terrorism-2019-2.pdf.

88 Gouvernement des États-Unis, Maison Blanche, « National strategy for counterterrorism of the United States of America », octobre 2018 : 22. Disponible à l'adresse : https://www.whitehouse.gov/wp-content/uploads/2018/10/NSCT.pdf.

89 Gouvernement des États-Unis, Département de la Sécurité intérieure, « Strategic framework for countering terrorism and targeted violence », septembre 2019 : 24. Disponible à l'adresse : https://www.dhs.gov/sites/default/files/publications/19_0920_plcy_strategic-framework-countering-terrorism-targeted-violence.pdf.

90 Daphne Keller *et al.*, « Regulating Online Terrorist Content: A Discussion With Stanford CIS Experts About New EU Proposals », SLS Blogs, 25 avril 2019. Disponible à l'adresse : https://law.stanford.edu/2019/04/25/regulating-online-terrorist-content-a-discussion-with-stanford-cis-directors-about-new-eu-proposals/.

91 Ibid.

92 Victoria L. Killion, « Terrorism, Violent Extremism, and the Internet: Free Speech Considerations », Congressional Research Service, 6 mai 2019 : i. Disponible à l'adresse : https://fas.org/sgp/crs/terror/R45713.pdf.

Direction exécutive du Comité contre le terrorisme des Nations Unies

La Direction exécutive du Comité contre le terrorisme des Nations Unies (CTED) a été créée par la Résolution 1535 (2004) du Conseil de sécurité des Nations Unies comme organe spécialisé ayant pour mandat de soutenir le Comité contre le terrorisme du Conseil de sécurité (CTC)[93]. Son but initial était d'évaluer la mise en œuvre par les États membres des Nations Unies des résolutions du Conseil de sécurité en matière de lutte contre le terrorisme et d'appuyer leurs efforts par le dialogue. La CTED travaille en étroite collaboration avec le Conseil de sécurité, les grandes sociétés technologiques via le GIFCT et les organisations de la société civile. Elle a mis en place Tech Against Terrorism, un partenariat public-privé visant à « aider le secteur mondial des technologies à répondre à l'utilisation d'Internet par les terroristes tout en respectant les droits humains »[94].

Plusieurs Résolutions du Conseil des Nations Unies portent sur l'utilisation abusive d'Internet à des fins terroristes. La Résolution 2129 (2013) du Conseil de sécurité observe la relation changeante entre terrorisme et TIC, de même que l'utilisation de technologies comme Internet pour commettre et faciliter des actes terroristes en favorisant l'incitation, le recrutement, la levée de fonds ou la planification d'actes terroristes[95]. Cette résolution renforce également le mandat de la CTED. Les Résolution 2354 (2017), 2395 (2017) et 2396 (2017) prient les États membres de coopérer pour empêcher les organisations terroristes d'exploiter Internet, et de travailler avec le secteur privé et la société civile pour élaborer des mesures effectives de prévention de l'utilisation abusive d'Internet à des fins terroristes[96]. Par ailleurs, en 2017, en collaboration avec la Corée du Sud, le GIFCT et Tech Against Terrorism, la CTED a lancé sa propre plateforme de partage de connaissances en ligne pour promouvoir le partage de bonnes pratiques[97]. Cette plateforme vise à aider les sociétés technologiques plus modestes à contrôler le contenu en ligne et lutter contre l'extrémisme violent[98].

Dans son alerte relative aux tendances de novembre 2018, la CTED a reconnu les défis auxquels étaient confrontés les fournisseurs de services Internet et plateformes plus modestes en matière de régulation des contenus terroristes, défis qui devraient être partiellement relevés grâce à la plateforme de partage des connaissances. Ses publications montrent les divergences entre les mesures prises par les États membres des Nations Unies et celles prises par les sociétés technologiques, ainsi que l'éventai très vaste de pratiques existantes[99]. Celles-ci peuvent aller de l'auto-régulation par les sociétés technologiques, en particulier les plus importantes comme Facebook et Twitter, à l'absence de lois instituant des mesures de type notification et action dans certains pays, comme les Pays-Bas.

93 Naureen Chowdhury Fink, « Meeting the Challenge: A Guide to United Nations Counterterrorism Activities », Institut international de la paix, 2012 : 45. Disponible à l'adresse : https://www.ipinst.org/wp-content/uploads/publications/ebook_guide_to_un_counterterrorism.pdf.
94 « March 2020 update », Tech Against Terrorism, 3 mars 2020. Disponible à l'adresse : https://www.techagainstterrorism.org/2020/04/03/march-2020-update/.
95 Nations Unies, Comité contre le terrorisme du Conseil de sécurité, « Public-private efforts to address terrorist content online: A year of progress – what's next? », 14 septembre 2018. Disponible à l'adresse : https://www.un.org/sc/ctc/news/event/public-private-efforts-address-terrorist-content-online-year-progress-whats-next/.
96 Ibid.
97 Ibid.
98 Conseil de sécurité des Nations Unies, Direction exécutive du Comité contre le terrorisme, 2018 : 2.
99 Voir, par exemple, Conseil de sécurité des Nations Unies, Direction exécutive du Comité contre le terrorisme, 2018.

Conclusion

Dans ce rapport, nous avons présenté le cadre politique instauré par neuf législateurs pour lutter contre les contenus à caractère terroriste en ligne. Les pays reconnaissent que l'utilisation abusive d'Internet par les groupes extrémistes et terroristes est un problème auquel il faut s'attaquer, à l'échelle des pays mais également au sein du secteur privé. Tous les pays que nous avons analysés ont élaboré ou élaborent actuellement des politiques et outils pour faire face à ce problème. Pourtant, si tous les législateurs sauf un semblent reconnaître publiquement la menace que représente l'utilisation d'Internet par les groupes extrémistes et terroristes pour leurs propres desseins, le consensus est toutefois moins net concernant la rigueur des exigences devant être imposées aux sociétés technologiques pour lutter contre cette menace. Nous observons également que les pays partagent des défis, législations et partenaires communs dans la lutte contre les contenus extrémistes en ligne. Il existe toutefois aussi des différences entre ces neuf législateurs, dans les moyens employés pour lutter contre ce type de contenu et la valeur accordée aux droits relatifs à la liberté d'expression. Les organisations et institutions multilatérales, comme l'UE et la CTED, ont parfois été mises en difficulté par le mandat qui leur a été octroyé par différents pays (CTED) ou les désaccords entre États membres concernant les mesures à prendre (UE). Les collaborations avec des initiatives mondiales, comme le GIFCT, et l'utilisation d'outils développés par des organisations, telles que Tech Against Terrorism, semblent être des mesures populaires.

Global Network
on Extremism & Technology

COORDONNÉES

Pour toute question, demande d'information et demande de copies supplémentaires du présent rapport, contacter :

ICSR
King's College London
Strand
Londres WC2R 2LS
Royaume-Uni

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**