unicri
United Nations
Interregional Crime and Justice
Research Institute

Handbook to combat CBRN disinformation

Handbook to combat CBRN disinformation

Disclaimer

The opinions, findings, conclusions and recommendations expressed herein are those of the authors and do not necessarily reflect the views and positions of the United Nations and the United Nations Interregional Crime and Justice Research Institute (UNICRI), or any other national, regional or international entity involved. Contents of the publication may be quoted or reproduced, provided that the source of information is acknowledged.

This publication was produced with the financial support of the Federal Bureau of Investigation (FBI) of the United States of America. Its contents do not necessarily reflect the views of the FBI.

Acknowledgements

This Handbook has been prepared by Mr. Francesco Marelli, Head of the Chemical, Biological, Radiological and Nuclear Risk Mitigation and Security Governance Unit at UNICRI and Ms. Mariana Diaz Garcia, Fellow at UNICRI. UNICRI would like to express its appreciation to Ms. Katy Carroll, Mr. Luca Cislaghi Lanza and Ms. Marta Zeloni for their research on disinformation.

Copyright

For more information:

UNICRI, Viale Maestri del Lavoro, 10, 10127 Torino – Italy
Tel: + 39 011-6537 111 / Fax: + 39 011-6313 368.
Website: www.unicri.org
Email: unicri.publicinfo@un.org

Foreword

Chemical, biological, radiological, and nuclear (CBRN) disinformation can be very damaging, with potentially dire consequences. This applies particularly to the present age when information is easily accessible and sharable with little consideration for its veracity or repercussions.

Manipulated and fraudulent information about a CBRN emergency, such as a terrorist attack or a pandemic, can mislead governments and international organizations, compromise response measures, result in misdirected or wasted resources, and cause alarm among populations. False information, including conspiracy theories, can also amplify fear and anxiety in targeted populations, for example, if it conveys the message that a CBRN event is out of control. It can even generate social chaos if part of the population wrongly believes that a CBRN event has been deliberately and consciously orchestrated as part of an evil conspiracy. False information can also be used to radicalize and recruit unwitting victims into terrorist activities as it can amplify fear and stir hatred among various populations.

This *Handbook to combat CBRN disinformation* describes different techniques to detect, analyse and debunk disinformation that is intentionally misleading. Using specific examples, the *Handbook* provides readers with the tools they need to avoid falling prey to disinformation about CBRN events.

In line with its mandate to advance justice and the rule of law in support of peace and sustainable development, UNICRI has been tackling the threat of CBRN disinformation in various ways.

Since 2020, UNICRI has been monitoring the malicious use of social media. In November 2020, UNICRI published the report *Stop the virus of disinformation*, which describes how malicious actors took advantage of the COVID-19 pandemic to jeopardize the efficacy and credibility of response measures by governments. UNICRI has also analysed existing technological options to detect and debunk false information (*e.g.*, Big Data, Artificial Intelligence tools and platforms, mobile apps and chatbots, etc.,) to understand the advantages and possible challenges of each technological option in the short and long term.

In 2021, UNICRI, in alliance with the World Health Organization (WHO) and in cooperation with the United States Federal Bureau of Investigation (FBI), began to develop strategies to increase awareness of CBRN disinformation and to deliver trainings to Member States. This *Handbook to combat CBRN disinformation* is the result of this action-oriented approach. Designed for individuals and agencies working in CBRN risk mitigation at different levels, which have been or could potentially be exposed to and targeted by disinformation, the *Handbook* equips practitioners with the competencies to effectively analyse, understand and respond to CBRN disinformation in the media and on social media platforms.

I hope this *Handbook* will enhance knowledge and understanding of this complex problem and encourage daily use of effective measures to prevent and combat the malicious use of social media. Efforts to curb CBRN disinformation and misinformation, as well as their undesirable consequences, contribute to a safer world for all of us.

Antonia Marie De Meo

Director
UNICRI

Table of Contents

1

Introduction

Chemical, biological, radiological, and nuclear (CBRN) disinformation is intentionally misleading and deceptive information about CBRN threats, that can potentially cause serious political, financial, and physical harm to governments, international organizations, the scientific community, academia, industry, and the population at large.[1]

CBRN disinformation in the media and on social media has become a significant problem in the last few years. Three main factors have contributed to its amplification. The first is the **particular nature of CBRN threats**, involving dangerous and sometimes invisible substances that make them a particularly suitable topic to generate fear and anxiety through social networks and the media. An example is the wave of disinformation related to the coronavirus disease (COVID-19) pandemic that proliferated immediately after the Wuhan Municipal

[1] Misinformation is information that is false but not created to cause harm. At the same time, disinformation is information that is false and deliberately created to harm a person, social group, organization, or country. See UNESCO (2018). *Journalism, 'fake news' and disinformation: A handbook for journalism education and training*. Available on the Internet. See also Fallis, D. (2015). What Is Disinformation? *Library Trends*, 63, 401 - 426.

Health Commission in China reported a cluster of cases of pneumonia on 31 December 2019.[2] The wave of disinformation began with groundless and harmful conspiracy theories about the origin of the virus (e.g., COVID-19 was manufactured as a bioweapon by governments and the pharmaceutical industry, or COVID-19 was a hoax that was used to enslave citizens and impose a dictatorship) and soon expanded to the transmission patterns of the virus, the treatments, prophylactics, and cures, the existence of the virus itself, and finally, the effectiveness and consequences of interventions and decisions made by authorities and institutions.[3]

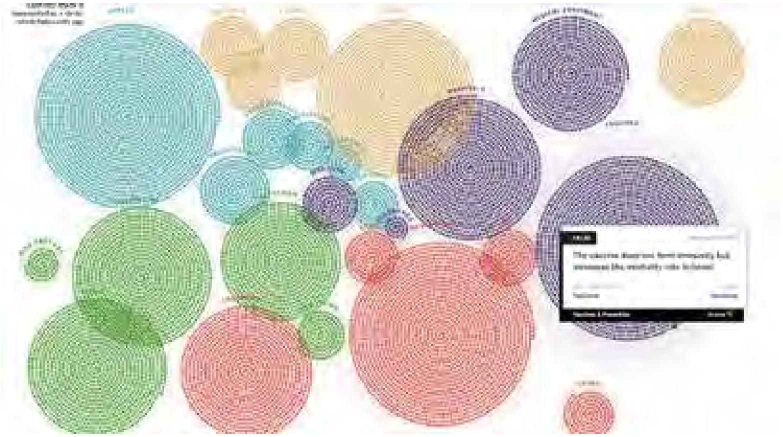


Figure 1: The graphic shows different thematic areas where false information has been detected by the #CoronaVirusFacts Alliance.
Source: Poynter (2022). Fighting the Infodemic: The #CoronaVirusFacts Alliance. Poynter. Available on the Internet

[2] WHO defines as ‘infodemic’ the overabundance of accurate and inaccurate information that takes place during an epidemic, and that spreads via digital and physical information systems, making it difficult to find trustworthy and reliable sources. See WHO (2020). *An ad hoc WHO technical consultation managing the COVID-19 infodemic: call for action*. Available on the Internet.

[3] For more information see UNICRI (2020). *Stop the virus of disinformation: the risk of malicious use of social media during COVID-19 and the technology options to fight it*. Available on the Internet.

The second factor is that **new digital platforms** have created new forms of communication and greater connection between millions of users. The global exchange of content in real-time on social media has modified citizens' behaviours regarding news consumption. Easy and rapid access to a great wealth of data and information has permitted citizens to expand their knowledge and introduced innovative journalistic practices. However, as so often with technological advancement, social media platforms have also posed new challenges. One of them is the proliferation of false information and conspiracy theories. Although misinformation and disinformation are nothing new in the history of humankind, social media platforms have been used by malicious actors who deliberately fabricate and disseminate false information to harm individuals, social groups, organizations, or countries.[4]

The third factor is the role of violent **non-state actors,** especially on social media. While in the past CBRN disinformation was often part of covert operations conducted by governments with the intention to influence the opinions and actions of individuals and Member States (disinformation campaigns and disinformation mitigation tactics), in recent years, terrorists, violent extremists, and organized criminal groups have shown the ability to exploit vulnerabilities in the social media ecosystem to deliberately disseminate conspiracy theories and manipulate people in relation to CBRN threats. Violent non-state actors can sometimes operate as voluntary or involuntary proxies of governments, but their direct involvement has introduced a new variable that significantly amplifies the spread of disinformation.

[4] Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation: A new ICFJ learning module. *International Center for Journalists*. Available on the Internet.

Today, disinformation related to CBRN threats has the potential to cause serious political, financial, and physical harm to governments, international organizations, the scientific community, academia, industry, and the population at large. More individuals and organizations than ever have been targeted by CBRN disinformation. Viral online and sometimes physical attacks have been conducted against almost every stakeholder operating in the area of CBRN risk mitigation, including policymakers, managers of institutions related to CBRN, researchers from universities and research centres, spokespersons of different government departments, in particular the public health sector, journalists and representatives of international organizations. In this respect, CBRN disinformation has become a serious new challenge.

Combating disinformation is not an easy task. A successful strategy requires a combination of different actions from governments, educational institutions, the media industry, and technology companies. This includes the monitoring of disinformation on social media, the debunking of false information and conspiracy theories, investment in technological tools to identify false news, education about media literacy, the adoption of adequate legal instruments without violating basic human rights and freedoms, and the training of law enforcement and prosecutorial agencies to investigate and prosecute crimes connected to disinformation.[5]

Different methods have been studied and developed to combat disinformation, including techniques to anticipate disinformation (pre-bunking), to detect and analyse disinformation, and to effectively respond to disinformation and demonstrate the falseness of information or a conspiracy theory (debunking).

[5] For more information about courses and guides to improve media literacy skills, see: training resources from First Draft and the resources shared by the International Center for Journalists (ICFJ). Available on the Internet.

This Handbook focuses mainly on debunking techniques. It has been designed for individuals or agencies working in CBRN risk mitigation at different levels (communication, decision-making, managerial, operational, technical, etc.) who have been or could potentially be exposed to and targeted by disinformation. The Handbook addresses the problem in two ways: the first is to understand the problem of CBRN disinformation on social media; the second is to develop a set of competencies to effectively prevent and respond to disinformation on media and social media platforms with a specific focus on techniques for debunking false information.[6]

The Handbook is divided into two sections. The first part offers an analysis of the problem by describing the strategic objectives of CBRN disinformation and the techniques used to manipulate audiences on social media. The second part identifies and describes techniques to effectively demonstrate the falseness of an idea, story, or theory. It also provides practical advice on how to analyse a situation in which an individual or an organization has been exposed to disinformation and how to decide whether to respond to a false allegation.

The Handbook provides several practical examples of techniques to disinform the audience or debunk false information. To produce the document, UNICRI has monitored several social media platforms, paying specific attention to the role of violent non-state actors. UNICRI has focused on three main non-state actors. The first is represented by violent extremists, particularly right-wing extremist groups – also referred to as the far-right. These groups do not represent a coherent or easily defined movement, but as stated by the United Nations Counter-Terrorism Committee Executive Directorate, they are rather a "shifting, complex and overlapping

[6] This is not a guide for professional journalists and fact checkers whose work is to ensure that a source of news or information maintains credibility and integrity. UN has produced guidelines for these categories such as the UNESCO handbook with Fondation Hirondelle entitled *Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training*.

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".[7] The malicious use of social media by right-wing extremists is not recent, since some of these groups have attempted to use the Internet to promote campaigns of hatred since the 1990s.[8] With the beginning of the COVID-19 outbreak, right-wing extremist groups further expanded their online presence and resorted to xenophobic, Islamophobic, or anti-Semitic narratives to disseminate conspiracy theories about the origin of COVID-19 and its possible cures.[9]

The second group of non-state actors is represented by terrorist organizations, particularly those associated with the Islamic State of Iraq and the Levant (ISIL, also known as Da'esh) and Al-Qaida.[10] A rich literature has analysed the online presence of these groups.[11] For example, the ISIL global network of supporters (whom they term *munasireen*) have been using Telegram for operational activities such as content

[7] United Nations Counter-Terrorism Committee Executive Directorate (CTED) (2020). *Trends Alert "Member States concerned by the growing and increasingly transnational threat of extreme right-wing terrorism"*, p. 2.

[8] Conway, M., Scrivens, R. & Macnair, L. (2019). Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends. *ICCT*. Available on the Internet.

[9] UNICRI (2020). *Stop the virus of disinformation: the risk of malicious use of social media during COVID-19 and the technology options to fight it*. Available on the Internet. See also Diaz Garcia, M. (2021). Infodemic: Right-wing extremist groups and the risk of disinformation during the COVID-19 pandemic. *Freedom from Fear (F3) magazine*. Available on the Internet.

[10] The Security Council Committee pursuant to resolutions 1267 (1999), 1989 (2011) and 2253 (2015) concerning ISIL (Da'esh), Al Qaida and associated individuals, groups, undertakings and entities (called Security Council ISIL (Da'esh) & Al-Qaida Sanctions Committee) regularly updates 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 (available on the Internet). For the purpose of this report, the list has been used as the main source to identify the organizations that are part of this second group of violent non-state actors.

