The role of fake news in fueling hate speech and extremism online; Promoting adequate measures for tackling the phenomenon

The American technology businessman and former CEO of Google, Eric Schmidt has once said, *"The Internet is the first thing that humanity has built that humanity doesn't understand, the largest experiment in anarchy that we have ever had"*. And indeed, while the importance of the internet in today's society as a driver of communication, gateway to vast amounts of information and enabler of socio-political participation to any type of social group is irrefutable, the drastic advancement in digital and communication technologies have posed numerous problems to State agencies, technology companies and researchers in its catalyzation and diffusion of disinformation, extremist content and hate speech. Those issues have become even more intricate given the development of social media platforms and search engines' algorithms, which curate and proliferate content based on the preference of users, magnifying already existing beliefs and thus causing group polarization.

The current submission will explore those matters further, first by providing a comprehensive definition of the term 'Fake News'. It will examine the role of disinformation campaigns in the fuelling of hate speech and sectarianist sentiments online and then evaluate to what extend that increases the radicalized views of certain individuals and could subsequently lead to extremist violence. It will use as case study the results of EFSAS' research in the region of Indian-administered Jammu & Kashmir where questionnaires were distributed among 139 young people, assessing their capacity of identifying hate speech online and their levels of social media critical thinking. The Paper will conclude with a list of recommendations on possible interventions for tackling the phenomenon.
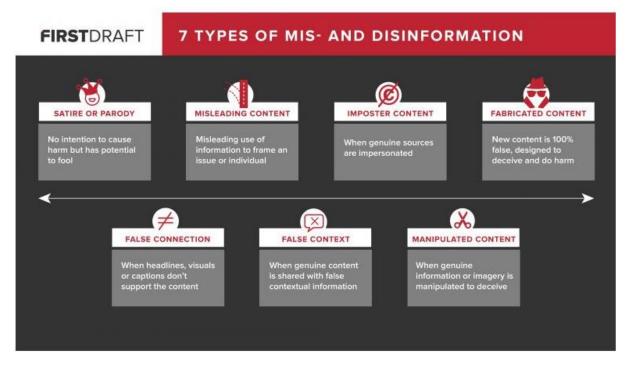
The 'Fake News' Phenomenon

In 2016, 'post-truth' was chosen as Word of the Year by the Oxford Dictionaries, yet as argued by Al-Radhan (2017), the term is symptomatic of an era rather than just a year: *"an era of boundless virtual communication, where politics thrives on a repudiation of facts and commonsense"* (n.p.)*.* As explained by the Oxford Dictionaries, post-truth is an adjective often associated with politics which is defined as *"relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief"* (2016, n.p.). Thus, post-truth politics translates into assertions, which allure to one's emotions and gut feeling, rather than having any basis on empirical evidence and valid information (Al-Radhan, 2017). As further argued by Keyes (2004), a post-truth era creates an ethical twilight zone, where the attached stigma to lying is lost, and lies could be told with impunity and with no consequences for one's reputation. That results in the creation

of rumours, 'fake news' and conspiracy theories, which could go viral in short time and give impetus to false realities and serve propaganda purposes (Al-Radhan, 2017).

In the case of fake news, the term has acquired a dual meaning: on one hand, as fabricated or 'false news', which circulate online, and on the other – as a polemic weapon used to discredit news media channels (Quandt, Frischlich, Boberg and Schatto-Eckrodt, 2019). We would focus first on the former interpretation.

According to Wardle (2017), the definition of fake news needs to be broken down as per the different types of content that are being created and shared, the motivations of those who create this content, and the ways in which this content is being disseminated. In line with that, she distinguishes seven types of fake news, namely: satire or parody, misleading content, imposter content, fabricated content, false connection, false context and manipulated content.

Source: Wardle (2017)

In a similar manner, Nielsen and Graves (2017) describe fake news as a landscape that consists of poor journalism, political propaganda, and misleading forms of advertising and sponsored content. Other authors, such as Lazer et al. (2018, p.1094), define fake news as *"fabricated information that mimics news media content in form but not in organizational process or intent".* Quandt et al. (2019, p.2) further summarise all the proposed definitions in a more systematized method by arguing that first a basic differentiation between (i) the core content of the information (including textual information, imagery, audio elements, etc.); (ii) accompanying meta-information (headlines/titles, author information, tags, and keywords); and (iii) contextual aspects (positioning, references to other articles, framing) needs to take

place. Subsequently, all of these elements could be exposed to different degrees of falsehood, or incongruities from factuality, such as: (a) misleading (but factually correct) information; (b) additions or deletions of information (e.g., "enrichment" of facts by misleading or wrong information, or a change of meaning by omitting or deleting relevant information); and (c) complete fabrications without any factual basis. In addition to that, combinations of those elements could also eventuate (ibid).

The second meaning of the term 'fake news', primarily used by former US President Donald Trump, stands for slandering news coverage that is unsympathetic and critical of one's argumentation or administration (Holan, 2017). Particularly in the case of Trump, the former President used to label media channels as fake news whenever they gave him unfavorable coverage, yet his delegitimization was never followed with any rebuttal consisting of factual evidence or data (ibid). Thus, labeling someone as fake news functions as discrediting one's story, diminishing trust in the media as a whole and obscuring the interpretation of the concept (Quandt et al., 2019). Historically, it has been considered a characteristic of authoritarian regimes to use such "Orwellian" technique in the appropriation of ordinary words and declaring their opposite in a bid to deprive their subjects from independent thinking and convince them in lies (Holan, 2017). A more appropriate term for such strategy nowadays is 'gaslighting', which stands for a psychological form of manipulations where *"a person orchestrates deceptions and inaccurately narrates events to the extent that their victim stops trusting their own judgments and perceptions"* (Jack, 2017, p.9).

Hate Speech

While until 2016 the concept of 'hate speech' existed within its own orbit, that year, the term started often arising simultaneously with 'fake news' (Gollatz, 2018). Although the two terms have distinctively different contexts, as per a conducted study in December 2016, out of 49 articles on hate speech published that month, 37 articles also dealt with fake news (ibid). As further explained by Gollatz (2018), the other factor that connected them was not only the similar incidents around which they appeared but also the same online milieu, particularly social media platforms such as Facebook.

