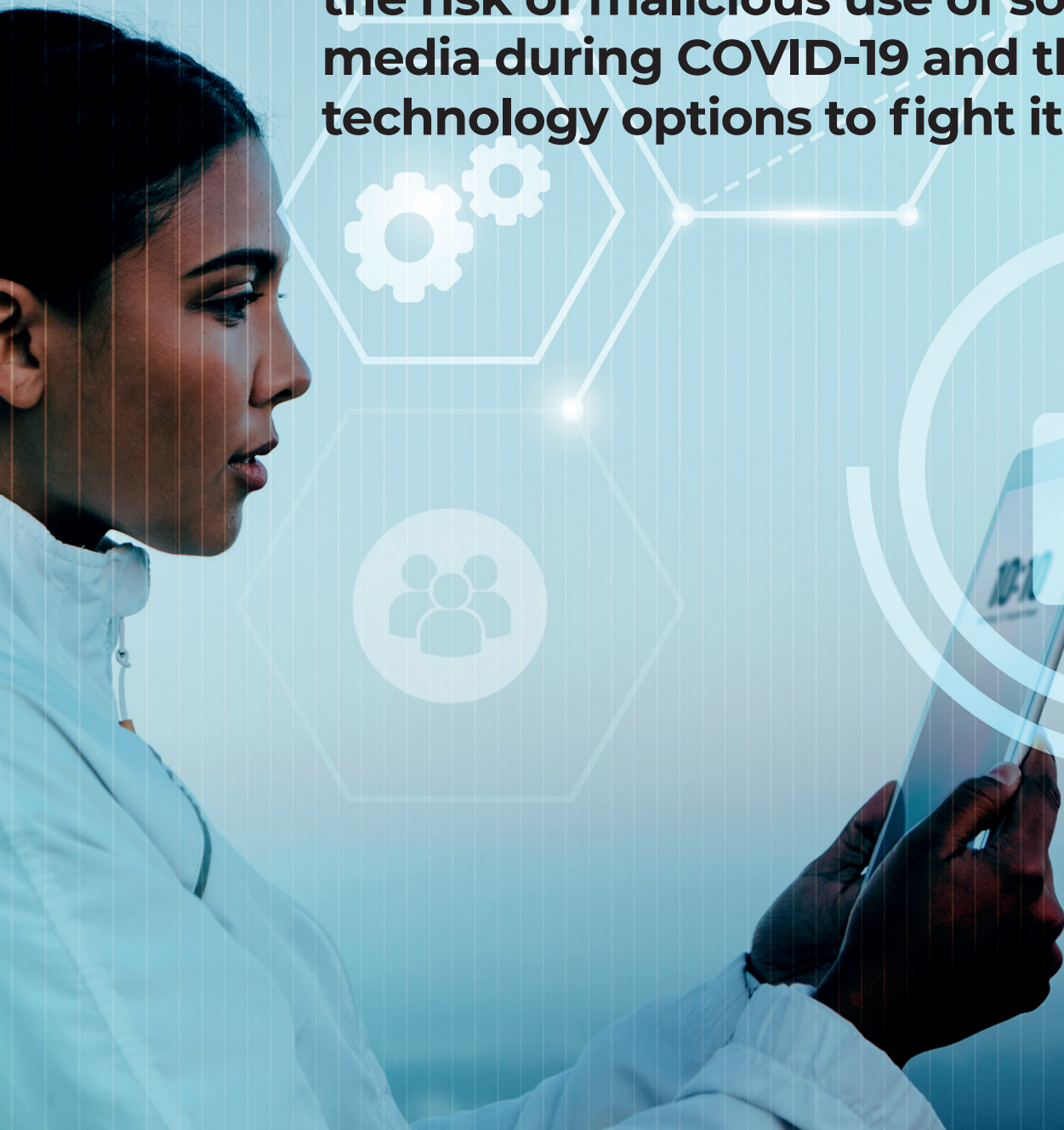UNICRI

United Nations
Interregional Crime and Justice
Research Institute

STOP THE VIRUS OF DISINFORMATION

the risk of malicious use of social media during COVID-19 and the technology options to fight it

STOP THE VIRUS OF DISINFORMATION

the risk of malicious use of social media during COVID-19 and the technology options to fight it

DISCLAIMER

The opinions, findings, and conclusions and recommendations expressed herein do not necessarily reflect the views of the United Nations Interregional Crime and Justice Research Institute (UNICRI) or any other the national, regional or international entity involved. The responsibility for opinions expressed in signed articles, websites, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by UNICRI of the opinions expressed in them. The designation employed and presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area of its authorities, or concerning the delimitation of its frontiers or boundaries. Contents of this publication may be quoted or reproduced, provided that the source of information is acknowledged.

COPYRIGHT

Foreword

This report describes how terrorist, violent extremist and organized criminal groups are trying to take advantage of the Coronavirus disease (COVID-19) pandemic to expand their activities and jeopardize the efficacy and credibility of response measures by governments.

Misinformation and disinformation on social media are not new problems, but the COVID-19 crisis has amplified them and created new opportunities for violent non-state actors. In recent months we have seen numerous cases of malicious use of social media to undermine trust in governments and, at the same time, to reinforce extremist narratives and recruitment strategies. Terrorist, violent extremist and organized criminal groups have successfully exploited vulnerabilities in the social media ecosystem to manipulate people and disseminate conspiracy theories about the origin of COVID-19, its mode of transmission and possible cures.

It is also alarming that some terrorist and violent extremist groups have attempted to misuse social media to incite potential terrorists to intentionally spread COVID-19 and to use it as an improvised form of a biological weapon. Moreover, some criminal organizations have taken advantage of the COVID-19 pandemic to portray a positive self-image to reinforce their presence and control in the territory and to further expand their illegal activities.

As a proud member of the United Nations Global Counter-Terrorism Coordination Compact, UNICRI prioritizes the prevention and countering of terrorism, radicalization and violent extremism in all its forms. We apply our expertise as the leading United Nations research and training institute in the field of criminal justice and crime prevention to constantly explore innovative ideas and tools to assist Member States and the international community at large to fight the scourge of terrorism. In this connection, this report analyses how technology can provide valid instruments to combat online disinformation and misinformation, highlighting both the advantages and the limitations.

It must be emphasized, however, that technology countermeasures alone cannot solve the problem of malicious use and abuse of social media. Technology can support but not replace human ability and skill to evaluate the veracity of online information. Moreover, the effective use of technology to detect and debunk disinformation should empower people using social media so that they can make their own informed decisions about what is verified and what is not. This, then, can empower people to build a closer relationship with truth and justice.

We hope this report enhances knowledge and understanding of this complex problem and encourages the identification of new approaches to prevent and combat the malicious use of social media by violent non-state actors.

Antonia Marie De Meo
UNICRI Director

Acknowledgements

Table of contents

BIG DATA IS
WATCHING YO

Introduction

The Coronavirus disease (COVID-19) pandemic and the resulting global crisis are posing unprecedented challenges to national and international communities due to the rapid and global impact on public health, social well-being, the economy and critical infrastructures of several countries.[1] After the World Health Organization (WHO) declared COVID-19 a pandemic on 11 March 2020, governments around the world adopted protective measures on their societies and economies that included quarantine, isolation and social distancing to limit the spread of the virus. These restrictions contributed to an immediate decline in global trade, an international economic downturn and social tensions.

In this environment, in which home isolation, health concerns about the virus and its socio-political and economic implications produced an explosive mixture of anxiety and fear, social media offered a window of opportunity for collective discussion and response to the coronavirus outbreak. Surveys of social media users in different countries show an increase in the use of popular social platforms such as

---

1   According to WHO, there are two official names: coronavirus disease (COVID-19) for the disease and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) for the virus. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

YouTube, Facebook, Instagram and Twitter during the period of physical distancing at home.[2]

Unfortunately, social media has not only been used to bring people closer and to share thoughts and opinions during the crisis, but also to spread false information. During the first days of the outbreak, a proliferation of fake news messages about COVID-19 went viral on private groups and messaging apps globally. These messages ranged from fake news about the origin of the virus (e.g. "the virus was maliciously created in a lab") or its mode of transmission (e.g. "the virus spreads through petrol pumps or the fifth-generation wireless 5G"), to methods of self-diagnoses (e.g. "breathe in deeply and hold your breath for 10 seconds").[3]

Misinformation and disinformation on social media have continuously become significant problems in the last few years.[4] Recent advances in digital media have permitted a greater connection between millions of users, empowering them to exchange original content in real-time on social media. Moreover, due to changing behaviours with regard to news consumption and the growth of mobile internet traffic, social media platforms are increasingly becoming the primary source of information for millions of people around the world.

However, the global network of computer-mediated communication (CMC) provided by different social media platforms, its decentralization and the limited oversight and restriction, have also facilitated the development and spread of false information.[5] These platforms tend to be widely accessible, easy to use and free, and in the event that a user's account is deactivated or banned, the user can move their content to other platforms or accounts in a short period at limited to no expense.

The COVID-19 crisis has further amplified the problem of malicious use of social media. Until April 2020, Reuters reported that almost four in ten (37%) individuals interviewed identified a lot or a great deal of misinformation about COVID-19 in social media like Facebook and Twitter, and 32% via messaging apps like WhatsApp.[6] Between January and March 2020, it was identified that 59% of misinformation was created by reconfiguring and recontextualizing information, while 38% was entirely fabricated; in most of the cases, the content was created with cheap and accessible software ("cheap-

---

2   Until March 2020, the worldwide social media consumption due to the coronavirus outbreak increased almost 45% due to the coronavirus outbreak, while an equal increase in 45% was reported on spending more time on messaging services. Survey was conducted among a number of individuals in different countries: 1,004 (Australia), 1,001 (Brazil),1,003 (China), 1,016 (France), 1,010 (Germany), 1,010 (Italy), 1,079 (Japan), 1,008 (Philippines), 1,008 (Singapore), 573 (South Africa), 1,005 (Spain), 1,040 (UK) and 1,088 (USA) internet users aged 16-64. Global Web Index (March 2020), *GWI Coronavirus Research | March 2020 Release 3: Multi-market research*. Available at: https://www.globalwebindex.com/hubfs/1.%20Coronavirus%20Research%20PDFs/GWI%20coronavirus%20findings%20March%202020%20-%20Multi-Market%20data%20(Release%203).pdf

3   The Center for Informed Democracy & Social - cybersecurity (IDeaS) of the Carnegie Mellon University has produced a list of stories containing inaccurate or misleading information on COVID-19. Available at: https://www.cmu.edu/ideas-social-cybersecurity/research/coronavirus.html

4   Disinformation is considered as "false and deliberately created to harm a person, social group, organization or country". Misinformation is considered as "false but not created with the intention of causing harm". See UNESCO (2018) *Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training*. Available at: https://en.unesco.org/fightfakenews

5   Computer-mediated communication (CMC) is defined as human communication that occurs through the use of electronic devices.