[11] Clifford, B., & Powell, H. (2019). Encrypted Extremism Inside the English-Speaking Islamic State Ecosystem on Telegram. *Program on Extremism of the George Washington University*. Available on the Internet. Waters, G., & Postings, R. (2018). Spiders of the Caliphate: Mapping the Islamic State's Global Support Network on Facebook. *The Counter Extremism Project*. Available on the Internet. Kruglova, A. (2022). Terrorist Recruitment, Propaganda and Branding Selling Terror Online. *Routledge*.

hosting, audience development, secure communication and financing.[12]

Organized criminal groups represent the third group.[13] The malicious use of social media by criminal groups, especially to promote their criminal activities in the areas where they operate and to intimidate and discourage rival groups, is not new. However, following the COVID-19 pandemic some of them have been very active in manipulating their audience and promoting a positive image of themselves as reliable "institutions".

[12] Clifford, B., & Powell, H. (2019). Encrypted Extremism Inside the English-Speaking Islamic State Ecosystem on Telegram. *Program on Extremism of the George Washington University*. Available on the Internet.

[13] Organized crime can be defined as a "continuing criminal enterprise that rationally works to profit from illicit activities that are often in great public demand. Its continuing existence is maintained through corruption of public officials and the use of intimidation, threats, or force to protect its operations". See UNODC (2018). Defining organized crime. *UNODC*. Available on the Internet.

Virus spread
tests pos
deaths
state of
for covid
economic
dead

2

CBRN Disinformation

This chapter analyses the strategic objectives of CBRN disinformation and illustrates some of the main effective disinformation techniques.

2.1 What are the strategic objectives of CBRN disinformation?

CBRN disinformation can be used to achieve three different strategic objectives:

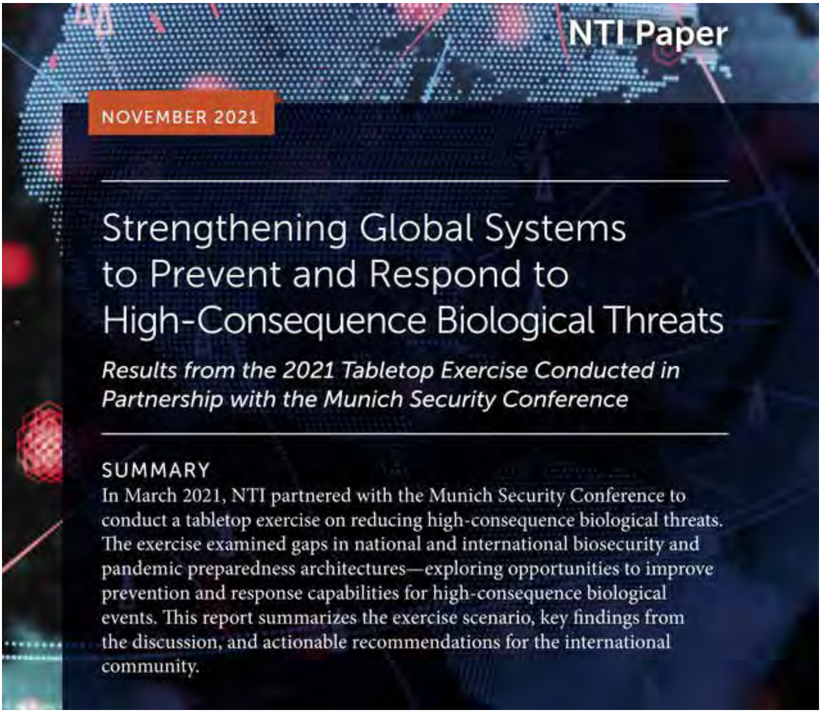
1. jeopardize the trust and credibility of institutions operating in CBRN risk mitigation,
2. incite fear, hatred and violence through the dissemination of radical narratives, and
3. obtain financial benefits.

2.1.1 Jeopardize the trust and credibility of institutions operating in the area of CBRN risk mitigation

A first strategic objective of CBRN disinformation on social media is to jeopardize the trust and credibility of institutions operating in the area of CBRN risk mitigation such as governmental and intergovernmental organizations, research centres, non-profit organizations, pharmaceutical companies, etc. A possible way to achieve this objective is to fabricate and spread false stories that accuse these institutions of lying about the origin of a CBRN emergency. For example, a document from the Nuclear Threat Initiative (NTI) summarizing the results of a 2021 tabletop exercise was shared on social media in May 2022 as "evidence" that the monkeypox outbreak was planned (see figure 2 and 3).

Tabletop exercises with fictional scenarios are commonly used to test the pandemic preparedness of governmental and intergovernmental organizations. Interestingly, the NTI tabletop exercise was real and envisaged a terrorist attack with an unusual strain of monkeypox, but this does not prove that the 2022 disease outbreak was predicted and deliberate. As explained by NTI, "the fact that several countries are currently experiencing an outbreak of monkeypox is purely a coincidence".[14]

[14] Reuters Fact Check (2022, 24 May). Fact Check-No evidence that 2021 Nuclear Threat Initiative exercise proves monkeypox outbreak was planned. *Reuters*. Available on the Internet.



NTI Paper

NOVEMBER 2021

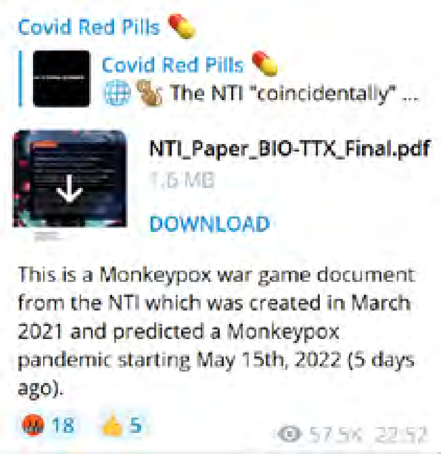
Strengthening Global Systems to Prevent and Respond to High-Consequence Biological Threats

Results from the 2021 Tabletop Exercise Conducted in Partnership with the Munich Security Conference

SUMMARY

In March 2021, NTI partnered with the Munich Security Conference to conduct a tabletop exercise on reducing high-consequence biological threats. The exercise examined gaps in national and international biosecurity and pandemic preparedness architectures—exploring opportunities to improve prevention and response capabilities for high-consequence biological events. This report summarizes the exercise scenario, key findings from the discussion, and actionable recommendations for the international community.

Figure 2: Cover of an NTI paper regarding a tabletop exercise that was falsely used as "evidence" that the monkeypox outbreak was planned.
Source: For the original document see: Yassif, J. M., O'Prey, K.P., Isaac, C. R. (November 2021), Strengthening Global Systems to Prevent and Respond to High-Consequence Biological Threats, NTI Paper, available on the Internet.



Covid Red Pills 💊

Covid Red Pills 💊
🌐 🐒 The NTI "coincidentally" ...

NTI_Paper_BIO-TTX_Final.pdf
1.6 MB
DOWNLOAD

This is a Monkeypox war game document from the NTI which was created in March 2021 and predicted a Monkeypox pandemic starting May 15th, 2022 (5 days ago).

Figure 3: Example of social media post that suggests that the monkeypox outbreak was planned.
Source: Telegram, channel "Covid Red Pills", posted on 20 May 2022.

Online disinformation can also attempt to **jeopardize the public health response and interfere with international cooperation** during a CBRN emergency. For example, figure 4, posted on a social media channel in September 2021, shows an attempt to maliciously mislead the population about the COVID-19 vaccination campaign. In the image, a doctor with a gun pointed at her head is telling a patient that "vaccines are safe and effective". The image insinuates that the doctor is forced by pharmaceutical companies (Pharma) to lie to the patient as part of an international conspiracy called "Agenda 21". Agenda 21 is a plan of action to meet the challenges of environment and development, which was adopted by 178 governments at the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro, Brazil on 13 June 1992. However, in the narrative of far-right extremists, Agenda 21 is a secret plan to reduce the world population to under 500 million, with COVID-19 vaccines being instrumental in reaching this objective.

In reality, there is no connection between Agenda 21 and COVID-19 vaccines. Agenda 21 recommended the development of improved vaccines for the prevention of diseases and, clearly, does not make any reference to COVID-19 vaccines since it was written in 1992. Yet the intention of the image was to maliciously cast doubts on the efficacy of the COVID-19 vaccination by falsely linking it to an alleged United Nations conspiracy to control the world population.

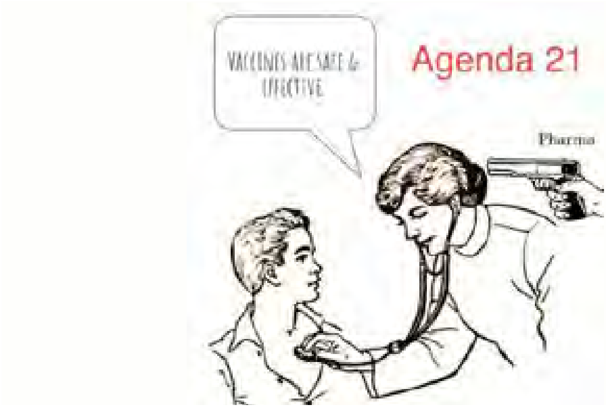


Figure 4: Example of false allegation against Agenda 21 and the United Nations.
Source: Telegram, channel "COVID-19 Agenda", posted on 15 September 2021.

Organized criminal groups can also mislead the population during a CBRN emergency by claiming that they are in "charge" of the emergency, rather than the government. For example, during the COVID-19 pandemic, drug cartels in Mexico shared images and videos on social media platforms in which armed group members distributed boxes and bags containing food, medicine, cleaning products, and toys (see figure 5). In some cases, the boxes were labelled with the logo or image of the cartel (see figure 6). Some criminal groups even 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.[15]

The criminal groups' intention was to minimize the role of the government, while projecting a positive image of themselves as viable replacements for healthcare institutions and responsible political actors during the pandemic. In this way, organized criminal groups could use the CBRN emergency to build the false image that they are a "state within the state", with the power to enforce territorial control through social approval. Regrettably, the main goal of these criminal groups is not to protect the local population during a CBRN emergency, but rather to protect their criminal business and social base in the territories. Their concern is that a large health crisis could trigger the arrival and active involvement of law enforcement agencies or the army in the areas where the criminals operate and, as a result, jeopardize their illegal activities.

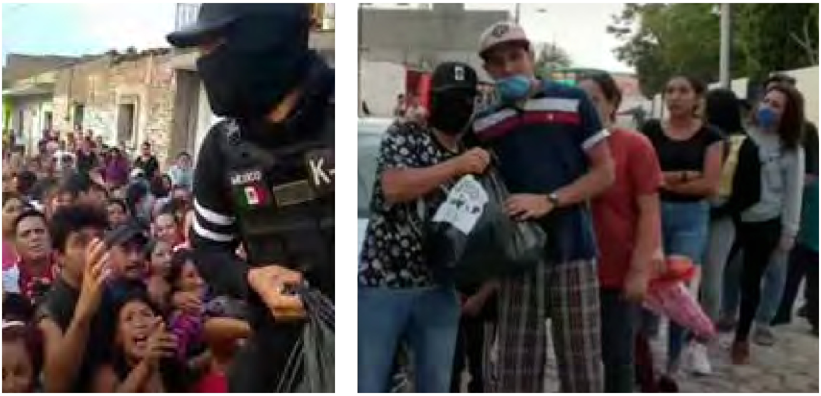


Figure 5: Examples of videos shared on TikTok by Cártel de Jalisco Nueva Generación (CJNG) while they were distributing bags with groceries (*despensas*) in Jalisco, Mexico (Infobae, 2020).
Source: Infobae (2020, 10 May). El narco en TikTok: el CJNG desafía al gobierno y alardea entregando despensas. *Infobae*. Available on the Internet.

[15] Dudley, S. & McDermott, J. (2020). GameChangers 2020: how organized crime survived the pandemic. *InSight Crime*. Available on the Internet.

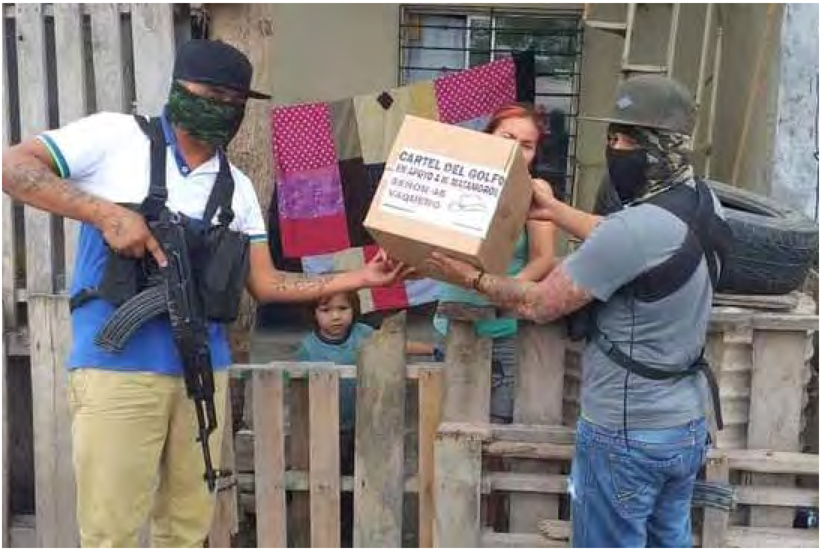


Figure 6: Cártel del Golfo distributing packages of groceries during the COVID-19 pandemic.
Source: Infobae (2020, 20 April). Narcos aprovechan coronavirus en México para repartir despensas y pelear territorio. *Infobae*. Available on the Internet.