According to the Council of Europe, hate speech refers to *"all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin"* (1997, p.107). In that sense, oftentimes fake news stories include biased and discriminatory content towards members of certain groups of belonging. As argued by Blanco-Herrero and Calderon (2019), the growing cases of hate speech against refugees and migrants is considerably owed to the circulation of fake news related to these groups in the social media space. The rise of nationalist right-wing parties and their derogatory rhetoric of portraying refugees and migrants as a threat have increased the cases of hate speech online, and also led to violence in real life, promoting hate crime (ibid).

Particularly, in the case of refugees and migrants, the intolerance expressed does not relate only to xenophobia and racism, but to a large extent to the fact that the majority of them profess Islam, triggering Islamophobic sentiments. When it comes to mainstream and social media, Islam and Muslims tend to be linked to negative images, oftentimes related to violence and extremism, implying a danger to national security and amplifying the 'us' versus 'them' dichotomy (Aguilera-Carnerero and Azeez, 2016). Occasionally, that further trespasses the boundaries of the mass communications domain, by translating into institutional Islamophobia, where anti-Muslim prejudices have been promulgated within Western societies under the disguise of laws and regulations foisted as being for the benefit of the general public, such as the ban on burqas and mosques in some countries (Aguilera-Carnerero and Azeez, 2016; Esposito, 2019). Far-right political parties such as the British National Party led by Nick Grifin, the Netherland's Party for Freedom of Geert Wilders, Marine Le Pen's National Front and other European nationalist and populist parties have further aided and abetted the passing of restrictive migrant policies and have triggered Islamophobic attitudes amongst the population through inaccurate and biased narratives about Muslims and Islam (Esposito, 2019).

While the role of Islamophobia is well-researched in regards to Islamist radicalization by being at the core of brewing outrage among some Muslims, which in turn allows terrorist groups to hijack those personal feelings of discrimination, marginalization and victimization and convert them into extremist narratives (Abbas, 2012), this Paper will focus on the relatively new and less explored phenomenon of far-right domestic terrorism and radicalization of white men as a response to fake news and conspiracy theories online, especially vis-à-vis anti-Muslim and anti-immigrant discourses.

The Online Disinformation-Terrorism Nexus

As argued by Piazza (2021) there is a very scarce amount of empirical research on the influence of disinformation on actual political violence, and hardly any on its connection to terrorism specifically. Thus, his latest study, which uses a sample of 150 countries for the period of 2000 and 2017, makes two key findings: On one hand, countries which tend to circulate propaganda and disinformation online through the social media channels of their governments, political parties or foreign governments are subject to higher levels of domestic terrorism. On the other, the deliberate dissemination of disinformation online by political actors increases the political polarization of the country.

In order to further illustrate those linkages, an analysis of the existing literature exhibits how while social media platforms were previously considered democracy's allies, have increasingly become its foe, given it is easier to discredit opponents by instead of limiting their speech or criticizing them, to respond with a jumble of misleading and false information, leaving the readers in disarray of what is going on (Beauchamp, 2019). Oftentimes, members of the general public, including researchers and journalists, could lack the resources and tools to fact-check every piece of information and thus verify statements (Deibert, 2019): *"By the time they do, the falsehoods may have already embedded themselves in the collective*

consciousness" (p.32)*.* Even worse, attempts to directly repudiate them could result in their multiplicity by providing them with attention (ibid). Given the downpour of information and cacophony of viewpoints and comments, being confronted with such overflow of online materials, consumers tend to make use of cognitive shortcuts, which navigate them towards opinions that already fit their beliefs (ibid). In addition to that, by being exposed to such myriad of information, users are more likely to start questioning the integrity of all media outlets, which often translates into cynicism and indifference (ibid). As a result, this increases the political apathy and undermines the faith in established democratic institutions, thus strengthening the support and tolerance for far-right, anti-establishment or radical actors, providing oxygen to authoritarian factions (Beauchamp, 2019).

Social media platforms disproportionately assist far-right political parties by helping them bolster social divisions (ibid). They tend to demonize and further marginalize out-group communities such as refugees, immigrants and foreigners (ibid). The major strategy is to portray those individuals as intimidating and dangerous in order for the general population to accumulate fear and hatred against them (ibid). Bilewicz and Soral (2020) explain how exposure to derogatory rhetoric against immigrants and minorities could lead the way to political radicalization and engagement in intergroup violence. They argue that frequent subjection to hate speech results in empathy being replaced with contempt vis-à-vis minority groups, which translates in the erosion of existing anti-discriminatory norms (ibid).

Prominent examples include the 'genocidal' propaganda against Rohingya Muslim minorities in Myanmar, disseminated by not only the general population but also by Army representatives and the spokesman for Burma's de facto leader, Aung Sang Suu Kyi (Washington Post, 2017); the disinformation and fake news campaigns against minorities in South Asia on behalf of some sections of right-wing groups, including false claims of cow slaughtering, child ritual sacrifices in religious places and attacks on Hindus (Vij, 2020); and, Hungary's right-wing PM Viktor Orban's government's conspiracy theories regarding asylum seekers integration in Europe (BBC, 2019).

Thus, as summarized by Piazza (2021), often these political agents use online communities to recruit followers, mold their standpoints and mobilize them to action. Disinformation helps to ferment and reinforce group grievances and opinions, deepening their sense of resentment and rage (ibid).

Particularly in the case of Donald Trump, statements such as *"Allowing the immigration to take place in Europe is a shame . . . I think it changed the fabric of Europe and, unless you act very quickly, it's never going to be what it was and I don't mean that in a positive way"*; *"the U.S. is ill-prepared for this invasion, and will not stand for it. They are causing crime and big problems in Mexico. Go home!"* have been largely discussed in the context of anti-immigration and xenophobic terrorist attacks such as the 2019 El Paso shooting, the 2019 Christchurch mosque shooting in New Zealand and the 2018 Pittsburgh synagogue shooting (Bilewicz and Soral, 2020, p.1).

As explained by Bilewicz and Soral (2020), the perpetrators of these attacks have been previously heavily exposed to anti-immigrant hate speech online and have used such

derogatory language as a justification for their actions. Thus, the role of social media in the dissemination of such fake news and disinformation should not be dismissed. The perpetrators of the Christchurch mosque and El Paso shootings, both prior to the attacks began, sent their manifestos or 'open letters' to several media outlets or social media platforms and shared links to them on 8chan (Wong, 2019). The latter is particularly important for the current article.