6   Newman N. et al (2020), *Reuters Institute Digital News Report 2020*, Reuters Institute for the Study of Journalism, p. 19. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

fakes") rather than highly technologically manipulated deepfakes.[7]

This report discussed the role of violent non-state actors, including terrorist, violent extremist and organized criminal groups, in maliciously using social media during COVID-19 to spread misinformation and disinformation. These actors have created and amplified misleading content on a large-scale, by taking advantage of vulnerabilities in the social media ecosystem and by manipulating people through conspiracy narratives and fake news.

The malicious use of social media by non-state actors, including terrorist, violent extremist and organized criminal groups is not new, as demonstrated by many past and well documented examples. In the last years, the Islamic State of Iraq and the Levant (ISIL, also known as Da'esh) has demonstrated its capacity to use social media to radicalize new affiliates, recruit foreign fighters and raise, move and channel funds. Al-Shabaab has created its first Twitter accounts and used them to provoke the Kenya Defence Force (KDF) and later the African Union Mission in Somalia (AMISOM) into long "Twitter duels". Several terrorist groups have created special units whose role is to amplify, glorify, and reinforce their messages, while other organizations have used social media to posture and self-promote online.[8] For example, organized criminal groups like the Sinaloa

cartel has a Twitter account (@carteidsinaloa) with more than 57,000 followers and the alleged account of El Chapo Guzman has more than half a million followers (@ elchap0guzman), whereas ISIL frequently uses *Ansar (*helpers) accounts for media distribution, operational coordination, and recruitment.[9]

This report further describes how terrorist, violent extremist and organized criminal groups are trying to take advantage of the coronavirus disease crisis to expand their activities and jeopardize the efficacy and credibility of governments' response measures.

The report is divided into three sections. The first section analyses how violent non-state actors, including terrorist, violent extremist and organized criminal groups, are spreading misinformation and disinformation. To prepare this section, UNICRI has been monitoring the malicious use of social media from January 2020 until the end of August 2020. In particular, UNICRI has analysed the activity of these groups on Facebook, Twitter, Telegram, YouTube, VK, Gab, BitChute, and their own websites.

The second section offers an analysis of existing technology options to detect and debunk false information. Although in the last years social media platforms and messaging apps have sought to remove and caution users about suspicious information,[10]

7   Sample of 225 pieces of misinformation rated false or misleading by fact checkers and published in English between January and the end of March 2020. See Brennen J. S., Simon F. M., Howard P. N., & Nielsen R. K. (2020), *Types, Sources, and Claims of COVID-19 Misinformation*, Reuters Institute for the Study of Journalism. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-04/Brennen%20-%20COVID%2019%20Misinformation%20FINAL%20(3).pdf

8   Reitano T. and Trabulsi, A. (2016), 'Virtually Illicit: The Use of Social Media in a Hyper-Connected World' in Matfess H. and Miklaucic M. (Edited by), *Beyond convergence: World without order*, pp. 215-233.

9   Ibid.

10 Social media platforms have also been committed to redirect users to accountable sources, creating special sections in their newsfeeds with updates on the COVID-19 situation. A case study in this regard is how Facebook, WhatsApp, and Instagram integrated this approach on their platforms (25 March 2020). Available at: https://about.fb.com/news/2020/03/combating-covid-19-misinformation/

increasing flows of misinformation and disinformation complicate this task.[11]

Of course, technology countermeasures alone are not sufficient to address the on-going problem of the malicious use of social media. In part, this is because such a challenge requires a direct engagement from civil society actors through empowerment measures aimed at increasing basic media literacy as well as critical thinking in approaching online contents. Moreover, violent non-state actors themselves have proven proficient at employing automated technology systems such as non-human accounts or social bots to deploy large-scale disinformation campaigns.

As a result, a technological race is emerging between those generating misleading content and those creating solutions to detect it.[12] Data scientists, journalists, fact-checkers, and developers are teaming up to combat disinformation and misinformation, assisting the public to identify false and distorted reports. The report has identified five different types of technology options to detect and debunk fake news.

The third and final section provides some practical recommendations to address the problem.

To validate results, UNICRI organized two virtual experts' meetings: the first was held on 7 April 2020 to discuss the malicious use of social media by violent non-state actors while the second was organized on 22 May 2020 to identify technology options and review their advantages and challenges. Both virtual meetings were attended by representatives from United Nations (UN) Member States, International Organizations, research institutions, technology companies and fact-checking and media companies.[13]

The report includes a number of samples of images produced by violent non-state actors. Those images contain hateful messages intended to incite discrimination towards minority groups, something that is fundamentally contrary to the core principles of the United Nations. We have, however, made the decision to include them in the report in an effort to help readers understand the nature and magnitude of the threat.

The report has been prepared by the UNICRI Knowledge Center *Security through Research, Technology and Innovation* (SIRIO).[14]

4

11 'Verified' initiative aims to flood digital space with facts amid Covid-19 crisis. UN Department of Global Communications (28 May 2020). Available at: https://www.un.org/en/coronavirus/%E2%80%98verified%E2%80%99-initiative-aims-flood-digital-space-facts-amid-covid-19-crisis

12 Hannah Murphy illustrates various threats posed by AI powered disinformation in Murphy H. (10 May 2020), 'The new AI tools spreading fake news in politics and business' in *the Financial Times*. Available at: https://www.ft.com/content/55a39e92-8357-11ea-b872-8db45d5f6714

13 The two virtual meetings were attended by representatives from Member States (Burundi, Georgia and United States of America), International Organizations (CERN, European Union Agency for Fundamental Rights, EUROPOL, Food and Agriculture Organization of the United Nations (FAO), Gavi the Vaccine Alliance, Geneva Centre for Security Policy (GCSP) and INTERPOL), private companies (Bodacea Light Industries, CDI Italia, Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE), Emerj Artificial Intelligence Research, Fondazione Bruno Kessler, International Alliance CBRN (INAC), Pierre Viaud Consulting, Strongpoint Security and THEOREM), Research Institutions (Cornell University, Organized Crime and Corruption Reporting Project, OSDIFE, Polytechnic University of Turin (POLITO),) as well as fact-checking and media companies (NewsGuard, Open and Poynter International Fact-Checking Network).

14 The scope of SIRIO is to analyze and understand the global impacts, opportunities and challenges of technological change, including in the areas of artificial intelligence (AI), robotics, augmented and virtual reality (AR, VR), big data analytics, digital biology and biotech, nanotech and digital printing, networks and computing systems, supply chain security and decentralized technologies such as blockchain.

> **Violent non-state actors, including terrorist, violent extremist and organized criminal groups, in maliciously using social media have created and amplified misleading content on a large-scale, by taking advantage of vulnerabilities in the social media ecosystem and by manipulating people through conspiracy narratives and fake news.**

Part 1

The threat

During the COVID-19 pandemic different violent non-state actors, including terrorist and violent extremist groups as well as criminal organizations have maliciously used social media. This section of the report analyses the profiles of these violent non-state actors, their types of messages, their strategic objectives and their tactics.

1.1 Who are the perpetrators?

The report analyses three groups of violent non-state actors which are particularly active in maliciously using social media during the pandemic: the right-wing extremist groups; the groups associated with the Islamic State

in Iraq and the Levant (ISIL or Da'esh) and Al-Qaida, and the organized crime groups.[15]

The right-wing extremist groups – also referred to as far-right – do not represent a coherent or easily defined movement, but, as stated by the United Nations Counter-Terrorism Committee Executive Directorate (CTED), they are rather a "shifting, complex and overlapping milieu of individuals, groups and movements (online and offline) espousing different but related ideologies, often linked by hatred and racism toward minorities, xenophobia, islamophobia or anti-Semitism".[16]

Groups associated with ISIL and Al-Qaida[17] have been also very present on social media, including organizations such as the Al-Qaida-aligned Thabat Media Agency,

---

15 This report does not consider groups and individuals that, although have actively contributed to the misinformation and disinformation on coronavirus through the social media, they do not represent a violent extremist organization or ideology.

16 United Nations Counter-Terrorism Committee Executive Directorate (CTED) (April 2020), *Trends Alert "Member States concerned by the growing and increasingly transnational threat of extreme right-wing terrorism"*, p. 2. See also the updated version of July 2020.

17 The Security Council Committee pursuant to resolutions 1267 (1999), 1989 (2011) and 2253 (2015) concerning ISIL (Da'esh), AlQaida and associated individuals, groups, undertakings and entities (called Security Council ISIL (Da'esh) & Al-Qaida Sanctions Committee) updates regularly a Sanction List of individuals and entities subject to the assets freeze, travel ban and arms embargo set out in paragraph 1 of Security Council resolution 2368 (2017), and adopted under Chapter VII of the Charter of the United Nations. On 16 July 2020, the Sanctions List contained the names of 261 individuals and 89 entities cf. https://www.un.org/securitycouncil/sanctions/1267/aq_sanctions_list. For the purpose of this report, the list has been considered as main source to identify the organizations that are part of this second group of violent non-state actors.

pro-ISIL Al-Qitaal Media Center, Al-Qaida Central's media outlet Al-Sahab Foundation, Abubakar Shekau (Boko Haram faction leader) and Al-Shabaab.