CBRN disinformation can also take the form of a disease hoax. For example, in May 2005 the government of New Zealand responded to a letter claiming to have deliberately released the foot-and-mouth disease (FMD) virus on Waiheke Island. Although the government considered the letter a hoax, all necessary steps to safeguard New Zealand's interests and public welfare were taken, including quarantining the island and 14 days of surveillance of susceptible animals.[16] In such a case, disinformation can have a high financial cost by forcing a government to mobilize resources and activate disease emergency responses.

[16] Mackereth, G. F. & Stone, M.A.B. (2006). Veterinary intelligence in response to a foot-and-mouth disease hoax on Waiheke Island, New Zealand. *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006.*

2.1.2 Radicalization, recruitment and incitement to hatred and violence

Another strategic objective of CBRN disinformation is to amplify the fear and anxiety of the population and stir up violence, especially during a CBRN emergency. Figures 7 and 8 show images posted in far-right channels to incite hatred and violence against the vaccines and their producers during the COVID-19 pandemic.

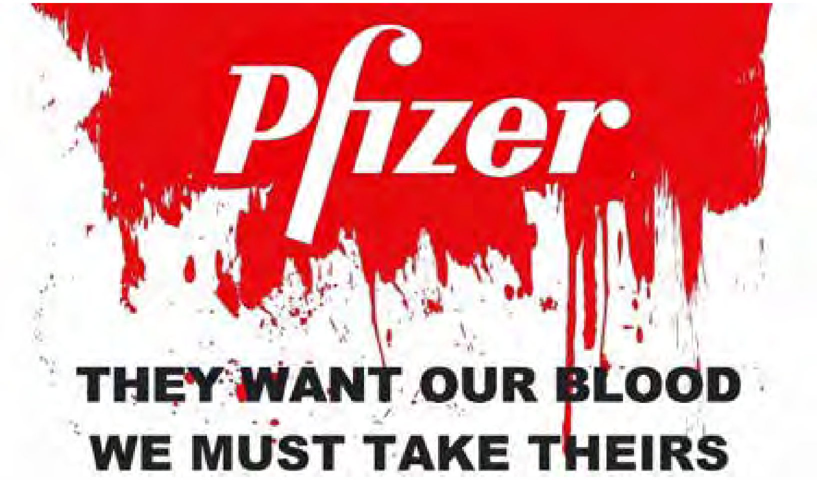


Figure 7: Example of a post that promotes violence.   against the authorities and the pharmaceutical companies that produce vaccines.
Source: Gab, channel “Nazi Society”, posted on 26 November 2021.

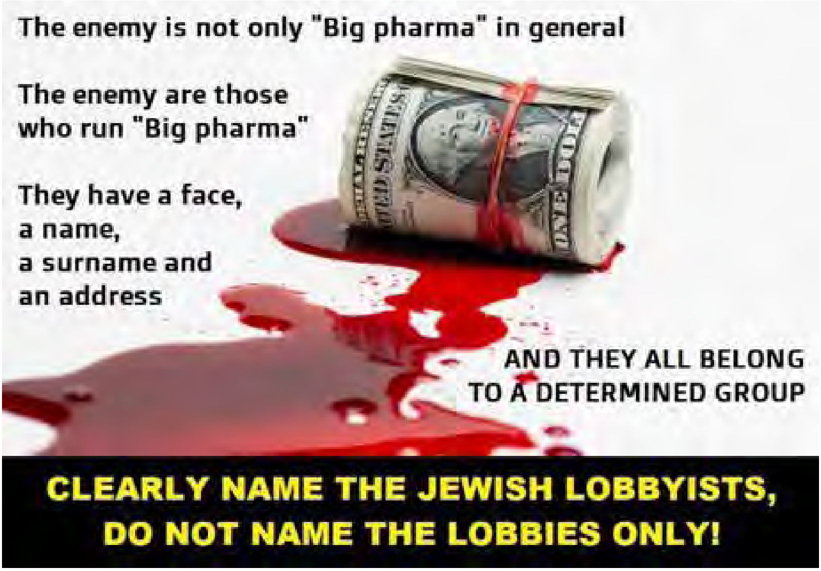

The enemy is not only "Big pharma" in general
The enemy are those who run "Big pharma"
They have a face, a name, a surname and an address
AND THEY ALL BELONG TO A DETERMINED GROUP
CLEARLY NAME THE JEWISH LOBBYISTS, DO NOT NAME THE LOBBIES ONLY!

Figure 8: Example of a post that promotes violence against the pharmaceutical companies that produce vaccines.
Source: Gab, channel "Nazi Society", posted on 4 October 2021.

In March 2020, the Federal Bureau of Investigation (FBI) reported that members of extremist groups were encouraging one another to spread COVID-19, if contracted, through bodily fluids and personal interactions (see figures 9 and 10).[17] This message was shared in multiple online channels and was adapted to different audiences.[18]

[17] Margolin, J. (2020). White supremacists encouraging their members to spread coronavirus to cops, Jews, FBI says. *ABC News*. Available on the Internet.

[18] UNICRI (2020). *Stop the virus of disinformation: the risk of malicious use of social media during COVID-19 and the technology options to fight it.* Available on the Internet.

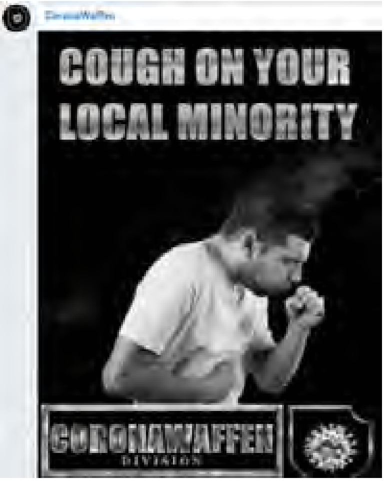
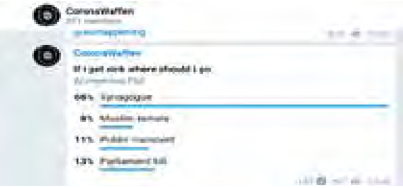


Figure 9: Example of a post that promotes the intentional spread of COVID-19 in a far-right channel.
Source: Telegram, channel "CoronaWaffen", posted in May 2020.



CoronaWaffen

CoronaWaffen

If i get sick where should i go

Anonymous Poll

68% Synagogue

8% Muslim temple

11% Public transport

13% Parliament hill

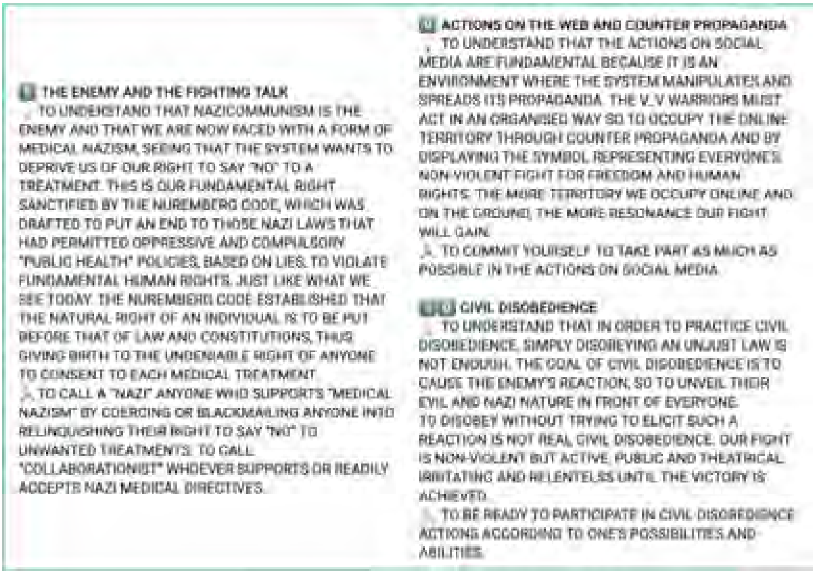
Figure 10: Example of a post that promotes the intentional spread of COVID-19 in a far-right channel.
Source: Telegram, channel "CoronaWaffen", posted in May 2020.

Amplifying fear and hatred during a CBRN emergency can also help to radicalize and recruit new members. For example, groups associated with ISIL/Da'esh and Al-Qaida have spread conspiracy theories that assert that the virus is a "soldier of Allah" that is punishing unbelievers and enemies of Islam. ISIL and Al-Qaida claimed that the virus is God's wrath upon the West.[19] Similarly, Al-Shabaab declared that the coronavirus disease was spread by "the crusader forces who have invaded the country and the disbelieving countries that support them".[20] Overall, by suggesting the coronavirus is a form of divine intervention, these extremist groups hope to radicalize and recruit new members who may perceive the coronavirus as 'proof' that their actions are sanctioned by Allah.

[19] Meek, J. G. (2020, 2 April). Terrorist groups spin COVID-19 as God's 'smallest soldier' attacking West. *ABCNews*. Available on the Internet.

[20] BBC News (2020, 1 March). Coronavirus: Fighting al-Shabab propaganda in Somalia. *BBC*. Available on the Internet.

Figure 11 shows some of the governing rules of the Viral Vendetta (V_V) movement. The V_V movement originated in Italy and France during the COVID-19 pandemic. It became very active in spreading conspiracy theories on social media, claiming that vaccines and other COVID-19 safety measures were a new form of "medical Nazism". The governing rules in the figure show that the movement tried to engage and eventually recruit followers through indoctrination, mainly by inculcating dogmatic ideas that should not be questioned and by asserting facts that should not be critically examined.[21]



THE ENEMY AND THE FIGHTING TALK

TO UNDERSTAND THAT NAZICOMMUNISM IS THE ENEMY AND THAT WE ARE NOW FACED WITH A FORM OF MEDICAL NAZISM, SEEING THAT THE SYSTEM WANTS TO DEPRIVE US OF OUR RIGHT TO SAY "NO" TO A TREATMENT. THIS IS OUR FUNDAMENTAL RIGHT SANCTIFIED BY THE NUREMBERG CODE, WHICH WAS DRAFTED TO PUT AN END TO THOSE NAZI LAWS THAT HAD PERMITTED OPPRESSIVE AND COMPULSORY "PUBLIC HEALTH" POLICIES, BASED ON LIES, TO VIOLATE FUNDAMENTAL HUMAN RIGHTS, JUST LIKE WHAT WE SEE TODAY. THE NUREMBERG CODE ESTABLISHED THAT THE NATURAL RIGHT OF AN INDIVIDUAL IS TO BE PUT BEFORE THAT OF LAW AND CONSTITUTIONS, THUS GIVING BIRTH TO THE UNDENIABLE RIGHT OF ANYONE TO CONSENT TO EACH MEDICAL TREATMENT.

TO CALL A "NAZI" ANYONE WHO SUPPORTS "MEDICAL NAZISM" BY COERCING OR BLACKMAILING ANYONE INTO RELINQUISHING THEIR RIGHT TO SAY "NO" TO UNWANTED TREATMENTS. TO CALL "COLLABORATIONIST" WHOEVER SUPPORTS OR READILY ACCEPTS NAZI MEDICAL DIRECTIVES.

ACTIONS ON THE WEB AND COUNTER PROPAGANDA

TO UNDERSTAND THAT THE ACTIONS ON SOCIAL MEDIA ARE FUNDAMENTAL BECAUSE IT IS AN ENVIRONMENT WHERE THE SYSTEM MANIPULATES AND SPREADS ITS PROPAGANDA. THE V_V WARRIORS MUST ACT IN AN ORGANISED WAY SO TO OCCUPY THE ONLINE TERRITORY THROUGH COUNTER PROPAGANDA AND BY DISPLAYING THE SYMBOL REPRESENTING EVERYONE'S NON-VIOLENT FIGHT FOR FREEDOM AND HUMAN RIGHTS. THE MORE TERRITORY WE OCCUPY ONLINE AND ON THE GROUND, THE MORE RESONANCE OUR FIGHT WILL GAIN.

TO COMMIT YOURSELF TO TAKE PART AS MUCH AS POSSIBLE IN THE ACTIONS ON SOCIAL MEDIA.

CIVIL DISOBEDIENCE

TO UNDERSTAND THAT IN ORDER TO PRACTICE CIVIL DISOBEDIENCE, SIMPLY DISOBEYING AN UNJUST LAW IS NOT ENOUGH. THE GOAL OF CIVIL DISOBEDIENCE IS TO CAUSE THE ENEMY'S REACTION, SO TO UNVEIL THEIR EVIL AND NAZI NATURE IN FRONT OF EVERYONE. TO DISOBEY WITHOUT TRYING TO ELICIT SUCH A REACTION IS NOT REAL CIVIL DISOBEDIENCE. OUR FIGHT IS NON-VIOLENT BUT ACTIVE, PUBLIC AND THEATRICAL, IRRITATING AND RELENTELSS UNTIL THE VICTORY IS ACHIEVED.