8chan [currently 8kun] is an imageboard website composed of user-created message boards, where individuals post anything of interest to them, almost entirely anonymous (ibid). Sometimes deemed a successor or offshoot of the much more popular imageboard 4chan, the 8chan website has been linked to extremist, bigoted, white supremacist, alt-right, neo-Nazi and anti-Semitic content, oftentimes being at the center of inciting hate crimes and mass shootings (ibid).

> *"8chan is almost like a bulletin board where the worst offenders go to share their terrible ideas",* argues Jonathan Greenblatt, the chief executive of the Anti-Defamation League. "*It's become a sounding board where people share ideas, and where these kinds of ideologies are amplified and expanded on, and ultimately, people are radicalized as a result"* (Roose, 2019, n.p.).

Although 8chan was removed from the Google search engine in an unprecedented move after being implicated in containing child pornography (Machkovech, 2015), the website still remains available on the web, especially after rebranding itself to 8kun. Particularly in recent years it became a prominent hub for the establishment and diffusion of the QAnon conspiracy theory.

When it comes to the links between radicalization, domestic terrorism and disinformation, QAnon is at the forefront of examples used by scholars and researchers in the field (Garry, Walther, Mohamed and Mohammed, 2021). QAnon is a collection of miscellaneous conspiracy theories, with the central one arguing that a cabal of political elites and prominent public figures are part of a Satan-worshipping pedophile ring, and Donald Trump is the only person who could defeat them, often portrayed as the nation's savior (ibid). The name originates from 'Q', as in "Q Clearance", which is a top-secret category of federal security clearances in the US, and 'Anon' as in "anonymous", arguing that this is an individual who drops to his followers clues to what is going to happen next, based on his access to highly confidential information at the government (ibid). While, QAnon incorporates various conspiracy narratives, its followers have managed to deduce concrete goals, translatable into actions, namely:

- *"A massive information dissemination program meant to:*

  o *Expose massive global corruption and conspiracy to the people.*

  o *Cause the people to research further to aid further in their "great awakening."*

  o *Root out corruption, fraud, and human rights violations worldwide.*

    o   *Return the Republic of the United States to the Constitutional rule of law and also return "the People" worldwide to their own rule."* (ibid, p. 160).

While conspiracy thinking and violent extremist ideologies share different categories, they could nonetheless intersect (ibid). Such overlap increases security concerns and establishes a dangerous mechanism when the conspiracy asserts that:

"(1) one group is superior to another (superiority versus inferiority);

(2) one group is under attack by the other (imminent threat); and

(3) the threat is apocalyptic in nature (existential threat)" (RAN, 2020, p.3).

Thus, in the case of QAnon, all of the abovementioned factors are present (Garry et al., 2021). Further research shows how when those features are combined with characteristics such as low self-control, law-relevant morality and self-efficacy, this could directly lead to violent extremist action (RAN, 2020). That was particularly the case of Edgar Welch, who in 2016 stormed a Washington-based pizza restaurant with an AR-15 rifle and started indiscriminately shooting, believing that the venue was the stage of a Hillary Clinton-run child sex network (LaFrance, 2020). While this particular case was known as the Pizzagate conspiracy theory, it largely gave the impetus of QAnon (ibid). The storming of the US Capitol earlier this year was also initiated by QAnon supporters (Argentino, 2021). The latter gave the impetus of currently ongoing Senate hearings, meant to recognize the security failures which led to those riots, including the responsibility of social media platforms (Wakefield, 2021).
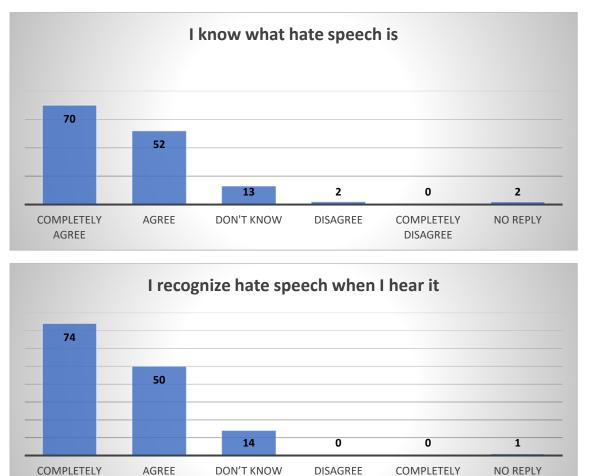
Johnson (2018) calls this process of self-radicalization of white men through fake news the result of masculinist paranoia that is built into the social processes of human and nonhuman communication and acts as a defense to a perceived threat, and is what gives oxygen to conspiracy theories to further proliferate, creating a vicious cycle. Thus, this paper will provide practical insights and recommendations on how to address and counter the diffusion of fake news, disinformation and conspiracy theories online. Before moving to that section, however, as a case study this Paper will present the results of a study amongst 139 youngsters in Indian-administered Jammu & Kashmir, which examined their levels of understanding and familiarity of key terminology regarding hate speech and violent extremism, and their levels of critical thinking towards social media platforms and information online, in order to develop better-informed strategies for debunking the abovementioned phenomena. While acknowledging the unique context of the study, which is not necessarily replicable in every social milieu, the quality of being the only one of its kind as per the current period was deemed highly significant for its dissemination.

Case Study

We received a total of 139 answers. All but one questions addressing indicators related to youths' capacity to identify hate speech and violent extremism are cross-sectional, i.e., they measure a point of departure rather than the impact of the project.