A third group of violent non-state actors is represented by organized criminal groups. Above all, narco cartels in Mexico such as the Gulf Cartel and the Cartel Jalisco Nueva Generación (CJNG) and criminal organizations in Italy such as Cosa Nostra and 'Ndrangheta have been active in the misuse of social media.

Obviously, there are many differences in the strategic objectives of these three groups. The right-wing extremists are promoting accelerationist terrorism to cause a race war, accelerate the "inevitable" collapse of society and replace it with a white ethnostate, a state limited to white people. Groups associated with ISIL and Al-Qaida also consider violence as a legitimate tool to advance and impose their values and ideas as they did with ISIL's 'caliphate' proclamation in Iraq. Organized criminal groups are engaged in illegal activity not for ideological reasons but rather for profit. However, as it will be shown, there are similarities in the way the three different categories of violent non-state actors have been using social media during the pandemic.

1.2 What types of messages?

Terrorist, violent extremist and organized criminal groups have maliciously used social media to disseminate **conspiracy theories** about the origin of COVID-19. These conspiracy theories usually attribute the origin of the virus to governments, religious or ethnic groups, secret networks, companies or businessmen who, according to these interpretations, are trying to push through secret agendas such as globalist depopulation, the control of the world or the generation of financial incomes through the sale of already produced vaccines and drug treatments.

The messages are often customised to match with the audience and ideologies of the violent non-state actors. For example, right-wing extremist groups have circulated conspiracy theories that blame immigrants and foreigners as the ones responsible for spreading the virus. The New Jersey European Heritage Association (NJEHA) shared a campaign where they placed stickers with slogans such as "Stop coronavirus – deport all illegal aliens", "migrants accepted no – we are infected", "open borders is the virus", "multicultural is the virus", "open borders spread disease" around the city.

► **Source:** Screenshot taken from the New Jersey European Heritage Association (NJEHA) Gab channel

► **Source:** Screenshot taken from the CoronaWaffen Telegram channel
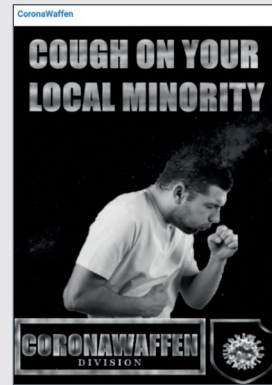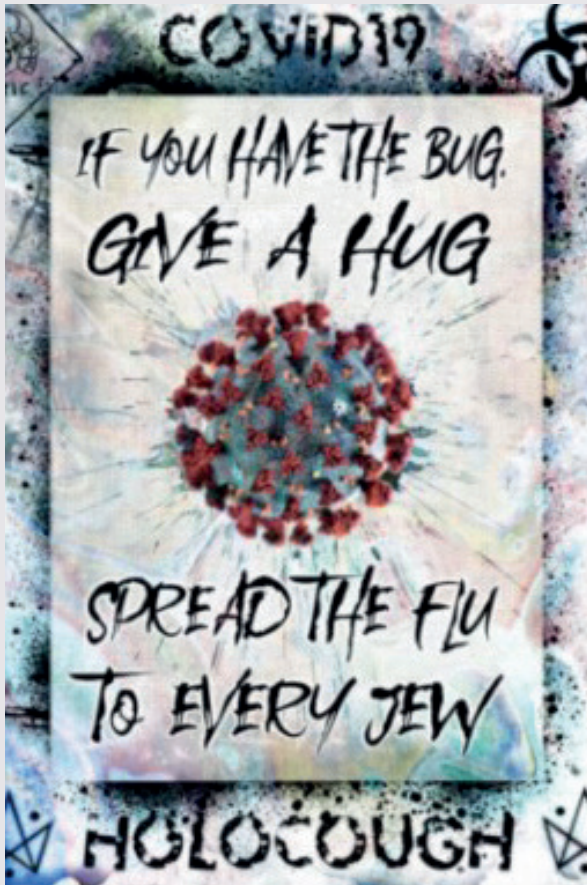
Other groups that have adopted a similar narrative include Blanche Europe and the online based Eco-Fascist Central, Corona Chan News and Corona Waffen.[18] They have also entwined the traditional anti-Semitic and Islamophobic narratives with fallacious theories related to the pandemic. Conspiracy theories such as Accelerationism, QAnon, and Boogaloo have been also used to misrepresent the COVID-19 pandemic.[19] The Atomwaffen Division (AWD) in the United States of America and its regional counterparts such as Feuerkrieg Division the Baltic states, Sonnenkrieg Division in the United Kingdom, AWD Deutschland, and the Northern Order in Canada also used their social media channels to spread accelerationist messages.[20] Other groups have also spewed fallacious theories that identified 5G broadband system as responsible for the

18 Blanche Europe has published articles in their website claiming that the solution to the current health crisis is "exterminating" immigrant populations and barring ethnic minorities from receiving medical treatment; the Telegram accounts Eco-Fascist Central, Corona Chan News and Corona Waffen continuously posted content spreading Sinophobia, anti-Semitism and racism against local minorities.

19 QAnon is a far-right conspiracy theory that claims that there is an on-going secret plot by an alleged "deep state". Boogaloo is an extreme right-wing anti-government movement referring to an impending civil war in the U.S.A. See Institute for Strategic Dialogue Digital Research Unit (9 April 2020), *COVID-19 Disinformation Briefing: Far-Right Mobilisation*. Available at: https://www.isdglobal.org/wp-content/uploads/2020/04/Covid-19-Briefing-PDF.pdf

20 Southern Poverty Law Center. (n.d.). *Atomwaffen Division*. Available at: https://www.splcenter.org/fighting-hate/extremist-files/group/atomwaffen-division

▶ **Source:** Screenshots taken from different Telegram channels

virus or that COVID-19 is "mother nature's terrorist".

Groups associated with ISIL and Al-Qaida have also spread conspiracy theories that assert that the virus is a "soldier of *Allah*" that is punishing the unbelievers and the enemies that have damaged Muslims over the last years. For example, ISIL and Al-Qaida claimed that the virus is God's wrath upon the West.[21] Similarly Al-Shabaab declared that the coronavirus disease is spread by "the crusader forces who have invaded the country and the disbelieving countries that support them".[22]

In all these cases, violent non-state actors claim to possess "real" knowledge about the origin of COVID-19 and prophesise that the virus will hasten the self-destructive tendencies of an existing governmental system causing ultimately its collapse and the creation of a new society in which their enemies will be eliminated.

21 Meek J. G. (2 April 2020), 'Terrorist groups spin COVID-19 as God's 'smallest soldier' attacking West' in *ABC News*. Available at: https://abcnews.go.com/International/terrorist-groups-spin-covid-19-gods-smallest-soldier/story?id=69930563

22 BBC News (1 March 2020), *Coronavirus: Fighting al-Shabab propaganda in Somalia*. Available at: https://www.bbc.com/news/world-africa-52103799

1.3 What are the strategic objectives and targets?

Non-state actors are trying to spread misinformation and disinformation to pursue different strategic objectives. The first objective is to **undermine trust in the government** and, at the same time, **reinforce non-state actors' extremist narratives and recruitment strategies**.

Some right-wing extremist groups such as the Nordic Resistance Movement (NRM) in Sweden, Kohti Vapautta! (KV) in Finland and Blanche Europe in France have shaped the pandemic in their "usual" narrative to reinforce their existing ideology and to increase recruitment. For example, the NRM claimed that COVID-19 can be used to strengthen the movement since what they define as "mismanagement" of the government (the financial support of immigrants, the anti-white system, corruption, and the lack of support of the white worker) will worsen the economic crisis. In this scenario, the NRM is asking people to join the group to fight the upcoming socio-political and economic crisis. KV stated that the current pandemic will originate an economic depression that will provide a great opportunity for national socialist movements to take advantage of the situation. Blanche Europe declared that the solution to the current health crisis is "exterminating" immigrant populations and barring ethnic minorities from receiving medical treatment. In addition, they blame the government of France for not acting quick enough to prevent the crisis and for not prioritizing "white" families.[23]

Groups associated with ISIL and Al-Qaida have shown similar strategic goals. The pro-ISIL Al-Qitaal Media Center shared a message in the second issue of the online magazine *Sawt al-Hind* (*Voice of Hind/India*) claiming that the virus is a divine punishment that would not affect the believers. The magazine is contributing to increase the hate speech and attacks against Hindus, while the message seeks to distort the nature of the virus by relabelling it as a "divine" matter and not a real health crisis. Al-Shabaab in Somalia also spread disinformation distorting the nature of the virus and using it as a motivation to continue its violent attacks. The spokesman asked for the expulsion of all foreign forces after stating that the unbelievers are deliberately spreading COVID-19. They particularly blamed the African Union Mission in Somalia (AMISOM) since the first cases of COVID-19 were reported in the Halane base.

A second objective of the non-state actors is to increase the **"inspired terrorism"** or the motivation of **self-radicalized terrorists** in order to perpetrate real attacks. There are cases in which right-wing extremist groups, like CoronaWaffen, explicitly asked their followers to spread the virus by coughing on their local minority or by attending to specific places where religious or racial minorities gather.

Other groups, such as Eco-Fascist Central, advocate to spread the coronavirus disease in countries with large populations or high levels of pollution. Additionally, CoronaWaffen has posted several online surveys on social media and messaging apps asking people if they had spread the virus.