TO BE READY TO PARTICIPATE IN CIVIL DISOBEDIENCE ACTIONS ACCORDING TO ONE'S POSSIBILITIES AND ABILITIES.

Figure 11: An example of three of the V_V movement's 12 governing rules, which have been translated into different languages and shared across Telegram.
Source: Graphika (2021). *Viral Vendetta. Inside the conspiratorial movement waging a cross-platform 'psychological warfare' campaign against Covid-19 vaccine advocates*. Available on the Internet.

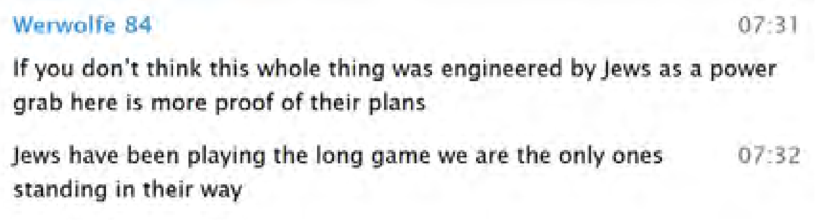
[21] As part of their online activities, the V_V movement organized online attacks against journalists, health workers, and public officials and orchestrated down-voting of social media posts that were advocating COVID-19 health measures. For more information, see: Graphika (2021). *Viral Vendetta. Inside the conspiratorial movement waging a cross-platform 'psychological warfare' campaign against Covid-19 vaccine advocates*. Available on the Internet.

Hate speeches and images on social media platforms can be precursors of real violence. During the COVID-19 pandemic, several episodes of violence were inspired by hateful rhetoric. The list of violent acts includes attacks against hospitals treating patients, vaccination centres, hospital supply chains, infrastructure considered as means of transmission of the virus (e.g., 5G cell towers) and individuals who were regarded as strong advocates of the restriction measures, all in relation to the COVID-19 pandemic.

For example, in March 2020, FBI agents shot and killed a domestic terror suspect in Missouri. The suspect allegedly planned to carry out an improvised explosive device (IED) attack against a hospital treating patients with COVID-19. According to court documents, he was distressed by the government's response to the COVID-19 crisis and motivated by racial, religious, and anti-government sentiment, which was fuelled by his association with two white supremacist groups – the National Socialist Movement (NSM) and the Vorherrschaft Division (VSD) – through the Telegram application (see figure 12). He was listed as one of the administrators of a Telegram chat where he argued that the government was using the COVID-19 pandemic as an "excuse to destroy our people".[22] He also claimed that he wanted to exploit the ongoing pandemic and the increased media coverage of the health sector as it provided unique opportunities.[23]

[22] Martin, N. R. (2020). Heartland terror. *The Informant*. Available on the Internet.

[23] Levine, M. (2020). FBI learned of coronavirus-inspired bomb plotter through radicalized US Army soldier. *ABC News*. Available on the Internet.



Werwolfe 84 07:31

If you don't think this whole thing was engineered by Jews as a power grab here is more proof of their plans

Jews have been playing the long game we are the only ones standing in their way 07:32

Figure 12: This is the last of the messages that the extremist posted on Telegram regarding the COVID-19 pandemic.

Source: Martin, N. R. (2020). Heartland terror. *The Informant*. Available on the Internet.

Hate speech and disinformation were also the precursors of extremists' attempts to infiltrate protests related to COVID-19 measures, with the aim to recruit new members and instigate violence. For example, in Italy, far-right extremists infiltrated a demonstration organized in the capital, Rome, on 9 October 2021, against vaccines and the Green Pass system that required citizens to provide digital proof of COVID-19 vaccination or immunity. On that occasion, a group of demonstrators, led by members of the neo-fascist organization Forza Nuova, broke away from the demonstration, clashed with police and burst into the national headquarters of the main Italian trade union organization, causing significant damage.[24]

[24] Al Jazeera (2021, 9 October). Clashes break out in Rome amid anger over COVID 'green pass'. *Al Jazeera*. Available on the Internet.

2.1.3 Financial gain

CBRN disinformation can also aim to generate profits. A very simple way is to use the Internet to promote and sell counterfeited or substandard products (e.g., personal protective equipment (PPE), medicines, and vaccines) during a CBRN emergency.

Figure 13 shows the example of a website selling fraudulent protective equipment during the COVID-19 pandemic. The website, managed by members of ISIL, was seized in the United States in August 2020. The website was selling N95 respirator masks by falsely claiming that they were approved by the United States Food and Drug Administration (FDA). It also claimed to have near unlimited supplies of the masks, even though the item was designated as scarce by the official authorities.[25]

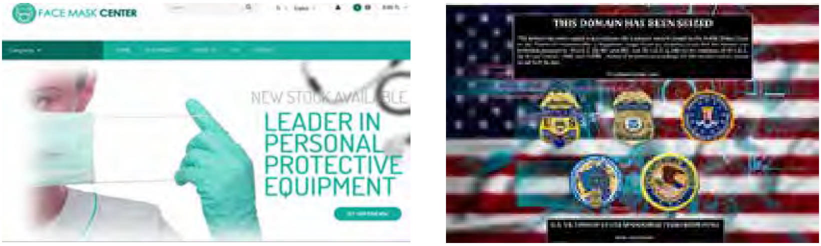
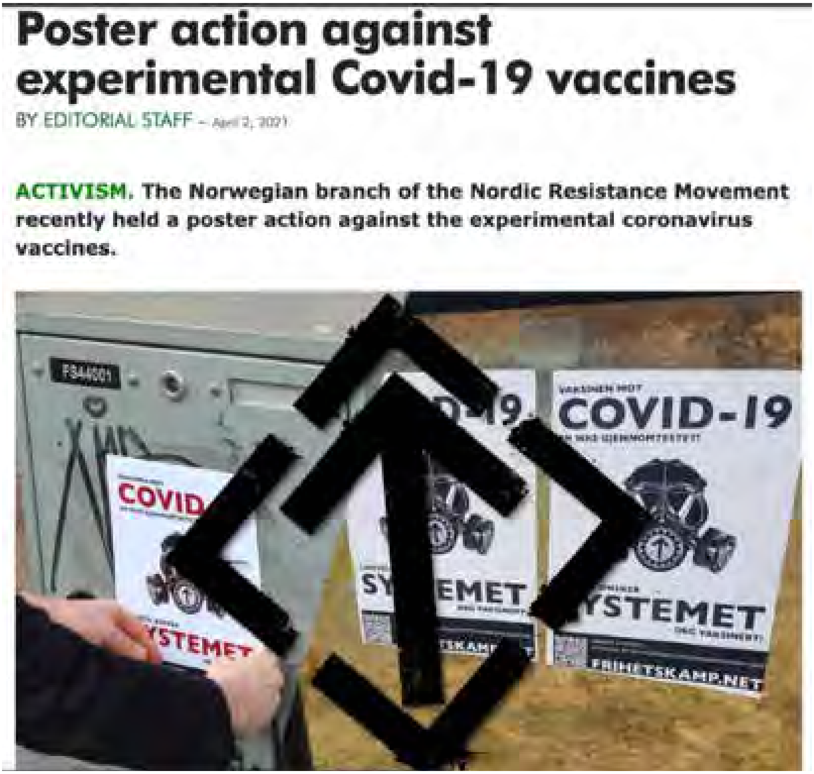

FACE MASK CENTER
NEW STOCK AVAILABLE
LEADER IN PERSONAL PROTECTIVE EQUIPMENT
THIS DOMAIN HAS BEEN SEIZED

Figure 13: Screenshot of the website used to sell fraudulent protective equipment to finance ISIL terrorist actions.
Source: The United States Department of Justice (2020). Global Disruption of Three Terror Finance Cyber-Enabled Campaigns. United States Department of Justice. Available on the Internet.

[25] The United States Department of Justice (2020). Global Disruption of Three Terror Finance Cyber-Enabled Campaigns. *United States Department of Justice*. Available on the Internet.

Moreover, a CBRN emergency can represent an opportunity to raise funds. For example, the Nordic Resistance Movement, a violent neo-Nazi organization, asked for donations to support their action against COVID-19 vaccines (see figures 14 and 15).[26] The group proposed different payment methods on their website, including a wide variety of cryptocurrencies.



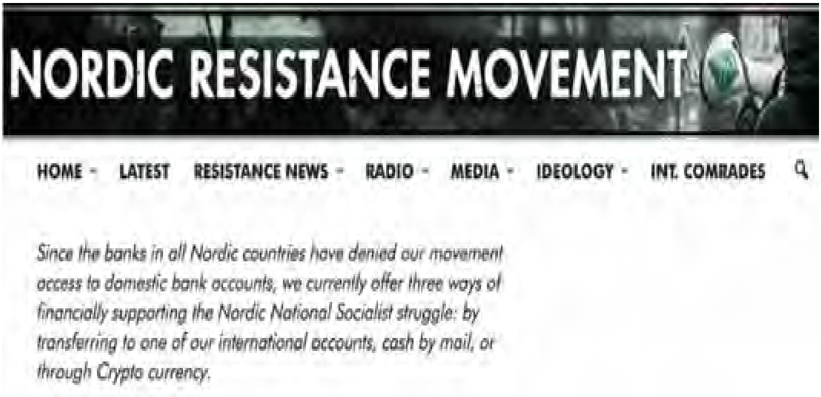
Poster action against experimental Covid-19 vaccines

BY EDITORIAL STAFF – April 2, 2021

ACTIVISM. The Norwegian branch of the Nordic Resistance Movement recently held a poster action against the experimental coronavirus vaccines.

Figure 14: Members of a far-right extremist movement promote their activities online.
Source: Website, The Nordic Resistance, posted in April 2021.

[26] Far-right groups based in the United States have successfully profited from the use of online platforms, particularly the streaming platform DLive, cryptocurrencies, and other fundraising methods, where they were able to raise up to $1.5 million dollars in a year. See Stone, P. (2021). US far-right extremists making millions via social media and cryptocurrency. *The Guardian*. Available on the Internet.



Since the banks in all Nordic countries have denied our movement access to domestic bank accounts, we currently offer three ways of financially supporting the Nordic National Socialist struggle: by transferring to one of our international accounts, cash by mail, or through Crypto currency.

Figure 15: Online donation options to financially support the far-right extremist group.
Source: Website, The Nordic Resistance, posted in April 2021.

Disinformation can also create "incentives" for criminal smuggling of CBRN materials. This is the case of "red mercury", a substance allegedly used to create nuclear weapons. As demonstrated by the forensic analysis of the sample seized by the police, red mercury is a hoax. However, some criminal groups have fuelled the myth that red mercury is a component of nuclear weapons since the early 1990s and tried to sell it as an exclusive product on the black market.[27]

[27] Chivers, C.J. (2015, 22 November). The Doomsday Scam. *The New York Times*. Available on the Internet.

2.2 Disinformation techniques and successful practices on social media and messaging applications

Different techniques can be used in social media and messaging applications to disseminate CBRN disinformation. This section provides an overview of them.

2.2.1 Content manipulation

In general terms, manipulated content on social media refers to genuine content that has been digitally manipulated using photo or video editing software.[28] Manipulated content can be created with relatively simple and accessible software. Figure 16 shows a poster, allegedly issued by the government of the United Kingdom, that encourages people to apply for compensation if they were not "made fully aware of the health risks from the COVID-19 vaccines". The poster was a fabrication that was circulated on the main social media platforms without the government's permission.[29] The original photo, of an elderly woman sitting with her hand on her forehead while a younger woman consoles her, was taken from a stock photography provider as shown in figure 17.[30]

Interestingly, there was some kernel of truth in the poster since the government of the United Kingdom had created a mechanism for one-off payments in the event that someone

[28] UNESCO (2018). Journalism, 'fake news' and disinformation: A handbook for journalism education and training in *UNESCO*. Available on the Internet.

[29] Reuters (2022, 20 May). Fact Check-Poster about UK vaccine compensation scheme is not from UK. *Reuters*. Available on the Internet.

[30] The image was most likely taken from the webpage of iStock available on the Internet at https://www.istockphoto.com.

were to become "severely disabled as a result of a vaccination against certain diseases".[31] However this mechanism, called the Vaccine Damage Payment, was not specifically designed for COVID-19 vaccines. The official government website also did not state that citizens were entitled to compensation "if they weren't made fully aware of the health risks from the COVID-19 vaccines" (as written in the fake poster).