Multiple-choice questions demonstrated that participants are relatively confident in their ability to detect hate speech, with 122 responding affirmatively to the question *I know what hate speech is,* and 124 either completely agreed (74) or agreed (50) that they be able to *recognize hate speech when [they] hear it.*

I know what hate speech is

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
70	52	13	2	0	2

I recognize hate speech when I hear it

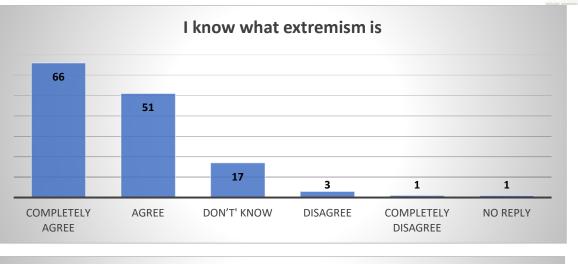| COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
74	50	14	0	0	1

When asked to describe the difference between hate speech and a legitimate arguments, however, respondents were less confident: 94 either gave no or invalid answers or stated that they did not know. 41 participants explained the difference between hate speech and legitimate argumentation in terms of the way opinions are brought forth, with many labelling hate speech as emotionally driven and legitimate arguments as driven by facts. Another 5 respondents drew the line along normative rationales arguing, for example, that 'hate speech is bad and a legitimate argument is healthy', that there is 'no moral in hate speech' or that 'hate speech is for personal benefits and has no community benefit'.
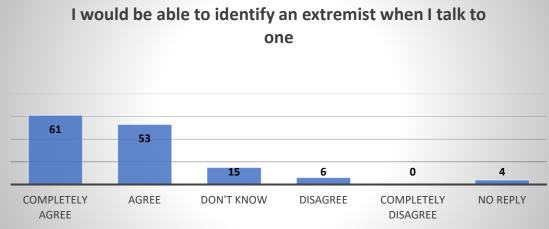
Participants also appeared confident with regard to their perception of their own understanding of the concept of 'extremism'. When asked whether they *know what extremism is,* 117 responded positively, with 66 agreeing completely and 51 agreeing and similarly, 114 stated that they *would be able to identify an extremist when [they] talk to one*.

I know what extremism is

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
66	51	17	3	1	1

I would be able to identify an extremist when I talk to one

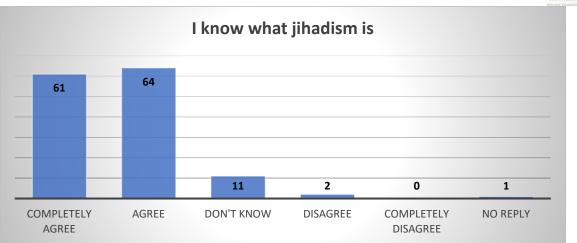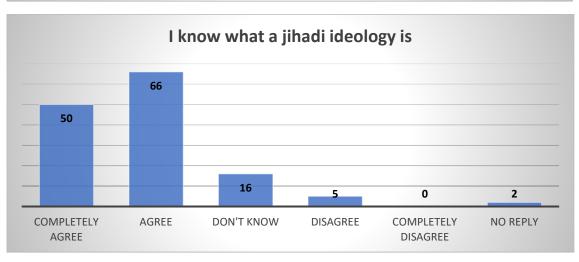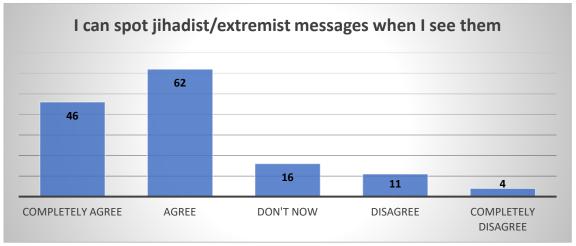| COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
61	53	15	6	0	4

However, in line with a pattern observable throughout the survey assessment, open-ended questions did not necessarily corroborate this confidence: when asked to *specify what, in [their] opinion, distinguishes an extremist from someone who does not have extreme views,* a full 114 failed to provide valid or relevant responses. Of those who did respond, 16 quoted C. Mudde (2000:12) on the difference between radicalism and extremism (hence failing to provide a concrete answer to the question posed), while others gave vague responses such as 'radical person', 'his vulgar ideas' or 'extremists are submissive while others are not'. These results indicate that while the target audience appears to have an intuitive understanding of hate speech and extremism, upcoming efforts focused on the target group will have to provide more concise clarifications of key concepts.

Participants were similarly confident with regard to the concept of 'jihadism', with 125 stating to agree completely (61) or to agree (64) with the question *I know what jihadism is* and 116 responding affirmatively to the question *I know what a jihadi ideology is.* Further, when asked whether they believed to be able to *spot jihadist/extremist messages when [they] see one,* 46 participants responded with 'completely agree' and 62 with 'agree' – making a total of 108 affirmative responses – while 16 disagreed or completely disagreed.

I know what jihadism is

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
61	64	11	2	0	1

I know what a jihadi ideology is

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
50	66	16	5	0	2

I can spot jihadist/extremist messages when I see them

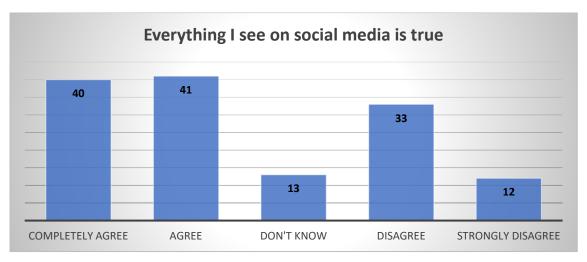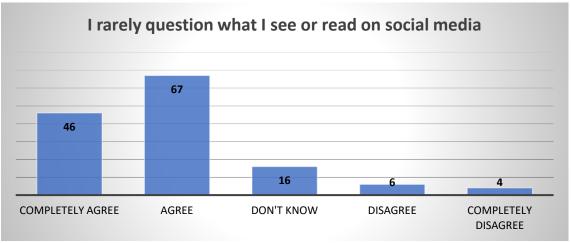| COMPLETELY AGREE	AGREE	DON'T NOW	DISAGREE	COMPLETELY DISAGREE
46	62	16	11	4

Measuring participants' critical thinking skills regarding social media contents, we asked a number of longitudinal-type questions, the responses to which suggest a rather low level of social media criticality within the target group.
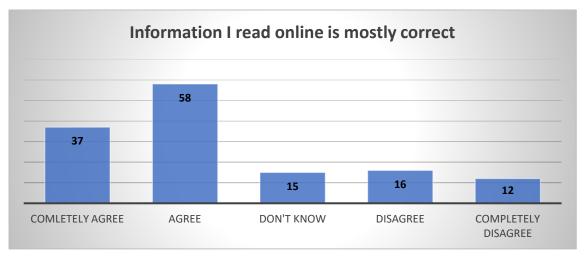
Upon being asked to state whether *everything [they] see on social media is true,* 81 participants responded affirmatively, while 33 disagreed and 12 disagreed completely, suggesting that more than half of the respondents do not question online contents. This lack of criticality is confirmed in the subsequent statement – *I rarely question what I see or read*

on social media – which is responded to affirmatively by 113 participants, while a mere 10 disagreed or completely disagreed. In a similar vein, 95 participants appeared to believe that *Information [they] read online is mostly correct,* while only 28 disagreed or completely disagreed. It is unclear whether this reflects participants' actual mindset, or whether it is the result of so-called acquiescence bias, i.e., the tendency of survey respondents to agree with research statements – especially since many participants have tended to tick the same boxes in a vertical line, suggesting either a lack of time or interest in responding to the survey.