Furthermore, the Global Fatwa Index has identified COVID-19 related messages from groups associated with ISIL and Al-Qaida, including non-official *fatwas* that called on ISIL members who have contracted COVID-19 to act as "biological bombs" by delib-

---

23 See Blanche Europe website available at: https://www.blancheurope.com
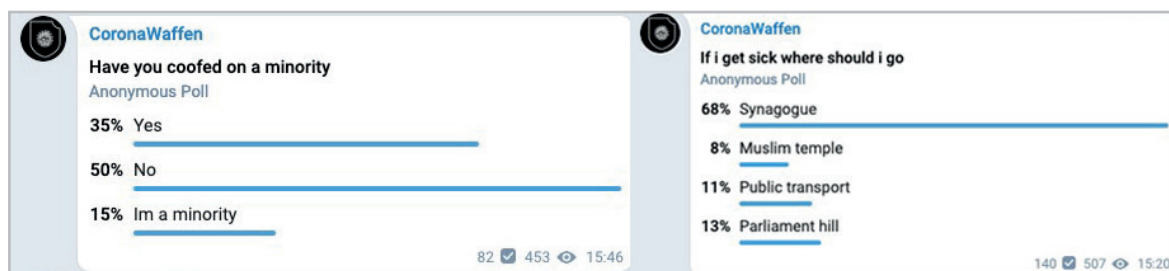
> ▸ **Source:** Screenshots taken from the CoronaWaffen Telegram channel

erately spreading the disease among the organization's enemies.[24]

An example of "inspired terrorism" is represented by Timothy Wilson who, on 24 March 2020, was shot by the United States Federal Bureau of Investigation (FBI) in Kansas City when he plotted to detonate a bomb in a hospital caring for coronavirus patients. Wilson was active in at least two neo-Nazi Telegram channels and kept communication with an Army infantry soldier who wanted to plan an attack on a major American news network and discussed targeting a Democratic presidential candidate.[25] Wilson's last online comment was an anti-Semitic message regarding the origin of COVID-19. The online influence of these extremist groups can also be observed in a series of cyberattacks perpetrated in April 2020, in which

25,000 email addresses and passwords supposedly belonging to the National Institutes of Health, the World Health Organization, the Gates Foundation and other groups working to combat the coronavirus disease pandemic were allegedly posted online by right-wing groups.[26]

A third strategic objective is to promote a **positive image** of the organizations during the pandemics as a viable replacement of healthcare institutions and as a responsible political actor. In some cases, non-state actors have exploited the additional grievances resulting from a fragile socio-economic situation due to the crisis, which has led to an increase of prices and reduction of food availability. To bolster their presence in their territories, Al-Shabab broadcasted on their

11

24 The Middle East Media Research Institute (2020), *Egypt's Official Fatwa-Issuing Body Warns Against Extremist Fatwas On Coronavirus, Calls To Follow Instructions Of Medical Establishment*. Available at: https://www.memri.org/reports/egypt%E2%80%99s-official-fatwa-issuing-body-warns-against-extremist-fatwas-coronavirus-calls-follow

25 Levine M. (26 March 2020), 'FBI learned of coronavirus-inspired bomb plotter through radicalized US Army soldier' in *ABC News*. Available at: https://abcnews.go.com/Politics/fbi-learned-coronavirus-inspired-bomb-plotter-radicalized-us/story?id=69818116 See also Silke A. (2020), *COVID-19 and terrorism: assessing the short-and long-term impacts*, p. 5. Available at: https://www.poolre.co.uk/wp-content/uploads/2020/05/COVID-19-and-Terrorsim-report-V1.pdf

26 Mekhennet S. and Timberg C. (22 April 2020), 'Nearly 25,000 email addresses and passwords allegedly from NIH, WHO, Gates Foundation and others are dumped online' in *The Washington Post*. Available at: https://www.washingtonpost.com/technology/2020/04/21/nearly-25000-email-addresses-passwords-allegedly-nih-who-gates-foundation-are-dumped-online/

radio station Andalus that they have opened a coronavirus treatment centre in Somalia.[27]

Although from a different perspective, criminal organizations have also taken advantage of COVID-19 to portray a positive image of themselves and reinforce their presence and control in the territory.

A number of organized criminal groups have traditionally attempted and succeeded in maintaining the monopoly of illegal activities at the local level, by imposing protection on all types of economic businesses and corrupting local political institutions. To enforce this form of territorial control, these criminal groups often need to build the image of a sort of "state within the state", able to provide aid and support to the local community and, at the same time, discourage any form of criticism and dissent through intimidation and violence.

In this regard, the COVID-19 crisis represented an excellent opportunity to reinforce this rhetoric and promote the idea that, in the moment of emergency, criminal groups can replace the government and perform its role. For example, some criminal groups have attempted to perform the role of the government and official institutions within territories where they have a strong presence by adopting strict health measures, such as lockdowns, or directly supporting the population with sanitizers and food. However the main goal of the criminal groups is not to protect the local population but rather to protect their criminal interests since they are

concerned that a large health crisis could cause the arrival and the active involvement of the law enforcement agencies or the army in the areas under organized crime control and, as a result, jeopardize their illegal activities.

These "services" have been largely publicized through social media. For example, in Mexico, criminal groups have labelled the aid boxes that they have distributed with the name or logo of the criminal organization and, subsequently, promote their actions in social media and in news reports. This is the case of the Gulf Cartel that distributed aid packages containing food and sanitizers in Tamaulipas, placing in each box a sticker that indicates the name of the cartel and of its leader.[28] One of the daughters of Joaquín Guzmán Loera (El Chapo), the historical leader of the Sinaloa Cartel, also distributed groceries with her father's image to citizens in Guadalajara, Jalisco.[29]

Mexican cartels have also posted pictures on social media taken while they were distributing aid packages. Similar cases have also taken place in other Latin American countries as well as in Italy, South Africa and Japan. After distributing food in a neighbourhood of Palermo, Sicily, the brother of a drug-trafficking leader appealed for others on Facebook to follow his example. When a journalist reported the news in an Italian newspaper, the brother of the criminal emphatically responded on Facebook that "the State does not want us to do charity because

27 Al Jazeera (14 June 2020), Al-Shabab sets up coronavirus treatment centre in Somalia. Available at https://www.aljazeera.com/news/2020/6/14/al-shabab-sets-up-coronavirus-treatment-centre-in-somalia.

28 Fajardo L. (21 April 2020), 'Coronavirus: Latin American crime gangs adapt to pandemic' in BBC News. Available at: https://www.bbc.com/news/world-latin-america-52367898

29 de Córdoba J. (20 April 2020), 'New Face of Mexico Charity: Drug Lord 'El Chapo' in The Wall Street Journal. Available at: https://www.wsj.com/articles/new-face-of-mexico-charity-drug-lord-el-chapo-11587391200
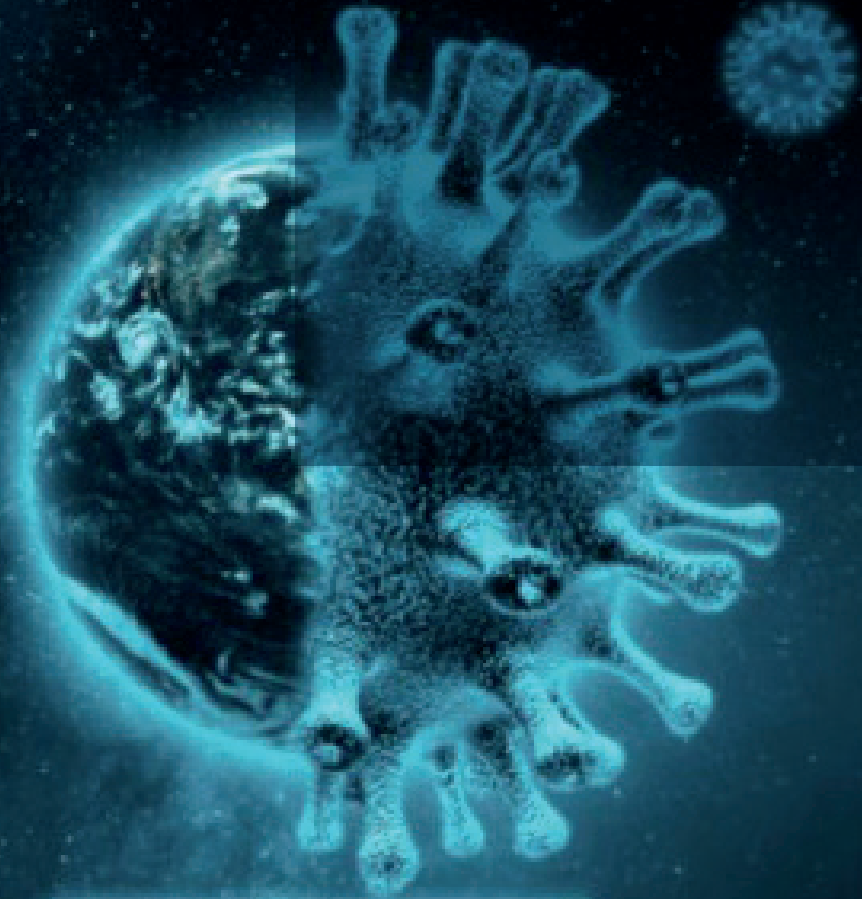
► Cartel del Golfo distributing packages. **Source:** Infobae[30]

► CJNG distributing packages in a TikTok video. **Source:** Infobae[31]

we are Mafiosi" and then he threatened the journalist. [32]

It is also important to stress that the current economic crisis created by COVID-19 is greatly facilitating the possibilities that criminal groups acquire direct or indirect control and ownership of legitimate companies. The financial crisis and the potential bankruptcy of several enterprises, shops and economic activities, could represent opportunities for organized crime to penetrate and infiltrate the legal economy. Cases have already been registered where criminal groups are approaching entrepreneurs facing financial difficulties to purchase their activity or to offer them a loan. In these cases, organized crime directly aims at obtaining control of economic activities by imposing "loan shark" at extremely high interest rates, that will be impossible to repay and will ultimately force entrepreneurs to alienate property or control of their company to a crime group affiliate or strawman. For example, in Mexico, the Cartel Jalisco Nueva Generación (CJNG), in collaboration with other cartels like La Nueva Familia Michoacana and Los Granados, increased the production of counterfeit medicines in the country with the intention of forcing small and medium pharmacies to sell their companies to the drug cartels in the states of Guanajuato, Jalisco, Guerrero and Michoacan.[33] A recent case in Italy also