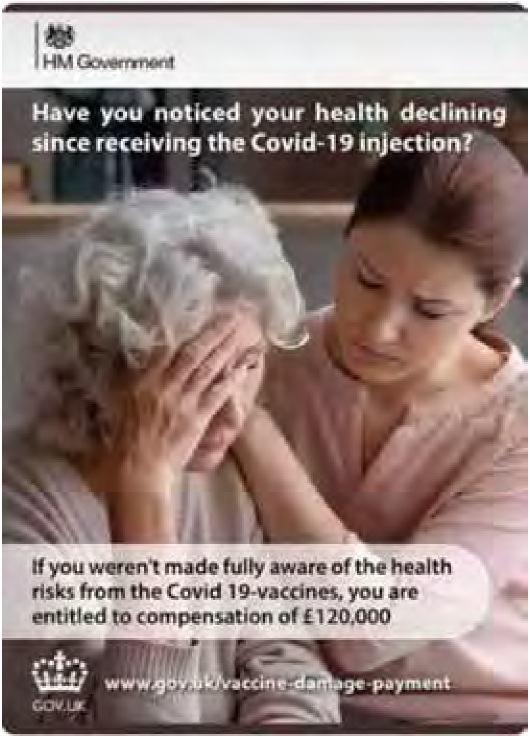

HM Government
Have you noticed your health declining since receiving the Covid-19 injection?
If you weren't made fully aware of the health risks from the Covid 19-vaccines, you are entitled to compensation of £120,000
GOV.UK
www.gov.uk/vaccine-damage-payment

Figure 16: Fake poster with the UK Government logo. The poster was shared on Facebook.
Source: AFP Fact Check (2022, 23 May). UK govt rejects fake Covid vaccine injury poster shared on Facebook. *AFP Fact Check*. Available on the Internet.

[31] Government of the United Kingdom (n.d.). Vaccine Damage Payment in *Government of the United Kingdom*. Available on the Internet.

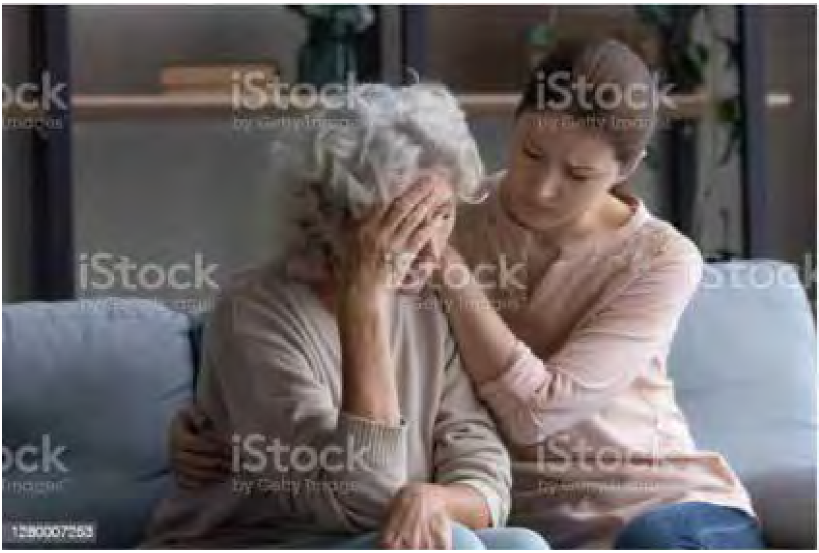


Figure 17: Original photo of the woman with her head in her hand while a younger woman consoles her. It was most likely taken from iStock.

Figure 18 is an example of another fake poster, featuring the logos of the United States Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO). Shared through different far-right online channels, the poster with a clear anti-Semitic, Islamophobic and racist message, instructs members of the group who test positive for COVID-19 to spread the virus around local minority communities.

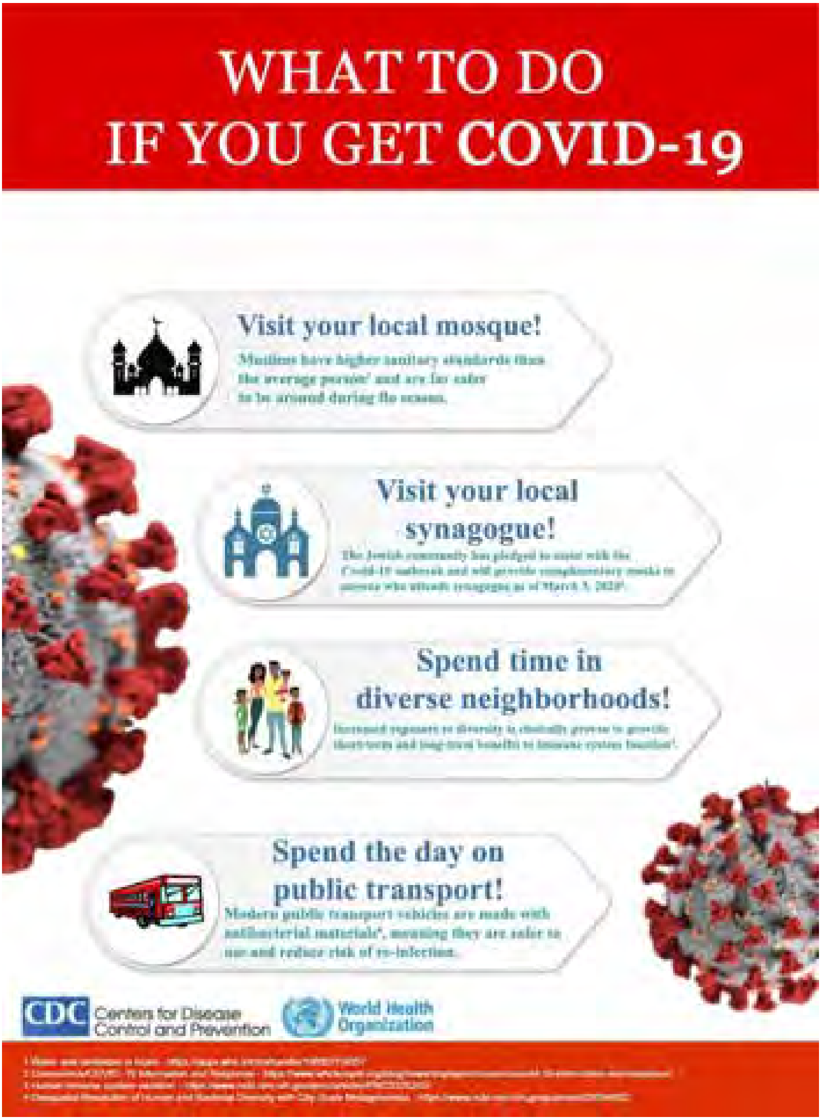

WHAT TO DO
IF YOU GET COVID-19
Visit your local mosque!
Muslims have higher sanitary standards than the average person¹ and are far safer to be around during flu season.
Visit your local synagogue!
The Jewish community has pledged to assist with the Covid-19 outbreak and will provide complimentary masks to anyone who attends synagogue as of March 5, 2020².
Spend time in diverse neighborhoods!
Increased exposure to diversity is clinically proven to provide short-term and long-term benefits to immune system function³.
Spend the day on public transport!
Modern public transport vehicles are made with antibacterial materials⁴, meaning they are safer to use and reduce risk of re-infection.
CDC Centers for Disease Control and Prevention
World Health Organization

Figure 18: False poster of CDC and WHO with disinformation targeting minorities and inciting people to spread COVID-19.
Source: Telegram, The British National Socialist Movement channel, posted in March 2020.

Figure 19 is another example of content manipulation, where two genuine images have been purposefully used in the wrong context and overlaid with misleading text. After an earthquake in Japan on 24 March 2022, a post on Facebook shared two old images. The first was the picture of a large fire at a Japanese oil refinery next to the Fukushima Daiichi Nuclear Plant following the terrible earthquake of 11 March 2011 (figure 20). The second was the picture of a group of firefighters wearing protective gear who intervened after a man injured at least 10 people in a knife and fire attack on a train in Tokyo on 31 October 2021.

With clear intent to mislead the population, the post overlaid the two images with the following text in Korean language: “[Breaking News] Japan’s Fukushima nuclear power plant swept up in a red blaze”.

In this example, the images of two real events (a fire next to a power plant as a result of the earthquake of 11 March 2011, and a group of firefighters gathering after a train attack on 31 October 2021) were not manipulated, but they were maliciously associated with a fake event due to the addition of misleading text (a fire at the Fukushima Daiichi Nuclear Plant on 24 March 2022). In terms of the kernel of truth in this example, a fire alarm did go off at the Fukushima Daiichi Nuclear Plant on 16 March 2022, but, according to the Japanese authorities, no fire or smoke was detected.[32]

[32] Shim, K. (2022, 29 March). Social media users share misleading Fukushima plant claim after Japan earthquake. *AFP Fact Check*. Available on the Internet.

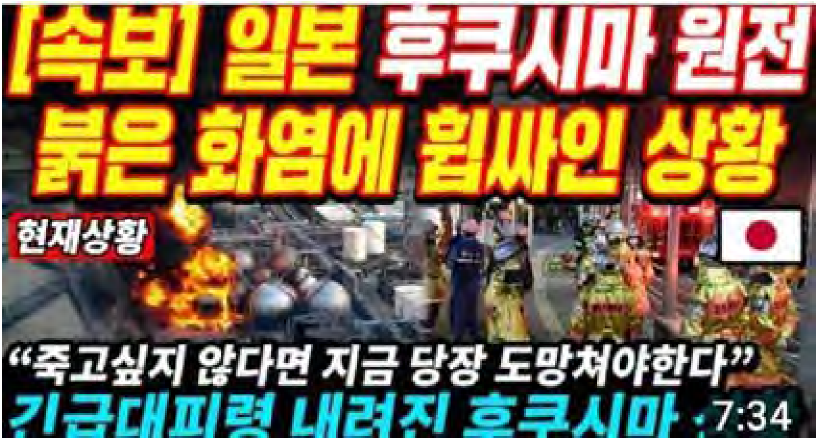

[속보] 일본 후쿠시마 원전
붉은 화염에 휩싸인 상황
현재상황
"죽고싶지 않다면 지금 당장 도망쳐야한다"
긴급대피령 내려진 후쿠시마
7:34

Figure 19: Misleading post falsely claiming that there is a large fire at the Fukushima Daiichi Nuclear Plant.
Source: Facebook, posted on 24 March 2022.

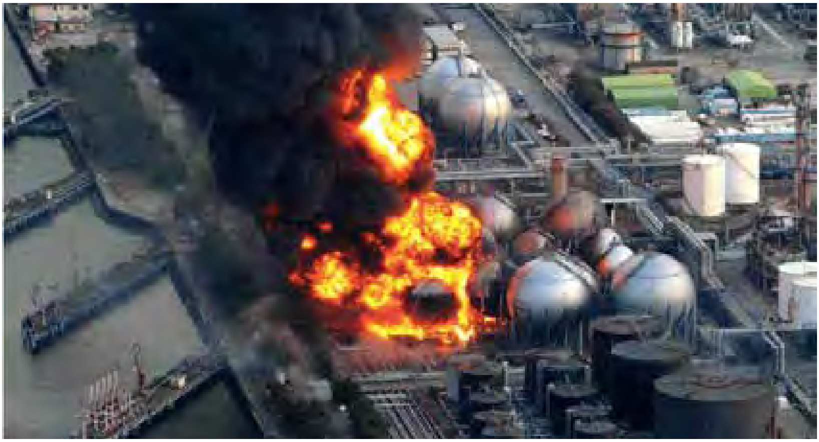
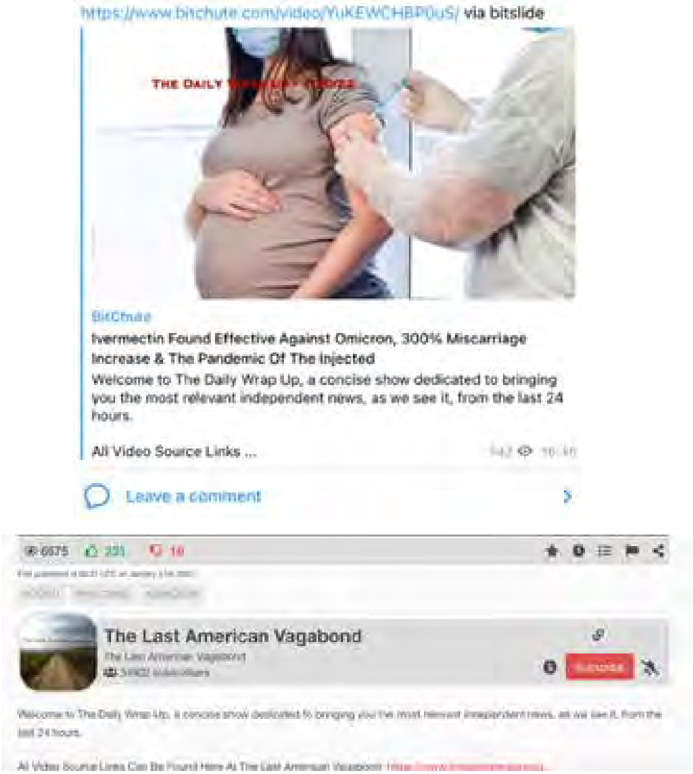


Figure 20: Original picture of a large fire at an oil refinery next to the Fukushima Daiichi Nuclear Plant following the terrible earthquake of 11 March 2011.