Everything I see on social media is true

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	STRONGLY DISAGREE
40	41	13	33	12

I rarely question what I see or read on social media

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE
46	67	16	6	4

Information I read online is mostly correct

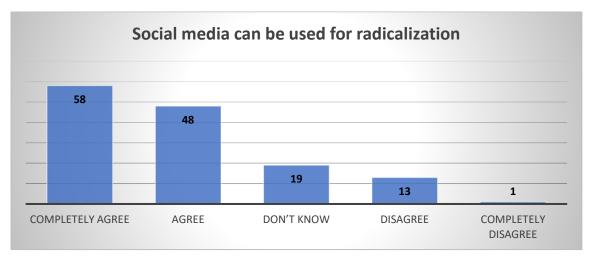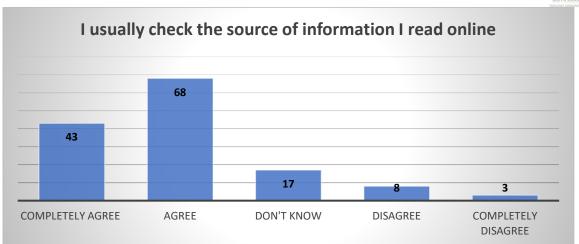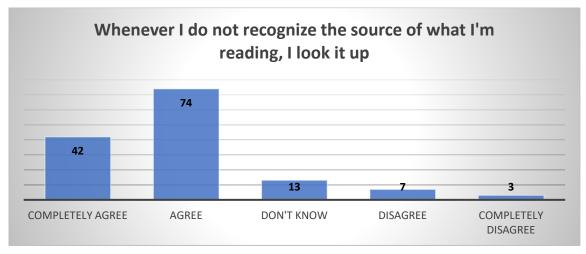| COMLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE
37	58	15	16	12

With regard to the link between social media and extremism, we asked participants to state whether they believed that *social media can be used for radicalization.* 106 stated that they did believe so (58 completely agreed, while 48 agreed), while 13 disagreed and one completely disagreed. When asked to state their opinion on the question *what are the potential dangers of social media?*, however, very few respondents explicitly linked between terrorism to social media: only 4 participants included extremism or terrorism in their response. Fake news, with a total of 37 references, made up a more sizeable portion, as did cyberbullying, invasion of privacy and identity theft (14 times) and riots (probably referring to popular mobilization; 21 times).

Social media can be used for radicalization

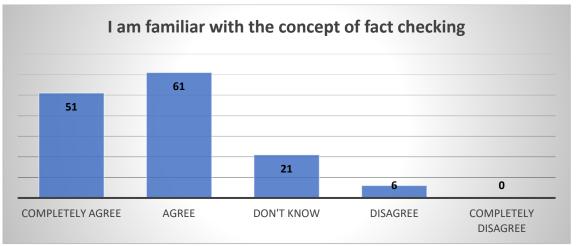|				
58	48	19	13	1
COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE

The ability of participants to verify online sources and contents was also assessed. To this end, we asked whether participant *usually check the source of information (publishing author or institution) [they] read online*, to which 111 participants completely agreed or agreed, while 8 disagreed and 3 completely disagreed. Furthermore, 116 completely agreed (42) or agreed (74) that *whenever [they] do not recognize the source of what [they're] reading, [they] look it up* and 112 stated that they were *familiar with the concept of fact checking.* In the open-ended question, formulated as *what techniques would you use to make sure information you receive online are correct?* a total of 41 participants responded that they would cross-check information read online through, for example, books (15) or other websites (2). Another 24 suggested to verify online statements interactively, e.g., through group discussion, while 54 failed to provide a relevant answer.

I usually check the source of information I read online

Bar chart values:
- COMPLETELY AGREE: 43
- AGREE: 68
- DON'T KNOW: 17
- DISAGREE: 8
- COMPLETELY DISAGREE: 3

Whenever I do not recognize the source of what I'm reading, I look it up

Bar chart values:
- COMPLETELY AGREE: 42
- AGREE: 74
- DON'T KNOW: 13
- DISAGREE: 7
- COMPLETELY DISAGREE: 3

I am familiar with the concept of fact checking

Bar chart values:
- COMPLETELY AGREE: 51
- AGREE: 61
- DON'T KNOW: 21
- DISAGREE: 6
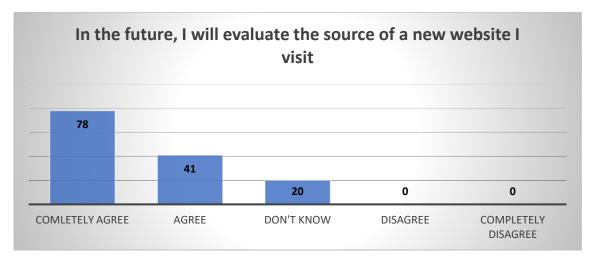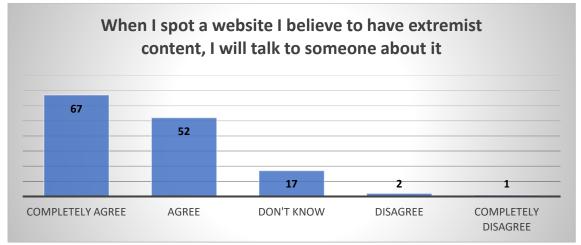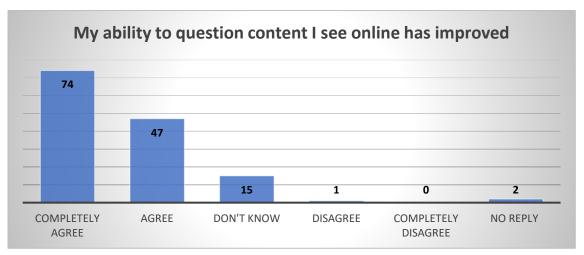- COMPLETELY DISAGREE: 0