14

30 Infobae (20 April 2020), Narcos aprovechan coronavirus en México para repartir despensas y pelear territorio. Available at: https://www.infobae.com/america/mexico/2020/04/20/narcos-aprovechan-coronavirus-en-mexico-para-repartir-despensas-y-pelear-territorio/

31 Infobae (10 May 2020), El narco en TikTok: el CJNG desafía al gobierno y alardea entregando despensas. Available at: https://www.infobae.com/america/mexico/2020/05/10/el-narco-en-tiktok-el-cjng-desafia-al-gobierno-y-alardea-entregando-despensas/

32 La Repubblica (8 April 2020), *Palermo, minacce all'inviato Salvo Palazzolo. Insulti dopo il post del fratello del boss. Decine di messaggi di solidarietà*. Available at: https://palermo.repubblica.it/cronaca/2020/04/08/news/il_fratello_del_boss_su_fb_orgoglioso_di_essere_mafioso_giornalisti_peggio_del_coronavirus_-253480726/

See also Roberts H. and Barigazzi J. (9 April 2020), 'Mafia plots post-coronavirus pounce' in *Politico*. Available at: https://www.politico.eu/article/mafia-plots-post-coronavirus-pounce/

33 Fiscalía General de la República, Seguridad y Defensa (17 March 2020), *CJNG, el principal distribuidor de medicamentos pirata del país*. Available at: https://seguridadydefensa.mx/nacionales/cjng-el-principal-distribuidor-de-medicamentos-pirata-del-pais/

shows how organized crime is applying this strategy, where more than 100 people were investigated by the police on accounts of usury and extortion targeting economic activities facing a crisis after the lockdown.[34]

1.4 What are their tactics?

The violent non-state actors have taken advantage of the inherent characteristics of social media and messaging apps where contents can be uploaded anonymously, instantly, and at no cost.

The most common tactic used especially by right-wing extremist groups to attract followers is to create simple and highly visual content, including "Internet memes".[35] An Internet meme consists of a phrase, image or video that spreads rapidly from person to person via the Internet through social media channels and messaging apps. Memes often intend to elicit humour to facilitate its spread.[36] Memes have become a social phenomenon to promote ideas, behaviour

or style, and this is the main reason why violent non-state actors have been attracted by them. Unfortunately, they use memes as a method to share antisemitic, xenophobic and radical content with a potential large audience. For example, CoronaWaffen has used this format to suggest the intentional transmission of the virus in a *meme* format that was circulating online.

During the pandemic, several groups used simple tactics to transmit their propaganda. For example, criminal organizations such as the Cartel Jalisco Nueva Generación (CJNG) have uploaded videos and pictures to promote a positive image of themselves, whereas in another example, the Turkestan Islamic Party (TIP) distributed a video stating that COVID-19 is a divine punishment for China's treatment of Uyghurs. Similarly, groups such as pro-ISIL Al-Qitaal Media Center[37], Al-Qaida-aligned Thabat Media Agency[38], Al-Qaida's Al-Sahab Foundation[39] and ISIL[40] produced messages regarding the nature of the virus. In Africa, Al-Shabaab[41] and Abubakar Shekau, a Boko Haram fac-

34 La Repubblica (20 May 2020), *Usura, nel Barese oltre 100 denunce durante emergenza Covid: "Scenario allarmante"*. Available at: https://bari.repubblica.it/cronaca/2020/05/20/news/usura_100_denunce_bari-257161552/. Cases include La Unión Tepito in Mexico City and Camorra in the south-western region of Campania. Camorra had a similar distribution strategy with a model with the intent of laundering money in legal businesses.

35 The evolutionary biologist Richard Dawkins introduced the term meme (from the Greek mimema, meaning "imitated") in 1976 as a unit of cultural transmission spread by imitation.

36 Puche-Navarro R. (2004), 'Graphic Jokes and Children's Mind: An Unusual Way to Approach Children's Representational Activity' in *Scandinavian Journal of Psychology*, pp. 45, 343-355.

37 Tony Blair Institute for Global Change (April 2020), *Snapshot: How Extremist Groups are Responding to Covid-19*. Available at: https://institute.global/policy/snapshot-how-extremist-groups-are-responding-covid-19-9-april-2020

38 Mazzoni V. (20 March 2020), 'Coronavirus: How Islamist Militants Are Reacting to the Outbreak' in *European Eye on Radicalization*. Available at: https://eeradicalization.com/coronavirus-how-islamist-militants-are-reacting-to-the-outbreak/

39 The Middle East Media Research Institute (1 April 2020), *Al-Qaeda Central: COVID-19 Is Divine Punishment For Sins Of Mankind; Muslims Must Repent, West Must Embrace Islam*. Available at: https://www.memri.org/jttm/al-qaeda-central-covid-19-divine-punishment-sins-mankind-muslims-must-repent-west-must-embrace

40 Tony Blair Institute for Global Change (April 2020), *Snapshot: How Extremist Groups are Responding to Covid-19*. Available at: https://institute.global/policy/snapshot-how-extremist-groups-are-responding-covid-19-9-april-2020

41 The Middle East Media Research Institute (29 April 2020), *Al-Shabab Spokesman Says Unbelievers Intentionally Spread Coronavirus In Somalia, Calls For Expulsion Of Foreign Forces*. Available at: https://www.memri.org/jttm/al-shabab-spokesman-says-unbelievers-intentionally-spread-coronavirus-somalia-calls-expulsion

tion leader, distributed audios with similar content.[42]

Violent non-state actors can also exploit functions and services provided by the social networking sites' platforms. For example, mainstream social media platforms, such as Facebook or Twitter, work with algorithms that suggest expanding your own network and finding new contacts ("friends") based on criteria such as mutual friends, work and education. As has been already done by ISIL in the recent years, violent non-state actors, in particular right-wing groups, have taken advantage of these algorithms during the pandemic to contact "suggested friends" and recruit new members. As a result, members of the violent extremist groups frequently participate in different forums and groups on social media in an attempt to radicalize individuals and find other groups or individuals with whom they share similar extremist views.

Some groups are also trying to bypass control measures in the main social platforms by avoiding the use of certain words or symbols that can be easily spotted as part of the "extremist language" and even attempting to look legitimate to a large audience.[43]

Some violent non-state actors have also been shown resilience by spreading disinformation even after their accounts have been eliminated. A possible tactic is to reproduce the same contents by creating new accounts on social media platforms. Another tactic is to redirect followers and visitors to less-controlled and encrypted channels (out-linking), such as Telegram, VK, Gab or websites. This

the case of the 'boogaloo movement' that first posted contents on Facebook to attract followers, and then invited them to join discussions on messaging app networks such as Telegram channels, where it was possible to share more extreme materials and attempt to radicalize and recruit users.

The use of less-controlled and encrypted channels started before the COVID-19 crisis. According to a report by the European Council on Foreign Relations, Twitter's suspension of more than 200,000 extremists' accounts in August 2016 resulted in an "online exodus to end-to-end encrypted messaging services like Telegram, WhatsApp and Viber".[44] Violent extremists have been attracted by the significant freedom in terms of contents provided by these messages services which, at the same time, make it more difficult for law enforcement agencies to monitor extremist activities. It is likely that this trend will be further reinforced with the continuation of the pandemic.

Another important aspect is the malicious use of "social bots" or "chatbots". A social bot is a computer algorithm that automatically produces content, interacts with humans on social media and attempts to influence their opinion and behaviour. Social bots are increasingly becoming an essential tool for orchestrated, large-scale disinformation campaigns on social media. A recent study from Carnegie Mellon University observed that 45% of the Twitter accounts sharing more than 200 million messages on coronavirus are likely to be social bots. Those accounts have fuelled over 100 false narratives about the pandemic between January and May

16

42 Campbell J. (April 2020), 'Boko Haram's Shekau Labels Anti-COVID-19 Measures an Attack on Islam in Nigeria' in *Council of Foreign Relations*. Available at: https://www.cfr.org/blog/boko-harams-shekau-labels-anti-covid-19-measures-attack-islam-nigeria

43 For example see Miller-Idriss C. (2017), *The Extreme Gone Mainstream: Commercialization and Far Right Youth Culture in Germany*. See also Ebner J. (2020), *Going Dark: The Secret Social Lives of Extremists*.

44 Soesanto S. & D'Incau, F. (2017), 'Countering online radicalisation' in *European Council on Foreign Relations*. Available at: https://www.ecfr.eu/article/commentary_countering_digital_radicalisation_7216
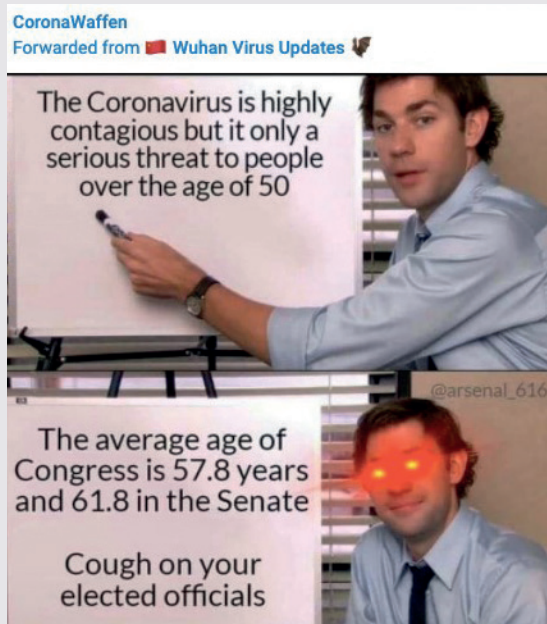
▶ **Source:** Screenshot taken from the CoronaWaffen Telegram channel

▶ **Source:** Screenshot taken from the Eco-Fascist Central Telegram channel

2020.[45] Several studies demonstrate that right-wing extremist groups and groups associated with ISIL and Al-Qaida possess sufficient skills and knowledge to maliciously use social bots to promote their rhetoric and radicalise new affiliates.[46]

---

45 Allyn B. (20 May 2020), 'Researchers: Nearly Half of Accounts Tweeting About Coronavirus Are Likely Bots' in *NPR*. Available at: https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots?t=1591623690495

46 See for example Gambetta D. and Hertog S. (2016), *Engineers of Jihad: the curious connection between violent extremism and education* or Caldarelli G., De Nicola, R., Del Vigna, F. et al. (2020), 'The role of bot squads in the political propaganda on Twitter' in *Commun Phys* 3, 81 (2020). https://doi.org/10.1038/s42005-020-0340-4

18

Technology options to combat online disinformation and misinformation

Accurate, accountable, and reliable information remains critical to step up collective efforts to contain the transmission of COVID-19. Considering that social media has become a primary source of information for millions of people around the world, an effective anti-COVID strategy must include the use of technology to prevent and combat online disinformation and misinformation.