Another way to manipulate content is to create a misleading link that takes users to a disinformation website. For example, the post in figure 21 falsely claims that COVID-19 vaccines increase the risk of pregnancy loss. The message refers to a source ("All Video Source Links"), but when the users click on the hyperlink they are redirected to a website (The Last American Vagabond) that spreads conspiracy theories and extremist content.



https://www.bitchute.com/video/YuKEWCHBP0uS/ via bitslide

BitChute

Ivermectin Found Effective Against Omicron, 300% Miscarriage Increase & The Pandemic Of The Injected

Welcome to The Daily Wrap Up, a concise show dedicated to bringing you the most relevant independent news, as we see it, from the last 24 hours.

All Video Source Links ...

The Last American Vagabond

The Last American Vagabond

Welcome to The Daily Wrap Up, a concise show dedicated to bringing you the most relevant independent news, as we see it, from the last 24 hours.

All Video Source Links Can Be Found Here At The Last American Vagabond: [illegible]

Figure 21: The post falsely claims that COVID-19 vaccines increase the risk of pregnancy loss. The message refers to a source (“All Video Source Links”). However, when users click on the hyperlink they are redirected to a website (The Last American Vagabond) that spreads conspiracy theories and content related to other extremist views.
Source: Telegram, channel “Plandemic”, posted on 31 January 2022 and Bitchute, The Last American Vagabond posted on 31 January 2022.

Recent technological advancements have made it possible to use sophisticated techniques for manipulating content such as website defacement and deepfake videos. Website defacement is an attack on a website to change its appearance and content. The attackers (called defacers) break into the web server and replace the hosted website with the one they have created.[33] A deepfake video is the product of an Artificial Intelligence (AI) technique for human image synthesis that combines and superimposes existing images and videos onto source images or videos. These videos or photographs can misrepresent people by generating images that are nearly indistinguishable from the original. If combined with speech synthesis systems (that learn to imitate individuals' voices), deepfake videos can misrepresent people by reproducing not only their voices, but also their cadence and expressions.[34] In this manner, AI techniques can produce fake news reports, including realistic video and audio, to influence public opinion, affect political campaigns and erode trust in government (e.g., in the area of vaccines).[35]

Figure 22 shows an example of deepfake video produced by researchers at the University of California, Berkeley and the University of Southern California as part of a study to create new techniques to detect deepfakes of political leaders.

[33] Ferreira, S., Antunes, M., & Correia, M. E. (2021). Exposing Manipulated Photos and Videos in Digital Forensics Analysis. *Journal of Imaging*, 7(7), 102.

[34] Allen, G. & Chan, T. (2017). Artificial Intelligence and National Security. *Belfer Center Study.* Available on the Internet.

[35] Larson, H. J. (2018, 16 October). The biggest pandemic risk? Viral misinformation. *Nature*. Available on the Internet. See also Gambetta, D. & Hertog, S. (2017). *Engineers of Jihad: The Curious Connection between Violent Extremism and Education*; Kahneman, D. (2011). *Thinking, Fast and Slow*.

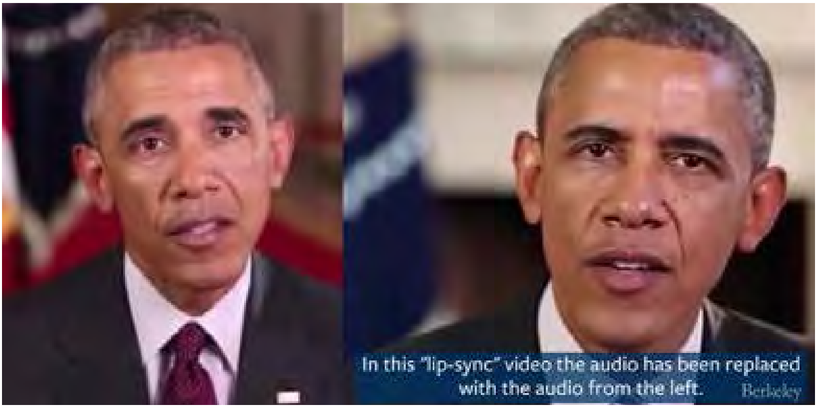

In this "lip-sync" video the audio has been replaced with the audio from the left.
Berkeley

Figure 22: Deepfake video of the former President of the United States Barack Obama produced by researchers at the University of California, Berkeley and the University of Southern California.
Source: Manke, K. (18 June 2019), Researchers use facial quirks to unmask 'deepfakes' in Berkeley News. Available on the Internet.

2.2.2 Mimicking scientific debate

Another disinformation technique on social media is the imitation of scientific debate. Scientific credibility refers to the recognition of science as a source of reliable information, particularly because the information is considered to come from a trustworthy methodology (scientific method)[36] or has undergone a rigorous peer review to be recognized as scientific literature.[37]

Malicious actors can manipulate readers of an online discussion by referring to imagery or debate falsely associated with a reliable source of information. For example, online

[36] "The scientific method is the process of objectively establishing facts through testing and experimentation. The basic process involves making an observation, forming a hypothesis, making a prediction, conducting an experiment and finally analyzing the results." For more information see: Wright, G. & Lavery, T (n.d.). Scientific Method. *TechTarget.* Available on the Internet.

[37] Bocking, S. (2004). *Nature's experts: science, politics, and the environment*. p. 164.

pseudo-scientific debate can quote individuals referred to as “scientists” who are not affiliated with any educational or scientific institutions or who are not recognized as experts in the scientific community. Malicious actors can share false claims online by quoting alleged doctors or experts, when these individuals do not hold any official credentials or in some cases, do not even exist in real life but are invented and mentioned with the purpose of adding false credibility to the disinformation.

Another misleading technique is to mimic scientific discussion by making citations from or references to articles published on the Internet. The fallacy of this technique is that the quoted articles have not been published in peer-reviewed journals and therefore have not undergone a rigorous editorial review process. The same technique can be used in the production of documentaries. Figure 23 shows the poster for the film “Plandemic: Indoctrination” released in August 2020. The film claims that the COVID-19 pandemic is part of a conspiracy orchestrated by organizations such as the United States CDC and Google to control humanity and generate profits. The film purports to be a science documentary, but fact-checkers have discredited most of Plandemic’s claims.[38]

[38] Dunlop, W.G. (2020, 19 August). New ‘Plandemic’ film promotes coronavirus conspiracy theory. *AFP Fact Check*. Available on the Internet.

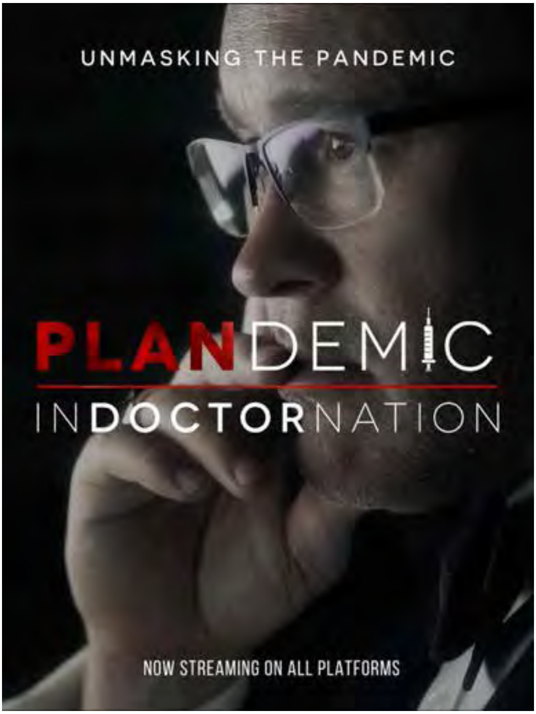


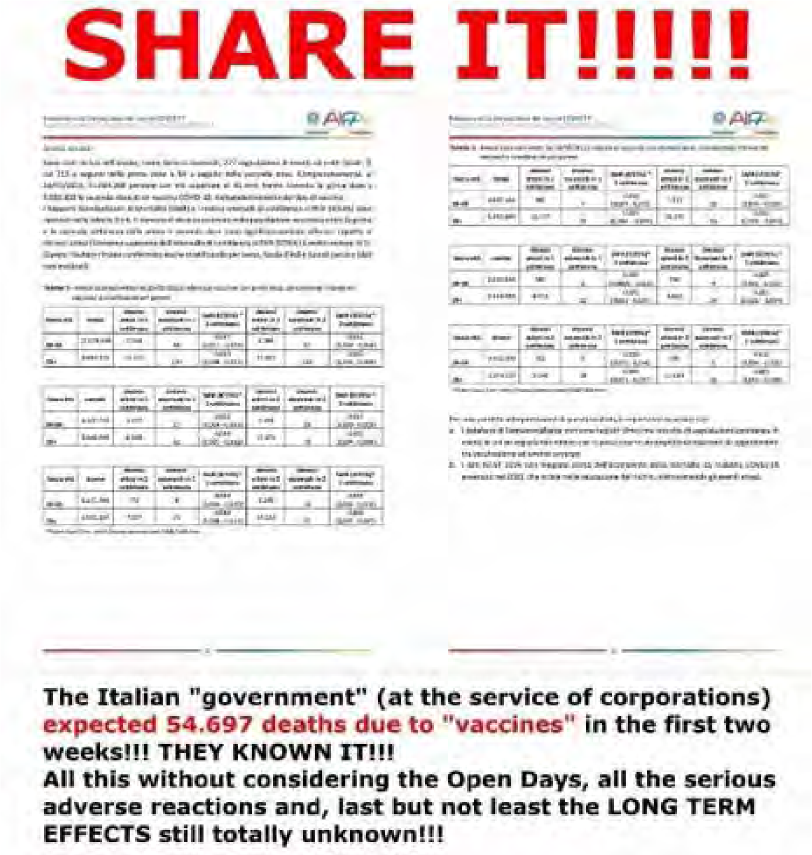
Figure 23: Cover image of the "Plandemic" conspiracy documentary. Far-right extremist groups have referred to this video as scientific evidence.

Misreading of statistics and data manipulation have also been a tactic used to spread disinformation on social media.[39] This is the case of an official document by the Italian Medicines Agency (AIFA) from May 2021 that provided statistics about deaths recorded after the first and the second doses of COVID-19 vaccines in Italy (see figure 24). A social media post by a right-wing extremist group invited all followers to share the statistics from AIFA, falsely claiming that the document

[39] For more information on How to identify false or incorrect statistics see: Otis, C. (2020). *True or False: A CIA Analyst's Guide to Spotting Fake News.*

revealed that the Italian government was expecting 54,697 deaths after the first and second doses.

However, the AIFA document says something completely different. It explains that the number of observed deaths in the 14 days following the first or second administration of COVID-19 vaccines was 277, a number that must be compared with the total expected deaths for Italian citizens in the same period (equivalent to 52,665, a number which was calculated based on statistics of 2019). As clearly explained by the document from AIFA, this type of analysis is important to understand if there is a potential statistical association between the administration of a vaccine and the number of deaths that would have occurred independently of it. When the number of cases observed after the vaccine administration is lower than the expected deaths (as happened in Italy), the association between the two is unlikely to be a coincidence.



SHARE IT!!!!!

The Italian "government" (at the service of corporations) expected 54.697 deaths due to "vaccines" in the first two weeks!!! THEY KNOWN IT!!!
All this without considering the Open Days, all the serious adverse reactions and, last but not least the LONG TERM EFFECTS still totally unknown!!!

Figure 24: Example of misinterpretation of official data related to COVID-19.
Source: Gab, channel "RAM: Right-wing Alt Media", posted on 6 December 2021. For the original document see: Italian Medicines Agency (AIFA), *COVID-19 Vaccine Surveillance Report 5*, period 27/12/2020 - 26/05/2021.

Box 1. How to identify false or incorrect statistics

When identifying false or incorrect statistics the following elements should be considered:

1. The statistics should refer to a source that can be accessed.
2. The information should come from a reliable source (scientific article, government, recognized media outlet, among others).
3. It should be possible to cross-reference correct and/or reliable data with other reliable sources.
4. Graphs should clearly indicate what is being measured.
5. Reliable sources usually clarify how information was collected or selected to create graphs and statistics. This can also help in analysing if the information was taken out of context.

2.2.3 Creation and dissemination of conspiracy theories

The development and spread of conspiracy theories on social media can also be an effective technique to disinform the general public. Conspiracy theories are explanations of complex events or situations without the support of credible evidence.[40] These theories are characterized by the existence of a conspiracy organized by powerful people who have managed to conceal their role and have deliberately and consciously orchestrated the events by following a secret agenda.[41] To be more "credible", conspiracy theories are often based on some elements of truth.