To assess the immediate impact the workshops conducted by the author had on participants, it was asked whether *in the future, [they] will evaluate the source of a new website [they] visit*, to which 119 responded affirmatively while none disagreed or completely disagreed. In the same vein, 119 responded that they completely agreed (67) or agreed (52) that *when [they] spot a website [they] believe to have extremist content, [they] will talk to someone about it*, with only 3 disagreeing or completely disagreeing. More generally, 111 participants confirmed that their *ability to question content [they] see online has improved* (with only one
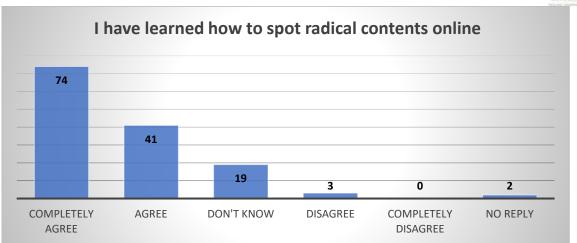
disagreeing), and 115 stated that they *have learned how to spot radical contents online* (with 3 disagreeing).

In the future, I will evaluate the source of a new website I visit

COMLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE
78	41	20	0	0

When I spot a website I believe to have extremist content, I will talk to someone about it

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE
67	52	17	2	1

My ability to question content I see online has improved

COMPLETELY AGREE	AGREE	DON'T KNOW	DISAGREE	COMPLETELY DISAGREE	NO REPLY
74	47	15	1	0	2

I have learned how to spot radical contents online

Recommendations

From the abovementioned section it becomes visible how while young people might intuitively have an understanding about what hate speech and extremist messages online look like, when it comes to exact terminology and more specific nuances, the issue is more complex. Even more alarmingly, our test group argued that they tend to believe information online unquestioningly and rarely fact-check its content. Thus, such disquieting statistics call for the adoption of comprehensive counter strategies.

According to a report issued by the Institute for Strategic Dialogue (2016, p.8), approaches to countering hate speech and extremist content originating from disinformation online could be divided into three major categories:

"(1) Efforts to restrict availability of extremist and disinformation content and access to hate speech on the internet by reporting, filtering and removing content, and taking appropriate measures or invoking legal regulations;

(2) Efforts to compete with such content by providing a broader spectrum of perspectives through counter or alternative narratives, or more recently by fact checking; and

(3) Efforts to bolster the resilience of internet users through digital and civic education (typically of young people) and broad public awareness campaigns".

Internet platforms have become the central conduits for the dissemination of fake news, since it is financially inexpensive to create websites that promote disinformation content, and rather simple to create tools such as ads for sharing them on different social media channels (Lazer et al., 2018). While strategies such as taking down websites, filtering materials and other forms of blocking online content are well known measures on behalf of concerned parties and are certainly necessary remedies in the current farrago, given the numerous hurdles related to that including the political repercussions and potential loss of legitimacy on democratic governments, which could be seen as 'censoring' parts of the internet, acting impartially or curbing 'freedoms of speech and expression', we try to offer few more sustainable approaches.

For instance, social media channels could include indicators of source quality into its shared content, working in collaboration with independent researchers and scholars on such categorizations (Lazer et al., 2018).That is particularly the case with Twitter, which started putting warning messages to tweets that might contain misleading information (BBC, 2020).The social media platform further launched its initiative Birdwatch, which encourages individuals to provide informative context to tweets they believe are fallacious (Coleman, 2021). In a similar vein, WhatsApp imposed limits on the forwarding of messages in a bid of slowing down the dissemination of fake news, thus under the new regulations, a viral message – one which has been forwarded more than five times – will be allowed to be sent on to only one single chat at a time (Hern, 2020). Due to the encryption of the application, the company cannot oversee the contents of the messages, yet this way it hopes to introduce restrictions on mass mailing (ibid). In addition to that, platforms should try suppressing the automated diffusion of fake news trough bots, which as per Facebook account to up to 60 million of their users (ibid).

Considering the omnipresence of the Internet, people will inevitably stumble across questionable or upsetting materials, however it is the attitude to those materials which will define their actions – whether they will engage with it or report it. Thus, the second point of engaging internet consumers with alternative perspectives is imperative to tackling disinformation. Alternative narratives present alternative messages, instead of engaging with the same extremist content, focusing primarily on positive values, such as communal harmony, diversity, tolerance, social inclusion, freedom and democracy.

In order to select which of the three approaches is most suitable for the given situation, van Ginkel (2015) argues that one must answer the question, *"Who is in control of the narrative?"* (p.1). In order to further clarify that, there are certain recurring elements that need to be taken into account: *target group, message, messenger* and *channel used for communicating the message*. It is imperative for the purposes of having a successful campaign to consider all four factors and coordinate them accordingly.

As communicated by the Radicalization Awareness Network's Communication and Narratives working group (RAN C&N), which has developed the GAMMMA+ model for implementing efficient alternative and counter narrative campaigns, understanding profoundly the needs, priorities, motivations, beliefs and influences of the target audience, how and where they communicate them to each other, and why they would be inclined to responding to one's campaign is the basis of any intervention (RAN, 2019a). Therefore, knowing the group's demographic information, their online and offline social networking patterns, their educational and professional occupations, alongside with their general interests is vitally important for engaging effectively with them (van Ginkel, 2015).

Second, the message promoted during the campaign must be highly relevant to the target group's needs and carry a positive social currency (RAN, 2019a). Campaigns that count on constant stream of content, which fosters interaction with the target audience, and in that sense have to rely on quantity and authenticity rather than simply technical quality, appear more successful in generating impact (ibid). In addition to that, the language selected for

undertaking the campaign must resonate with the lingo of the target audience (van Ginkel, 2015).

Even if the message of the campaign is designed to perfection, the lack of a credible messenger to disseminate it, all efforts of addressing and influencing positively the target audience will remain futile (ibid). The credibility of the messenger correlates with the extent to which he or she is being seen as a trustworthy individual by the target group (ibid). Here, van Ginkel (2015) outlines the main actors involved in the delivery of such messages:

- *Government actors* are best positioned in the formal communication of messages which fall in line with the government strategic communication campaigns, which predominantly focus on narratives of rule-of-law-based societies, which respect and protect human rights, diversity and pluralism. They further inform their audiences of their foreign policy decisions and measures adopted to mitigate any risks in terms of security. Government representatives could explain the rationales behind such operations and mitigate any misconceptions of governmental activities. The government could further play a small role in providing alternative narratives, yet that is rather limited, as it is better left to other actors.