In this section of the report, UNICRI has identified five different types of technology to detect and debunk fake news amid the COVID-19 crisis, while also illustrating the advantages and possible challenges of each technology option in the short and long term.

2.1 Data science/Big Data visualization to identify the spread of large-scale disinformation

Experts are developing new solutions to identify the spread of large-scale disinformation using **data science, big data visualization, and machine learning algorithms**. These technologies help researchers visualize the spread of disinformation and, potentially, track down the origin of false narratives. Data-driven techniques allow experts to extract information from millions of human and social bots, find similar texts and visualize misinformation themes. In many cases, social media platforms[47] as well as independent data scientists,[48] are training and using machine learning to identify specific patterns of language and using it to detect a higher volume of texts and multimedia

---

47 Lyons T. (21 July 2018), 'Increasing Our Efforts to Fight False News' in *Facebook*. Available at: https://about.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/

48 Lupi V. (10 March 2020), 'Covid-19 and Fake News in the Social Media' in *FBK*. Available at: https://www.fbk.eu/en/press-releases/covid-19-and-fake-news-in-the-social-media/

artifacts. The analysis is done by researchers that examine the collected information, using standards and specific diagnostic frameworks for categorizing orchestrated disinformation attacks.

This technology option enables fact-checkers, journalists, and content moderators from social media platforms to visualize the spreading of systematic false information and super-spreaders quickly, and make timely decisions to verify, flag, or remove content. However, one of the main challenges of this type of technology is that machine-learning algorithms may get wrong results, especially if the misinformation campaigns adopt more sophisticated strategies. For instance, AI-powered text collection may encounter data quality issues as AI text generators may be able to defy detection.[49] In addition, human biases, especially during the process of tagging data (data labelling), may impair algorithmic decisions. Therefore, the accuracy of the results needs to be continuously assessed. Practitioners recognize that automated systems alone cannot assure full detection of disinformation and misinformation.

As disinformation campaigns grow faster and turn into more complex systematic operations, a future trend could be the development of machine-assisted and crowd-based strategies to scale the quantity and the quality of verified information.[50] The involvement of qualified fact-checkers with solid journalistic ethics in this process should not be overlooked.

2.2 Artificial Intelligence tools and platforms to detect fake news online

Due to the COVID-19 pandemic, news consumption on social media has reached new heights. In the early stages of the outbreak, the World Health Organization warned the international community about a growing "infodemic" in the social media ecosystem, noting that hoaxes circulated faster than the virus itself. To counter this threat, developers have created AI systems to guide the public in detecting misinformation and disinformation as well as sharing accurate news reports.

A noteworthy trend in this field is the growing presence of **websites and platforms** that allow individuals to read AI-filtered information, by performing content verification with specific algorithms. These algorithms often validate news with a credibility score that is produced by a variety of factors inspired by journalistic standards, including the history of the media, author's expertise, sources present in the article, tone of the item (partial/impartial), and the political leaning of the author.[51]

The advantage of this technology option is that it is a scalable solution to detect in the shortest time possible thousands of relevant and accurate articles in real-time, as well as potentially malicious content, including deep-fake videos. The majority of these platforms target the general audience and, in particular, those who directly consume

49 Ter, S. (18 April 2020), 'Disinformation and AI for Good' in *Medium*. Available at: https://medium.com/@sara-jayneterp/disinformation-and-ai-for-good-d4e525be239

50 As stated by experts during the UNICRI Teleworkshop: *Technology solutions to combat fake news during COVID-19* (22 May 2020).

51 A case study in this field is the Factual website works: https://www.thefactual.com/how-it-works.html

essential information via their social media newsfeeds. Therefore, most of these sites deliver the best user experience, featuring mobile-friendly websites/apps that guarantee a smooth reading experience.

One of the main challenges this tech option faces is how to incorporate human judgment in verifying the data. Some developers are removing design journalists and human judgment from the fact-checking equation as if algorithms were better equipped than humans to verify false information. This assumption could be wrong as AI algorithms may include human biased decisions affecting the machine learning process. Human bias can affect algorithms depending on how the problem has been framed, the training data collected, and which attributes are considered.[52] Furthermore, as several social media companies have relied increasingly on AI systems to combat the spreading of fake news, instead of human moderators, the quality of news verification has not increased necessarily.[53] Many companies that are still experimenting these services do not always guarantee full transparency on how their algorithms select and promote news, and to what extent the privacy of their users is respected.

2.3 Mobile apps and chatbots powered by fact-checkers targeting the general public

On private messaging applications, fake news on COVID-19 have gone viral, putting at risk many lives. Orchestrated disinformation campaigns, misleading and doctored content on how to prevent and cure the coronavirus were amplified also by users without malicious intent.[54] As a consequence, a misinformation ecosystem rose to spread fake alarmist messages, screenshots, doctored documents from authorities, decontextualized images, and videos.[55]

To combat this worrying trend, public health agencies, fact-checking organizations, and messaging platforms developed automated **chatbots** to debunk false information. These chatbots, readily available on the most popular messaging platforms, target the general public and aim at helping users differentiate between facts and falsehood related to the coronavirus disease pandemic. In short, chatbots are software programs designed to interact with humans and work autonomously. The majority of these applications programmed during the COVID-19

20

---

52 Hao K. (4 February 2019), 'This is how AI bias really happens – and why it's so hard to fix' in *MIT Technology Review*. Available at: https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

53 Pazzanese C. (8 May 2020), 'Battling the 'pandemic of misinformation'' in *The Harvard Gazette*. Available at: https://news.harvard.edu/gazette/story/2020/05/social-media-used-to-spread-create-covid-19-falsehoods/

54 WhatsApp statement: "We've seen a significant increase in the amount of forwarding which users have told us can feel overwhelming and can contribute to the spread of misinformation. We believe it's important to slow the spread of these messages down to keep WhatsApp a place for personal conversation." WhatsApp Blog (7 April 2020), *Keeping WhatsApp Personal and Private*. Available at: https://blog.whatsapp.com/Keeping-WhatsApp-Personal-and-Private

New WhatsApp guidelines issued in April 2020 allow users to forward a message with up to five chats at one time. In 2018, a user could pass on a forwarded message to 250 groups at once. See Hern A. (7 April 2020), 'WhatsApp to impose new limit on forwarding to fight fake news' in *The Guardian*. Available at: https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news#maincontent

55 EU Disinfo Lab Blog (2 April 2020), *COVID-19 Disinformation: Narratives, Trends, and Strategies in Europe*. Available at: https://www.disinfo.eu/publications/covid-19-disinformation-narratives-trends-and-strategies-in-europe/

crisis allow users to learn more about viral fake news, mythbusters, and good practices to prevent the spread of the virus. Users can connect with chatbots by inserting a standard phone number in the messaging app, and then start texting. The user interface features a short numerical menu that allows users to interact with it by typing text or emoticons. Accountable organizations such as WHO and the Poynter's International Fact-Checking Network power their chatbots with an updated database of verified reports [56] and thousands of debunked fake news.[57]

The most significant advantage of this technology option is that it is available for all smartphones and popular messaging services, making the dissemination of accurate information scalable and available 24h/7 for billions of people. However, as AI does not power most of these chatbots, the interaction between humans and machines is based on fixed patterns rather than being conversational. As a consequence, if the user goes off script and does not receive a compelling reply from the chatbot, they may get frustrated and quit the chat. Overall, there is still room for improvement in designing AI-powered chatbots that can tailor comments and deliver better response customization in this field.

Furthermore, governments and big tech giants are consistently expanding their funding for fact-checking organizations to combat COVID-19-related misinformation.[58] Nevertheless, sustainability remains an existential issue for the fact-checking community and for the technology option. In fact, it is possible that, once the pandemic will be over, preventive measures against the next "infodemic" might be defunded.

2.4 Web-browser extension for the general public

The objective of web-browser extensions and platforms is to monitor and verify the reliability of **online news sites and influential social media accounts**. The overall focus of these tools aims at providing accurate content information and tracking its expansion online. For instance, although extremely difficult, some organizations have been successful in tracking false claims relating to the coronavirus back to its source.[59]

Extensions are software programs that can modify the user interface and provide additional features within the browsing experience. Browser add-ons range from translating content, identifying security threats to tracking cookies. In this case, integrated extensions can help users in tracking the accountability of news sites and social media feeds, by reviewing and examining the validity of the information presented.

21

56 WHO (15 April 2020), 'WHO launches chatbot on Facebook Messenger to combat COVID-19 misinformation' in *WHO Newsroom*. Available at: https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation

57 Grau M. (4 May 2020), 'New WhatsApp chatbot unleashes power of worldwide fact-checking organizations to fight COVID-19 misinformation on the platform' in *Poynter*. Available at: https://www.poynter.org/fact-checking/2020/poynters-international-fact-checking-network-launches-whatsapp-chatbot-to-fight-covid-19-misinformation-leveraging-database-of-more-than-4000-hoaxes/

58 The International Fact-Checking Network (17 March 2020), 'Flash grants of up to $50K are now available for fact-checkers fighting coronavirus misinformation' in *Poynter*. Available at: https://www.poynter.org/fact-checking/2020/flash-grants-of-up-to-50k-are-now-available-for-fact-checkers-fighting-coronavirus-misinformation/

59 Deutch G. (19 March 2020), 'How One Particular Coronavirus Myth Went Viral' in *Wired*. Available at: https://www.wired.com/story/opinion-how-one-particular-coronavirus-myth-went-viral/

Web-browser extensions can support effective strategies in providing users with the necessary information to detect and thus *pre-debunk* disinformation present on a plethora of websites in an efficient manner. Once installed, the add-on will assign a visual credibility score to news links and information sources, including some highlights on the transparency of the publication and its history.