For example, a conspiracy theory was created after the devastating blast at the Port of Beirut on 4 August 2020. The blast was caused by the explosion of a large amount of ammonium nitrate that was stored in a warehouse near the port without proper safety measures. However, several posts that appeared on different social media platforms claimed that it was a nuclear attack. Figure 25 shows an example of a social media post that falsely associated the blast to an attack with a tactical nuclear bomb. The text "right on top of a Rothschilds bank in Beirut, Lebanon" at the bottom of the post maliciously suggests that the Rothschilds, the influential European banking family, were linked to the supposed nuclear attack. According to a false narrative which is very popular through far-

[40] There are diverse types of conspiracy theories, and they are frequently tailored to the beliefs of the group, however, they have some structural consistency and often overlap significantly. The theories can be classified as conspiracies of control (where the world, the nation, the state, the media, or the establishment are under the control of a unitary body of collaborators), plots against the group (attempts by this elite group to destroy the group), specific event theories (interpretation of the events influenced by their belief in conspiracies of control or from a belief in a concerted campaign against the group itself), and complex or supernatural conspiracies. Brown, M. (2020). Fact check: Did Gates Foundation fund and does Pirbright Institute Own Coronavirus Patent? *Southwest Times Record*. Available on the Internet.

[41] Miller, C., & Bartlett, J. (2010). The Power of Unreason: Conspiracy Theories, Extremism and Counter-Terrorism. *Academia*. Available on the Internet.

right groups, the Rothschilds family owns the Central Bank of Lebanon. Since the headquarters of the Central Bank is located close to the Port of Beirut, the conspiracy theory concludes that this is not a coincidence: the explosion of 4 August was a nuclear attack orchestrated by the banking family.[42]

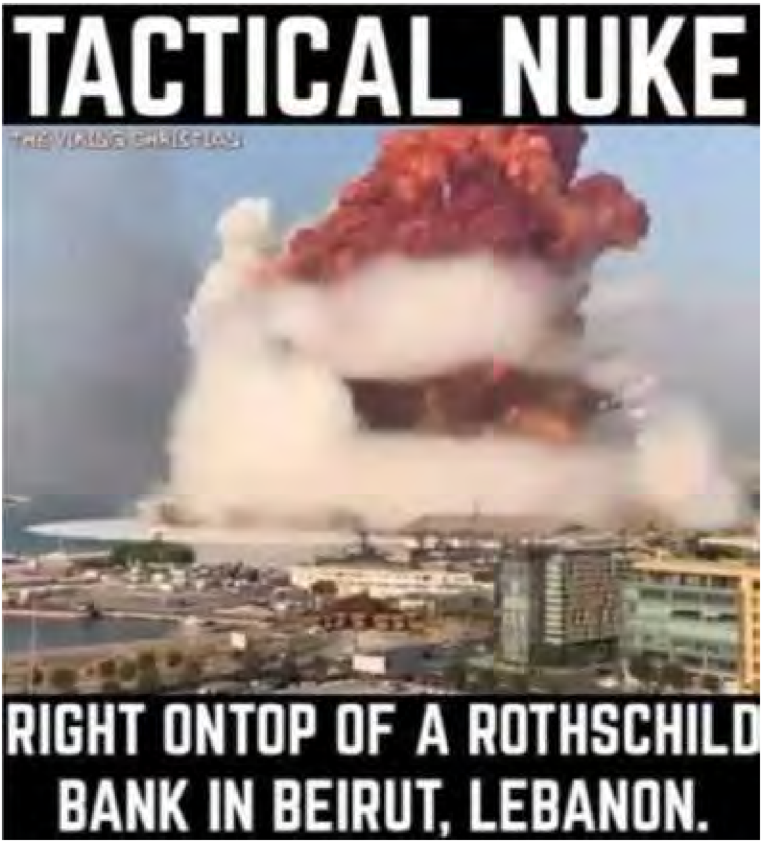


Figure 25: Example of a social media post claiming that the blast at the port of Beirut on 4 August 2020 was a nuclear attack.
Source: Facebook, captured on 5 August 2020.

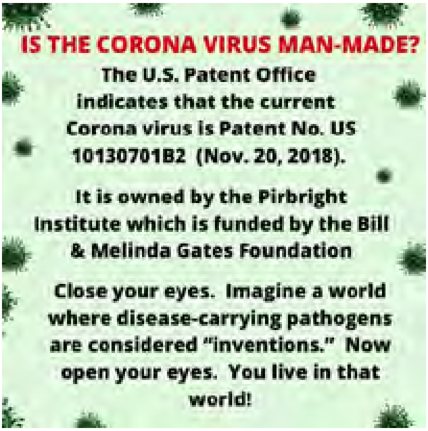
[42] Dunlop, W.G. (2020, 5 August). Beirut blast was not a nuclear explosion. *AFP Fact Check*. Available on the Internet. See also McCarthy, B. (2020, 6 August). QAnon conspiracy theorists seek to link Beirut explosion to Rothschilds. *PolitiFact*. Available on the Internet.

Another example is a conspiracy theory that targeted the Pirbright Institute in 2019. The Pirbright Institute is a biological research organization that holds a patent (from 2018) for a coronavirus that primarily affects chickens and could potentially be used as a vaccine to prevent respiratory diseases in birds (IBV). Even though this patent could not be used for the development of a COVID-19 vaccine, a conspiracy theory published for the first time by the website Humans Are Free in January 2020 claimed that the Pirbright Institute engineered and patented COVID-19 in 2018, well before a cluster of cases was reported in China (in December 2019). Therefore, according to this groundless conspiracy theory, the Pirbright Institute artificially created both the COVID-19 virus and its cure. Once the virus became endemic, the Institute started profiting from the vaccine (see figure 26).[43]

This conspiracy theory further evolved by associating the Pirbright Institute with Bill Gates, falsely claiming that the alleged COVID-19 patent had been developed with funds of the Bill and Melinda Gates Foundation (see figure 27). The truth is that the Foundation was one of the Institute's donors, but they did not fund research connected to coronavirus, including those related to the patent for a possible vaccine to prevent respiratory diseases in birds (IBV).[44]

[43] For example, a website that centres on the dissemination of conspiracy theories and other forms of misinformation and disinformation, published an article falsely claiming that the Institute had patented the COVID-19 stream in an attempt to "create a weaponized viral strain designed to sell more useless, deadly vaccines, while at the same time killing off a few thousand, or perhaps a few million, people". Brown, M. (2020). Fact check: Did Gates Foundation Fund and Does Pirbright Institute Own Coronavirus Patent? *Southwest Times Record*. Available on the Internet.

[44] The Bill and Melinda Gates Foundation has provided two grants to Pirbright, the first in November 2013 for research into diseases affecting livestock and again in June 2016 for research into a universal flu vaccine.



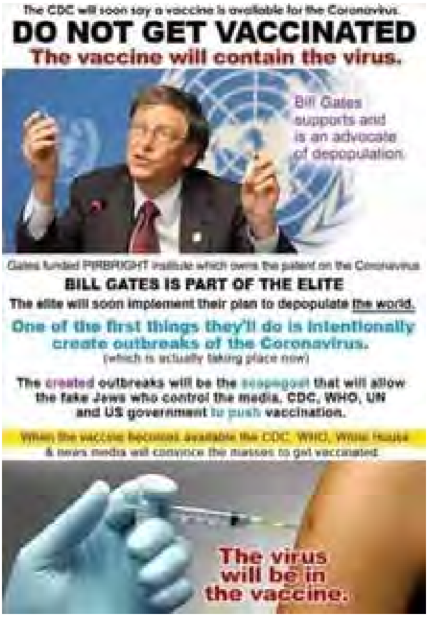
IS THE CORONA VIRUS MAN-MADE?

The U.S. Patent Office indicates that the current Corona virus is Patent No. US 10130701B2 (Nov. 20, 2018).

It is owned by the Pirbright Institute which is funded by the Bill & Melinda Gates Foundation

Close your eyes. Imagine a world where disease-carrying pathogens are considered "inventions." Now open your eyes. You live in that world!

Figure 26: Image obtained from a user's account on Facebook. **Source**: Facebook, user, posted in July 2020.



The CDC will soon say a vaccine is available for the Coronavirus.

DO NOT GET VACCINATED

The vaccine will contain the virus.

Bill Gates supports and is an advocate of depopulation.

Gates funded PIRBRIGHT institute which owns the patent on the Coronavirus

BILL GATES IS PART OF THE ELITE

The elite will soon implement their plan to depopulate the world.

One of the first things they'll do is intentionally create outbreaks of the Coronavirus.

(which is actually taking place now)

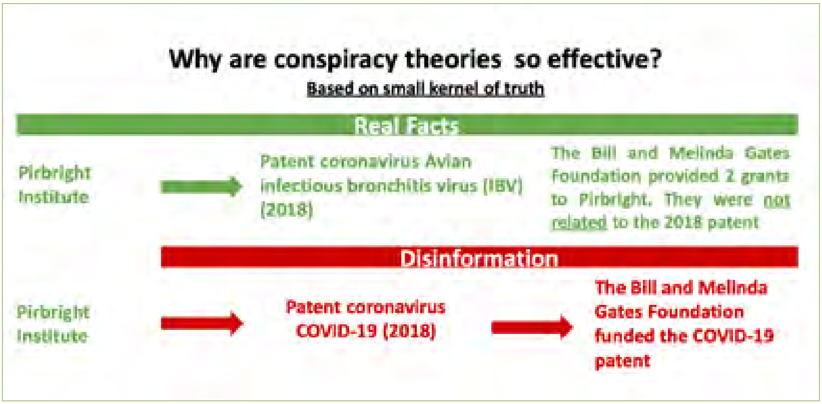
The created outbreaks will be the scapegoat that will allow the fake Jews who control the media, CDC, WHO, UN and US government to push vaccination.

When the vaccine becomes available the CDC, WHO, White House & news media will convince the masses to get vaccinated.

The virus will be in the vaccine.

Figure 27: Image obtained from a user's account on Facebook. **Source**: Facebook, user, posted in July 2020.

As shown in figure 28, this conspiracy theory contained an element of truth (the Pirbright Institute holds a patent for a coronavirus and was funded by the Bill and Melinda Gates Foundation), but the whole story was manipulated and distorted so that the Institute appeared to be the evil mastermind behind COVID-19. The presence of real elements made the conspiracy theory more appealing and, as a result, it rapidly spread on social media. It circulated on different platforms and disinformation outlets, including on a Twitter thread from a well-known QAnon YouTuber who claimed that patents connected to the Pirbright Institute were for the novel coronavirus.[45]



Why are conspiracy theories so effective?

Based on small kernel of truth

Real Facts

Pirbright Institute → Patent coronavirus Avian infectious bronchitis virus (IBV) (2018)

The Bill and Melinda Gates Foundation provided 2 grants to Pirbright. They were not related to the 2018 patent

Disinformation

Pirbright Institute → Patent coronavirus COVID-19 (2018) → The Bill and Melinda Gates Foundation funded the COVID-19 patent

Figure 28: Diagram illustrating the interaction of facts and disinformation in the Pirbright Institute conspiracy theory.
Source: UNICRI

[45] ibid

Conspiracy theories can increase radicalization because their narratives appeal to feelings of alienation from society and foster a “me vs. others” approach.[46] During the COVID-19 pandemic, social media channels populated by violent extremists became fertile ground for conspiracy theories that sometimes converged into some mega-conspiracy theories that combined the most disparate actors including local minorities, public health authorities, the United Nations, secret global elites and even aliens from other planets. Such convergence can have a multiplier effect and mobilize hostility and resistance against any government response to a CBRN emergency.[47]

In some cases, conspiracy theories can evolve to adapt to new contexts, which might include combining multiple conspiracy theories. These mega-conspiracy theories attempt to address different types of events throughout history and frequently refer to worldwide spread conspiracies. An example of a mega-conspiracy theory that was used during the COVID-19 pandemic is The New World Order. This conspiracy theory with anti-Semitic origins states that an elite is working behind the scenes to orchestrate global events and is secretly implementing a dystopian international governing structure that will allow them to take control of the global populace. Followers of this conspiracy theory claim that the members of the elite try to achieve the New World Order through the manufacturing of global events and controlling their associated narratives to sow civil unrest.[48]

[46] Emberland, T. (2020). Why conspiracy theories can act as radicalization multipliers of far-right ideals. *Center for Research on Extremism*. Available on the Internet.

[47] Reuters Fact Check (2021). Fact check-list of claims about Bill Gates includes falsities. *Reuters*. Available on the Internet. FactCheck.org. (n.d.). Person: Bill Gates. *FactCheck.org*. Available on the Internet. FactCheck.org. (n.d.). Person: Dr. Anthony Fauci. *FactCheck.org*. Available on the Internet.

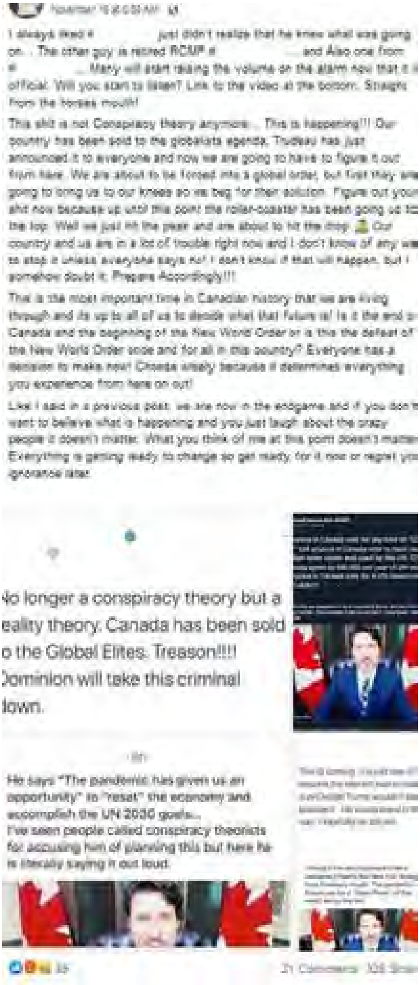
[48] Flores, M. (2022). The New World Order: The Historical Origins of a Dangerous Modern Conspiracy Theory. *Middlebury Institute of International Studies at Monterey*. Available on the Internet.