- *Semi-public actors or front-line practitioners* are actors, including youth workers, social workers and medical practitioners who often enjoy one-to-one communication with individuals from the target audience and are thus best situated in providing alternative narratives, alongside with counter-narratives. Their efficiency highly relies on their ability to build trust with the person in question.

- *Religious leaders and religious associations* are best placed in addressing extremist misinterpretation of Islam, thus directly tackling jihadist narratives and providing the individual with a correct reading of religion, thus offering both counter and alternative narratives.

- *Associations representing minority groups or migrants* are in similar position to communicating alternative and counter narratives as religious leaders.

- *Role models and youth leaders* are often public figures which are well respected in society and looked up by young people. They could play a positive role in setting up an example and inspiring others to follow it, thus desisting from the path of radicalization. They are most effective in communicating alternative narratives.

- *Former terrorists* are also very important voices in the delivery of credible messages. Based on their own experience they could provide both counter and alternative narratives, by exposing the misleading and treacherous line of terrorist groups and fostering others to question their real intentions.

- *Victims of terrorism* also carry significant role in providing counter-narratives by displaying the inhumane and barbaric nature of terrorist groups.

- *Educators* are essential in recognizing early signs of indoctrination and radicalization and thus addressing the underlying causes through special educational programs or open discussions and providing alternative narratives.

- *Family members and direct neighbours and friends* are also uniquely situated in recognizing early signs of radicalization and thus having an open dialogue with the individual in countering the narrative and offering alternative solutions.

The last step is choosing the correct medium of communication (ibid). In the same manner that extremists are using multiple platforms to communicate their message, successful public campaigns need to determine and resort to the right mediums for the distribution of their message. It is essential to consider using the right medium for the right purpose (RAN, 2019a). For example, the target audience could use Facebook as a source of news, but be more interactive and responsive on Instagram. Hence, if the aim of the campaign is to create short informative videos, then it would be best to disseminate it on Facebook. But if the aim of the campaign is to collect qualitative data, then using Instagram stories for polls, questions etc. could be the better option.

RAN Policy & Practice Workshop on Narratives and Strategies of Far-Right and Islamist Extremists published in 2019 presents the example of several well-known frames which far-right groups exploit in order to attract followers and offers guidelines in dealing with those narratives through counter and alternative communication strategies (RAN, 2019b).

- *"national identities are under threat"* – This is an often used narrative, which tries to imply that the white race is under threat from Muslims and migrants. Thus, terms such as *"white genocide", "demographic jihad"* and *"Islamisation"* are utilized to justify the actions of far-right outfits. Therefore, one's identity and physical appearance become major tenet for engagement in violence. As such, any strategies aiming to provide an alternative narrative should emphasise on the existence of multiple identities, breaking away from the binary discourse employed by such extremist groups.

In addition to that, as argued by RAN (2019b), given the tricky interplay between far-right and Islamist narratives, recognizing how both extremist groups feed off each other's discourse, such strategies should avoid stigmatization of one particular group and thus try bringing them together through common interests.

When it comes to the last section, media literacy classes become an essential weapon. The term stands for the ability to identify different types of media and assess them critically and analytically vis-à-vis the message they are aiming to convey and their authenticity. As argued by Hobbs (2010), digital and media literacy encompasses *"the full range of cognitive, emotional, and social competencies that include the use of text, tools and technologies; the skills of critical thinking and analysis; the practice of messaging composition and creativity; the ability to engage in reflection and ethical thinking; as well as active participation through teamwork and collaboration"* (p.17).

Some of the main strategies include:

- Recognizing mis- and disinformation (content analysis) based on non-formal logic, discerning media bias.

- Exploring different types of fallacious argumentation and learning how to detect them.

- Recognizing emotional appeals online.

- Addressing social representations in mainstream and social media.

- Using Multiple Sources. Source check, authorship, fact-checking.

- Gauging Tone and Language, source, intent and purpose, beliefs, values (social representations). Audience analysis, ex. left-right spectrum.

- Questioning Numbers and Figures, determining importance, synthesizing information.

- Understanding Images - visual content analysis, semiotic analysis.

- Developing Multimedia Skills. Creating the Media Ourselves, creativity, filming, editing, and writing. Web 2.0.

While very few studies actually evaluate the effectiveness of such interventions in practice, Guess et al.'s (2020) data from preregistered survey experiments conducted around elections in the United States and India indicated how exposure to this exercise decreased the perceived accuracy of both mainstream and false news headlines, with effects on the latter being significantly larger. As a consequence, their study improved the differentiation between mainstream and fake news headlines by 26.5% amongst a nationally representative sample in the US and by 17.5% amongst the highly educated population in India.

Thus, more attention should be paid to critical online content analysis in order to strengthen young people's safeguarding mechanisms vis-à-vis fake news messages. The intricate media and information landscape are in need of critical minds in the public in order to continue to serve its purpose properly.

The current Paper aimed to concisely portray the phenomenon of fake news, its role in fueling hate speech and extremist messages online, and thus its prospects for leading individuals on the path of radicalization. Particularly with the rapid development of technologies such as artificial intelligence and the genesis of 'deep fakes', the challenges of combatting disinformation have reached a new high. Therefore, this article sought to offer the needed context and policy recommendations for tackling the issue.

References

Abbas, T. (2012). The symbiotic relationship between Islamophobia and radicalisation. *Critical Studies on Terrorism*, *5*(3), pp.345-358.

Aguilera-Carnerero, C. and Azeez, A.H. (2016). 'Islamonausea, not Islamophobia': The many faces of cyber hate speech. *Journal of Arab & Muslim media research*, *9*(1), pp.21-40.

Al-Rodhan, N. (2017). Post-truth politics, the fifth estate and the securitization of fake news. *Global Policy*.

Argentino, M. (2021). QAnon and the storm of the U.S. Capitol: The offline effect of online conspiracy theories. *The Conversation.* Available at: https://theconversation.com/qanon-and-the-storm-of-the-u-s-capitol-the-offline-effect-of-online-conspiracy-theories-152815

Baldauf, J., Ebner, J. and Guhl, J. (2019). Hate Speech and radicalisation Online: The OCCI research report. *Institute for Strategic Dialogue*.