This technology option, differently from others, is essentially built on the work of hundreds of fact-checkers that based their accountability analysis on journalistic standards and ethics. This allows users to gain essential information through an approach that holistically evaluates news websites and social media performance across various accounts. Web-browser extensions and verification tools can thus have a positive and scalable impact in allowing users to freely make their *own* assumptions as well as allowing for concise and critical conclusions based on the information provided on different platforms.[60] These tools could further encourage individuals to think critically about the sources of information and guide them in understanding that not all information online is created equally or fairly.[61]

One of the challenges faced by this technology is its adoption: browser extensions are usually employed by an already quite informed audience. As a result, it is likely that the target groups directly affected from disinformation campaigns would not download the tool due to their limited digital literacy. Therefore, it is key to improve

interoperability and user experience (UX) to increase its adoption, especially on popular social media platforms. Furthermore, as most of these tools address only certain regions of the world, it is necessary to extend the outreach efforts of rating sources to a more global scale. Finally, a worrying trend is that the very same journalists and fact-checkers that are debunking false information are incurring in the risks associated with cyberbullying and retaliation, putting in serious danger their lives.[62]

2.5 Digital media information literacy platforms and tools

Digital media literacy tools can play a significant role in the 'digital age' by fighting disinformation through engaging users, with interactive, unique formats. Through modes such as online gaming, fact-checking, newsletters and crowdsourcing, users can think critically about the sources of information encountered online, and thus are able to interpret and identify false information.[63] By providing the necessary tools and correct information services to allow users to independently assess misleading content online, **digital media literacy tools and platforms** can be positive and productive modes of engagement amongst individuals of all levels and backgrounds. As users develop and grasp the skills needed to effectively evaluate and interpret information online, they are able to proactively engage in critical thinking and thus in meaningful

---

60 Examples of existing tools are NewsGuard, Crowdtangle, WeVerify

61 As stated by experts during the UNICRI Teleworkshop: Technology solutions to combat fake news during COVID-19 (22 May 2020).

62 Funke D. (31 August 2018), 'This Italian fact-checker is getting death threats for debunking hoaxes' in *Poynter*. Available at: https://www.poynter.org/fact-checking/2018/this-italian-fact-checker-is-getting-death-threats-for-debunking-hoaxes/

63 Examples of existing tools are: Checkology, Open.online, Bad News, Poynter.

exchanges both in various contexts online and offline, contributing to a powerful and positive sharing of information.

It is necessary to keep in mind that as these technologies constantly evolve and change based on users' needs' and on the target audience, the sustainability of digital media literacy tools and their long-term benefits and impacts will have to be assessed and examined accordingly. As an increasing number of users from various age and demographic groups, including professional, academic backgrounds, interact with different tools and platforms, the efficiency of these tools (i.e. time, resources, scalability) must also be taken in consideration when addressing the needs of a specific audience.[64]

Integrating various modes of instruction and training in different curricula at a global scale could represent a vital and necessary component in overcoming the challenge related to the accessibility of such tools. Addressing this obstacle – including that of geographic location, working to provide multi-language support, etc. – empowers users at a large scale to gain the needed analytical tools in order to properly tackle disinfor-

mation. Lastly, a challenge presented when considering the role of fact-checking at the internet scale, is in the attempt to build infrastructure amongst fact-checkers and various platforms, so as to increase the collaboration and visibility with these groups.[65] The importance of involving diverse actors and modes of instruction in this specific technology option is necessary so as to increase the verification of sources and further equip users to productively interpret online information, and thus effectively contribute to fighting disinformation. This aspect becomes ever more necessary especially when considering the current situation surrounding the coronavirus disease pandemic. Mobilizing various international stakeholders in further promoting media literacy tools and practices, could be seen as an effective strategy to countering the disinformation following COVID-19.[66] Some initiatives can turn into global counter-narratives to contain the "infodemic." In this regard, the UN has launched in June 2020 "Verified," a global project to share accurate information and, at the same time, "stories from the best of humanity".[67]

23

---

64 Based on UNICRI evaluation criteria for technology options.

65 Poynter Presentation, Baybars Örsek during the second UNICRI virtual experts' meetings (22 May 2020).

66 Organizations such as UNESCO has joined forces with the members of the UNESCO-led Global Alliance for Partnership on Media and Information Literacy (GAPMIL) to launch the MIL Alliance Response to COVID-19.

67 "There has never been a greater need for accurate, verified information – About page." Available at: https://www.shareverified.com/en/about

Technology Option	Objective	Advantages
Data science/Big Data visualization	Identifying the spread of large-scale disinformation	*Effectiveness*: visualizing disinformation layers, showing interconnected narratives, detecting sentiment and human bot-accounts, taking timely decisions based on visual insights.
Artificial Intelligence tools and platforms	AI systems using algorithms to guide the public in detecting mis/disinformation	*Efficiency* (funds & time): scalable solution to detect in the shortest time possible deep fakes and thousands of relevant and accurate articles.
Mobile apps and chatbots powered by fact-checkers	Through AI systems and fact-checking, aims to help users differentiate between facts and falsehoods	*Connectedness*: available on smartphones and popular messaging services (i.e. Whatsapp, Viber, FB Messenger).
Web-browser extension for the general public	Monitoring reliability of online news sites when browsing and scrolling social media news-feeds	*Effectiveness*: pre-debunking fake news from hundreds of websites. *Ethics, privacy & transparency*: accountable ratings by trained journalists.
Digital media information literacy platforms and tools	Fighting fake news by helping users in identifying and avoiding false information online	*Effectiveness*: unique formats (i.e, online gaming, fact-checking, newsletters, crowdsourcing) to assess and provide accurate information.

24

Challenges	Example of existing tools/org
Sustainability: challenges with possible data quality issues (i.e., similar texts vs text generators). *Accuracy* needs to constantly be assessed (i.e., "arms race"). *Technology* is still under development.	• BotSlayer • AltoAnalytics • F. Bruno Kessler • CogSec Collab • IFCN
Accuracy/Privacy & Transparency issues: reasoning behind each detection may lack of human judgment.	• Mindzilla • The Factual • BubbleNets • Emerj
Effectiveness: some chatbots are not powered by AI. *Technology issues:* UX and lack of tailored comments call for better response customisation. Funding *sustainability*.	• Logically • Whatsapp Chatbot by Poynter's IFCN and WHO • Viber Chatbot • Carina
Coherence for consistency of methods used. *Interoperability and UX* on social media platforms. Funding *sustainability*. Risks of *cyberbullying/retaliation* towards journalists.	• WeVerify • CrowdTangle • NewsGuard
Sustainability: long-term benefits and impacts to be assessed. *Efficiency (time, resources, scalability)* of tools in addressing needs of a specific audience. Fact-checking at internet scale.	• Poynter • Checkology • EuvsDisinfo Project • Open.online • Bad News (Game) • Science Feedback

25

26

Part 3

Conclusions and the way forward

There are two main conclusions that we can draw from this study. The first is that violent non-state actors, including terrorist, violent extremist and organized criminal groups, have been maliciously using social media during COVID-19. There are three types of groups who have been particularly active. The first two types, the right-wing extremist groups and the groups associated with ISIL and Al-Qaida, have tried to use the pandemic to reinforce their narratives (either racist, anti-Semitic, Islamophobic and antiimmigrant or against democracy and modernisation). They have also attempted to misuse social media to incite potential terrorists to intentionally spread COVID-19 and to use it as an improvised form of a biological weapon for example by coughing in crowded places.[68] The third type, being the organized criminal groups, has been trying to take advantage of the pandemic mainly to portray a positive image of their organizations to expand their activities and penetrate the legal economy.

It is very likely that the corrosive action of these violent non-state actors will not be limited to the present crisis, but will continue during its aftermath, as terrorist, violent extremist and criminal groups will seek to influence post-COVID-19 policies for their own benefit.

Moreover, there is a dangerous convergence of different conspiracy theories that continue to spread around the world. These theories put together different and often contradictory stories such as the identification of the 5G mobile phone signal as a vehicle to transmit the virus, or the false claim that the pandemic has been masterminded by Bill Gates to implant microchips into human beings, or the false idea that the virus is a hoax and does not exist.

This convergence of conspiracy theories can have a multiplier effect, reinforcing intergroup polarization and isolation and, as a result, reducing the observation of government coronavirus guidance on COVID-19 (like the refusal of masks, lockdown and fu-

---

68 See also *Eleventh report of the Secretary-General on the threat posed by ISIL (Da'esh) to international peace and security and the range of United Nations efforts in support of Member States in countering the threat*, p. 2. Available at: https://undocs.org/S/2020/774

ture vaccines) and potentially increasing the risks of terrorist attacks against people or against infrastructures that are falsely connected to COVID-19.[69]

The second conclusion is that technology can play an important role in assisting UN Member States in countering these complex threats. It is however important to stress the following points:

➕ **Technology alone cannot solve the problem**: Technology can assist but not replace human judgment in countering these complex threats. Technology increases speed and scalability to analyse data, identify trends and spot suspicious news, but the human ability to evaluate the veracity of a text remain (for the moment) unchallenged. Therefore technology, including AI, can help humans to improve performance, but not replace them.

➕ **Negative effect of technology solutions**: Technology-based countermeasures offer advantages, but at the same time, can bring negative effects, including the risks that technological tools infringe freedom of expression and privacy, or that they contain hidden biases reflecting social inequities.

➕ **Use technology to engage users**: Users are the strength of the Internet. Some violent non-state actors are particularly skilled in creating social bubbles to promote antisocial behavior and, at the same time, create distrust towards traditional media. An effective use of technology to detect and debunk fake news should aim to engage and empower users so that they can understand the problem and make their own informed decision about what is true and what is not.

---

69 Goodman J. and Carmichael F. (27 June 2020), 'Coronavirus: 5G and microchip conspiracies around the world' in *BBC News*. Available at: https://www.bbc.com/news/53191523. See also CTED (July 2020), Trends Alert, cited, p. 3; Emberland T. (24 February 2020), *Why conspiracy theories can act as radicalization multipliers of far-right ideals*, Center for Research on Extremism. Available at: https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2020/conspiracy-theories-radicalization-multipliers.html University of Oxford (22 May 2020), *Conspiracy beliefs reduce the following of government coronavirus guidance*. Available at: https://www.ox.ac.uk/news/2020-05-22-conspiracy-beliefs-reduces-following-government-coronavirus-guidance

The report has illustrated five different options: Data science/Big Data visualization to identify the spread of large-scale disinformation, Artificial Intelligence tools and platforms to detect fake news online, Mobile apps and chatbots powered by fact-checkers targeting the general public, Web-browser extension for the general public, and Digital media information literacy platforms and tools. Each option presents some advantages and limitations as illustrated by the table at the end of part 2.

What is the way forward?

UNICRI will use its Knowledge Center "*Security through Research, Technology and Innovation (SIRIO)*" in Geneva, Switzerland, to address the problem in cooperation with the UN Member States, international organizations, the private sector, civil society, the scientific and academic community, and other relevant stakeholders.

UNICRI will be committed to performing the following activities:

1. **Keep monitoring the malicious use of social media by violent non-state actors and evaluate new potential threats**, focusing in particular on:

   ✳ Strategies and tactics to maliciously use social media during the pandemic, considering also that some violent non-state actors are showing capacities to quickly incorporate modern technologies to put into practice their extremist visions;

   ✳ Attempts to use social media to inspire or organize terrorist attacks to intentionally transmit COVID-19 and other dangerous pathogens;

   ✳ Attempts to use social media to jeopardize the credibility of governments and public health authorities in relation to the response to the virus, in-

cluding social media attacks against the rapid deployment of vaccines;

   ✳ Attempts to use social media to expand criminal activities;

   ✳ Consequences and potential damage that violent non-state actors can cause not only from a security point of view but also from a social, economic and health perspective (including vaccine refusals).

To facilitate this task, UNICRI will prepare risk scenarios based on a systematic analysis of key drivers/risk factors and information from existing case studies. The risk scenarios will describe plausible future events with the intention to stimulate understanding and discussion with national and international experts (see Annex I). The risk scenarios will help weigh out potential threats, assess vulnerabilities and identify effective policy choices.

2. **Keep mapping technology options to combat online disinformation and misinformation** with the aim of understanding:

   ✳ How technology options fit or can be adapted to combat non-state actors' tactics and strategies to maliciously use social media;

   ✳ What are the strengths of the technology options;

   ✳ What are the possible limitations or risks related to the use of technology options, including the risk of infringing freedom of expression and violating privacy.

3. **Training onto monitoring malicious use of social media by violent non-state actors**, including:

   ✳ Training for security services, law enforcement and prosecutorial authorities.

4. **Raising awareness on malicious use of social media and technology options**, including:

   ✸ Social media campaign;

   ✸ Webinars to inform Member States;

   ✸ Briefings at the United Nations (Palais des Nations) in Geneva.

"

What is the way forward?

Annex: Sample of risk scenarios

This annex offers three different risk scenarios involving the malicious use of social media by a right-wing extremist group, a group associated with ISIL and Al-Qaida, and an organized criminal group. The scenarios describe plausible future events that are intended to stimulate understanding and discussion.

Although they are fictional, the three risk scenarios are based on real events (in line with the first part of the report) and attempt to explain how complex and different risks may impact on Member States during the pandemic.

These case studies do not represent a complete picture of the way violent non-state actors can maliciously misuse social media during a pandemic, but they represent samples to support innovative thinking about plausible future events.

RISK SCENARIO 1

THE CASE OF A RIGHT-WING EXTREMIST GROUP

A right-wing extremist group decides to take advantages of the COVID-19 pandemic to recruit new members and increase social tension. They take the following steps:

1. **Link COVID-19 to their extremist narrative**: the right-wing group creates a series of hashtags that attribute the pandemic to the Jewish communities and migrants. The hashtags promote fake news such as:

   a. the government wants to control the population with their anti-COVID-19 measures,

   b. the immigrants are "stealing" the resources that could help the white inhabitants,

   c. some Jewish scientists have created the virus in a lab and soon will try to sell a vaccine which has been already produced.

2. **Expansion on social media**: The group starts spreading fake news related to COVID-19 on social media, providing "examples" or "evidence" of their conspiracy theory. They use the video of a businessman, who talks about the importance to find a vaccine, as "evidence" that COVID-19 is a hoax and that business compa-

nies are trying to make profits by selling a fake cure. A series of memes start circulating across several social media channels embracing the same conspiracy theory: the Jewish community have created the coronavirus in order to gain power at the expense of others.

3. **Create a state of panic**: The group tries to contribute to the creation of a state of panic, combining real information with fake news and conspiracy theories. Dozens of groups in an instant messaging application create a newsfeed, in which they match scientifically accurate data with misleading and anti-Semitic comments. The materials are used to keep engaging the existing affiliates, and, at the same time, to radicalize new individuals.

4. **Inspire terrorism**: The group starts asking their followers to attack the Jewish community and groups of immigrants by intentionally spreading the virus or attacking their facilities. New memes, propaganda posters, hashtags with direct anti-Semitic and anti-immigrant messages start spreading across several social media platforms, instigating attacks against minorities (such as "clean the race" with "mother nature's terrorist").

5. **Terrorist attacks**: Self-radicalized individuals, who have contracted COVID-19, spit on food in Middle Eastern markets while other individuals with COVID-19 cough next to persons inside a synagogue. The police are immediately alerted and identify the perpetrators as extremists who, before the attacks, have posted anti-Semitic messages on social media.

6. **Consequence**: The two attacks attract international media attention. As a result, social tension increases. Right-wing political movements use the opportunity to attack on social media the national government on its migration strategy, alleging that the open borders and the current migration laws have caused the spread of the virus.

31

RISK SCENARIO 2

THE CASE OF AN EXTREMIST GROUP ASSOCIATED WITH ISIL AND AL-QAIDA

A group associated with ISIL and Al-Qaida decides to take advantage of the COVID-19 pandemic to spread the virus against their "enemies". They take the following steps:

1. **Link COVID-19 to their extremist narrative**: The group uses their official media channels to arbitrarily connect the origin of the virus to the mistreatment of a religious group in certain countries. They define COVID-19 as a "divine punishment" for the unbelievers and invite new followers to join the violent extremist group and spread the virus against their enemies.

2. **Build a positive image**: As the number of persons with COVID-19 increases, the group starts sharing on social media hygiene measures to protect the population. The group begins a campaign, teaching communities how to correctly wash their hands and how to implement social distancing correctly. They upload pictures of the campaign online and, at the same time, start sharing images and videos that accuse "unbelievers" of being responsible for the spreading of the virus.

3. **Inspire terrorism**: The group asks their followers to take advantage of the situation to plan attacks against their enemies. The group starts distributing videos and infographics that explain how the virus can be intentionally transmitted to other persons. They also encourage their followers to target vulnerable facilities, such as hospitals, through coordinated attacks and declare that those who sacrifice themselves will be considered as martyrs.

4. **Terrorist attacks**: Some of their followers successfully spread the virus in public spaces, including food markets.

5. **Aftermath**: The group uploads videos of the attacks to their media platforms declaring that they were a "divine punishment".

RISK SCENARIO 3

THE CASE OF AN ORGANIZED CRIMINAL GROUP

An organized criminal group wants to take advantage of the pandemic to expand their activities. They take the following steps:

1. **Imposing curfew during the pandemic**: The group is concerned that an increase of COVID-19 cases in "their" territory could attract public authorities and, as a result, jeopardise criminal activities. They decide to use social media platforms and instant messaging applications to alert the local community that a curfew has been enforced and that any person who will not respect it will be shot. The leader of the group also records an audio in which he threatens the inhabitants and warns them not to break the curfew, while audio and pictures of the members of the criminal group holding posters with the same information are distributed through instant messaging platforms. The message and pictures reach the national and international media.

2. **Reinforce the control of "their" territory**: The members of the group start patrolling the streets and, at the same time, they deliver 400 boxes with groceries and hand sanitizer. All boxes are labelled with a large sticker with the face of the leader of the criminal group. All members are wearing their uniform so that they can be easily identified as part of the group.

3. **Creation of consensus**: The local, national, and some international media cover the story, sharing pictures and interviews of the inhabitants of the communities. Part of the local community and several social media show appreciation. Some of them post messages that assert that the criminal group is more effective than the government in supporting the local population during COVID-19.

4. **Territorial expansion**: The criminal group takes advantage of the situation to expands their activities to other territories. The group distributes groceries in other cities outside their territory. The local government tries to intervene, but the criminal group uses social media to misrepresent the situation, showing videos in which they verbally confront the local police and accuse them of interfering with the "humanitarian" operation. The videos are distributed throughout social media accounts, and after a few hours they circulate in the national and international news.

33

5. **Expansion of activities:** The group expands illegal activities (such as illicit trafficking of drugs and racketeering) in the new territories. The group also offers financial support to local entrepreneurs who are facing the risk of bankruptcy as a result of the pandemic.

6. **Penetrate the legal economy**: The criminal group imposes "loan shark" at extremely high interest rates to the entrepreneurs who asked for their financial support. Ultimately, the criminal group forces the entrepreneurs to sell their companies whose properties are transferred to strawmen to conceal the true ownership.