BBC (2019). EU blasts Hungary 'fake news' on migrants. *BBC.* Available at: https://www.bbc.com/news/world-europe-47294183

BBC (2020). Coronavirus: Twitter will label Covid-19 fake news. *BBC.* Available at: https://www.bbc.com/news/technology-52632909

Beauchamp, Z. (2019). Social Media is Rotting Democracy from Within: How Social Platforms Enable Far-Right Politicians' Campaigns to Undermine Democracy. *Vox*. Available at: https://www.vox.com/policy-and-politics/2019/1/22/18177076/social-media-facebook-far-right-authoritarian-populism

Bilewicz, M. and Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*, pp.3-33.

Blanco-Herrero, D. and Calderón, C.A. (2019). Spread and reception of fake news promoting hate speech against migrants and refugees in social media. In *Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality-TEEM*.

Coleman, K. (2021). Introducing Birdwatch, a community-based approach to misinformation. *Twitter.* Available at: https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation.html

Deibert, R.J. (2019). The road to digital unfreedom: Three painful truths about social media. *Journal of Democracy*, *30*(1), pp.25-39.

Esposito, J.L. (2019). Islamophobia and radicalization: roots, impact and implications. In *Islamophobia and Radicalization* (pp. 15-33). Palgrave Macmillan, Cham.

Garry, A., Walther, S., Rukaya, R. and Mohammed, A. (2021). QAnon Conspiracy Theory: Examining its Evolution and Mechanisms of Radicalization. *Journal for Deradicalization*, (26), pp.152-216.

Gollatz, K. and Jenner, L. (2018). Hate speech and fake news - how two concepts became intertwined and politicized. *Zenodo*. Available at: https://zenodo.org/record/1197174#.YH1yi-77TIU

Gowen, A. and Bearak, M. (2017). Fake news on Facebook fans the flames of hate against the Rohingya in Burma. *The Washington Post*. Available at: https://www.washingtonpost.com/world/asia_pacific/fake-news-on-facebook-fans-the-flames-of-hate-against-the-rohingya-in-burma/2017/12/07/2c1fe830-ca1f-11e7-b506-8a10ed11ecf5_story.html

Guess, A.M., Lerner, M., Lyons, B., Montgomery, J.M., Nyhan, B., Reifler, J. and Sircar, N., 2020. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), pp.15536-15545.

Hern, A. (2020). WhatsApp to impose new limit on forwarding to fight fake news. *The Guardian*. Available at: https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news

Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action. A White Paper on the Digital and Media Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a Democracy*. Aspen Institute. 1 Dupont Circle NW Suite 700, Washington, DC 20036.

Holan, A. (2017). The media's definition of fake news vs. Donald Trump's. *First Amend. L. Rev.*, *16*.

Jack, C. (2017). Lexicon of lies: Terms for problematic information. *Data & Society*, *3*, p.22.

Johnson, J. (2018). The self-radicalization of white men: "Fake news" and the affective networking of paranoia. *Communication Culture & Critique*, *11*(1), pp.100-115.

Keyes, R. (2004). *The post-truth era: Dishonesty and deception in contemporary life*. Macmillan.

LaFrance, A. (2020). The Prophecies of Q: American conspiracy theories are entering a dangerous new phase. *The Atlantic*. Available at: https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/

Lazer, D., Baum, M., Benkler, Y., Berinsky, A., Greenhill, K., Menczer, F., Metzger, M., Nyhan, B., Pennycook, G., Rothschild, D. and Schudson, M. (2018). The Science of Fake News: Addressing Fake News Requires a Multidisciplinary Effort. *Science*, *359*(8).

Machkovech, S. (2015). 8chan-hosted content disappears from Google searches [Updated]. *Ars Technica*. Available at: https://arstechnica.com/tech-policy/2015/08/8chan-hosted-content-disappears-from-google-searches/

Nielsen, R. K., & Graves, L. (2017). *"News you don't believe": Audience perspectives on fake news*. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/news-you-dont-believe-audience-perspectives-fake-news

Oxford Dictionaries (2016). *Oxford Word of the Year 2016.* Available at: https://languages.oup.com/word-of-the-year/2016/

Piazza, J.A. (2021). Fake news: the effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, pp.1-23.

Quandt, T., Frischlich, L., Boberg, S. and Schatto-Eckrodt, T., 2019. Fake news. *The international encyclopedia of Journalism Studies*, pp.1-6.

Radicalisation Awareness Network. 2019a. *Effective Narratives: Updating the GAMMMA+ model. RAN.*

Radicalisation Awareness Network. 2019b. *RAN Policy & Practice Workshop – Narratives and Strategies of Far-Right and Islamist Extremists. RAN.*

Radicalisation Awareness Network (2021). Conclusion paper: The impact of conspiracy narratives on Violent RWE and LWE Narratives. *RAN*. Available at: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/about-ran/ran-c-and-n/docs/ran_c-n_concl_pap_impact_consp_narr_on_vrwe_vlwe_24-25_112021_en.pdf.

RECOMMENDATION, N.R. (1997). of the Committee of Ministers to Member States on" Hate Speech. *Adopted on*, *30*.

Roose, K. (2019). 'Shut the Site Down,'Says the Creator of 8chan, a Megaphone for Gunmen. *The New York Times*. Available at: https://www.nytimes.com/2019/08/04/technology/8chan-shooting-manifesto.html

Van Ginkel, B. T. 2015. *Responding to Cyber Jihad: Towards an Effective Counter Narrative.* International Centre for Counter-Terrorism Research Paper.

Vij, S. (2020). India's anti-Muslim fake news factories are following the anti-Semitic playbook. *The Print.* Available at: https://theprint.in/opinion/india-anti-muslim-fake-news-factories-anti-semitic-playbook/430332/

Wakefield, J. (2021). Google, Facebook Twitter grilled in US on fake news. *BBC*. Available at: https://www.bbc.com/news/technology-56523378

Wardle, C. (2017). Fake news: It's complicated. *First Draft*. Available at: https://firstdraftnews.com/fake-news-complicated/

Wong, J.C. (2019). 8chan: the far-right website linked to the rise in hate crimes. *The Guardian*. Available at: https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website