During the pandemic, this conspiracy theory evolved to incorporate the Great Reset conspiracy. The concept of the Great Reset originated from a non-binding international initiative launched by the World Economic Forum (WEF) in June 2020, which centres on promoting fairer outcomes and rethinking global investment and government expenditure to counter the negative economic effects of the pandemic.[49] The conspiracy theory that distorted the original meaning of the Great Reset linked this initiative to the New World Order conspiracy. The main assumption of the Great Reset conspiracy is that the global elite is using the COVID-19 pandemic as an opportunity to promote radical policies, such as forced vaccination, digital identity cards and the renunciation of private property.[50]

Figure 29 shows an example of how this Great Reset conspiracy theory spread on social media. In this post, the user refers to a speech given by the Canadian Prime Minister where he mentions the Great Reset initiative from the World Economic Forum (WEF). The user falsely claims that the Prime Minister is talking about the conspiracy theory, rather than the initiative.

[49] World Economic Forum (2020). The Great Reset. *World Economic Forum*. Available on the Internet.

[50] Slobodian, Q. (2020). How the 'great reset' of capitalism became an anti-lockdown conspiracy. *The Guardian*. Available on the Internet.



November 15 at 6:03 AM

I always liked # just didn't realize that he knew what was going on. . The other guy is retired RCMP # and Also one from # . Many will start raising the volume on the alarm now that it is official. Will you start to listen? Link to the video at the bottom. Straight from the horses mouth!

This shit is not Conspiracy theory anymore.. This is happening!!! Our country has been sold to the globalists agenda. Trudeau has just announced it to everyone and now we are going to have to figure it out from here. We are about to be forced into a global order, but first they are going to bring us to our knees so we beg for their solution. Figure out your shit now because up until this point the roller-coaster has been going up to the top. Well we just hit the peak and are about to hit the drop 🎢 Our country and us are in a lot of trouble right now and I don't know of any way to stop it unless everyone says No! I don't know if that will happen, but I somehow doubt it. Prepare Accordingly!!!

That is the most important time in Canadian history that we are living through and its up to all of us to decide what that future is! Is it the end of Canada and the beginning of the New World Order or is this the defeat of the New World Order once and for all in this country? Everyone has a decision to make now! Choose wisely because it determines everything you experience from here on out!

Like I said in a previous post, we are now in the endgame and if you don't want to believe what is happening and you just laugh about the crazy people it doesn't matter. What you think of me at this point doesn't matter. Everything is getting ready to change so get ready for it now or regret your ignorance later.

No longer a conspiracy theory but a reality theory. Canada has been sold to the Global Elites. Treason!!!! Dominion will take this criminal down.

He says "The pandemic has given us an opportunity" to "reset" the economy and accomplish the UN 2030 goals...
I've seen people called conspiracy theorists for accusing him of planning this but here he is literally saying it out loud.

35

21 Comments 108 Shares

Figure 29: Screenshots showing the dissemination of the Great Reset conspiracy theory online.
Source: AFP (2020). Justin Trudeau's UN speech is not proof of 'great reset' conspiracy. BOOM. Available on the Internet.

Box 2: How to identify a conspiracy theory?

1. The focus is very unclear and undetermined (e.g., secret global elites want to dominate the world).
2. There is no end to the story – the conspiracy theory keeps evolving with new elements.
3. The number of players involved is unlimited and they grow with the development of the conspiracy theory.
4. The conspiracy theory explains everything, leaving no room for alternative explanation.
5. The conspiracy theory is not confined to a specific historical era, but instead continues across different historical periods.

2.2.4 Hook people with false predictions and expectations

A technique to hook users with a conspiracy theory is to make future predictions and create expectations about the occurrence of a future event.

An example is provided by the QAnon conspiracy theory. It began in October 2017 when an anonymous figure called "Q" posted messages on the website "4chan" alleging that a cabal of Satanic paedophiles, composed of leading figures from the United States Democratic Party, was running an international child sex trafficking network and was conspiring against the United States government. Although Q made several failed predictions from the return of John F. Kennedy Jr. to an imminent civil war, the level of support for the conspiracy did not decrease.[51]

What happens when the predictions turn out to be wrong (as in the majority of cases)? Supporters of conspiracy theories often explain that the expected event did not happen because the secret conspirators impeded it. Therefore, the fact that the predictions were wrong and the expected events did not take place automatically become "evidence" that the conspiracy theory is true.

There are several studies that have attempted to understand why individuals tend to believe incorrect information when they face evidence that contradicts their beliefs. One of the first studies was conducted by the psychologist Leon Festinger who coined the term "cognitive dissonance" to describe the mental distress individuals feel when they try to simultaneously hold onto contradictory beliefs, values, or attitudes (e.g., when a smoker is faced with evidence that smoking is one of the biggest causes of illness and death). To cope with this stressful situation, one either has to change the beliefs and behaviour (e.g., giving up smoking) or else elaborate an explanation that eliminates the dissonance. In the second case, individuals can continue believing incorrect information by creating "alternative" explanations and responses that resolve the cognitive dissonance (e.g., "my neighbour has been smoking all his life and has never had health problems").[52]

[51] To know more about QAnon see: Forrest B. (2021, 4 February). What Is QAnon? What We Know About the Conspiracy-Theory Group. *The Wall Street Journal*. See also McDonald B. (2021, 31 March). How QAnon Reacts to Failed Predictions. *Global Network on Extremism and Technology*.

[52] Festinger, L. (1957). *A Theory of Cognitive Dissonance*.

2.2.5 Appealing to emotions

Users often exhibit a high emotional component on social media. Disinformation strategically depends on emotionally provocative content to cause or increase strong feelings and reactions. A CBRN emergency can be an opportunity to play on people's emotions, especially when there is insufficient clear information and guidance from the official authorities.

An example is offered by the message in figure 30 that was posted by a Thai right-wing group in October 2022. The intention of the post was to alarm readers and provoke a fearful emotional reaction by falsely claiming that the Finnish government had advised its citizens to urgently buy iodine tablets after the war in Ukraine. In fact, the country, which operates nuclear power plants, simply updated guidelines on the use of iodine to protect its most vulnerable citizens in case of an emergency resulting from a nuclear reactor accident.[53]

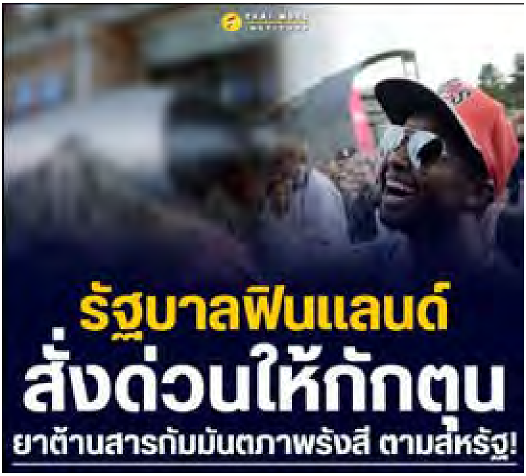


Figure 30: Screenshot of a Facebook user's post that intends to alarm readers by falsely claiming that the Finnish government had advised its citizens to urgently buy iodine tablets after the war in Ukraine. **Source:** Facebook, posted on 14 May 2022.

[53] Aemocha, P. (2022, 26 October). Finland did not advise citizens to 'urgently buy iodine tablets after escalation of war in Ukraine. *AFP Fact Check*. Available on the Internet.

Another common technique used to appeal to emotions is to find an “enemy”, such as blaming local minorities for a CBRN emergency. Figure 31 shows a poster that falsely accuses migrants, hidden in a Trojan Horse, of bringing COVID-19 to Europe.

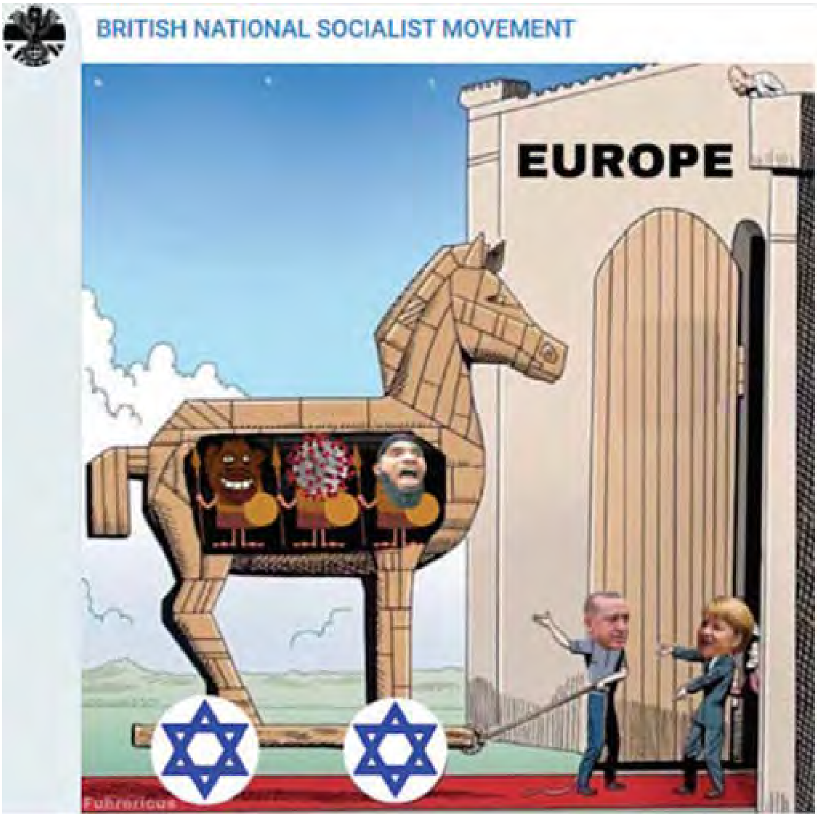


Figure 31: Meme shared by the neo-Nazi British National Socialist Movement.
Source: Telegram, channel “British National Socialist Movement”, posted in 2020.

2.2.6 Stealing cultural property

Another disinformation technique consists of stealing popular cultural symbols. Some extremist groups have used popular and often copyright-protected symbols such as logos and icons to disseminate disinformation without the permission of the copyright holders. The advantage of this technique is that a popular symbol can attract the attention of online users who are not necessarily aware that its use is decontextualized, illegitimate and fraudulent. In some cases, appropriating popular cultural symbols can even contribute to creating (or stealing) a sense of belonging by completely changing its meaning.

For example, far-right extremist groups have stolen the Pepe the Frog cartoon character. Pepe the Frog was created by Matt Furie, an artist and children's book author, as part of his "Boy's Club" series on MySpace in 2005. Even though the character of Pepe the Frog had no connection with violent extremist views, extremists started to use it to promote supremacist and alternative right content in 2016. During the COVID-19 pandemic, the character of Pepe the Frog was also stolen to promote online disinformation about immunization policies and other government measures. The author launched a campaign to "Save Pepe", sharing online content with "peaceful or nice" depictions of the character in an attempt to eliminate its association as a hate symbol. However, he eventually decided to "kill" Pepe the Frog off in a one-page strip for the Fantagraphics' Free Comic Book Day since he could not re-appropriate the character.[54] Figures 32-34 show the contrast between the original character and the manipulated images from violent extremist groups.

[54] Hunt, E. (2017). Pepe the Frog creator kills off internet meme co-opted by White Supremacists. *The Guardian*. Available on the Internet. See also Pettis, B. (2017). *Pepe the Frog: A Case Study of the Internet Meme and its Potential Subversive Power to Challenge Cultural Hegemonies*.

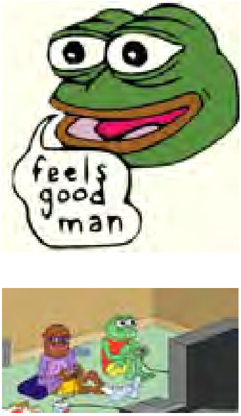


Figure 32: Original Pepe the Frog character with the phrase “feels good man”.

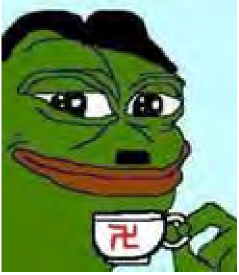


Figure 33: The far-right extremist groups’ misappropriation of Pepe the Frog, here represented with Nazi imagery.
Source: Goodyear, S. (28 September 2016). Pepe the Frog joins swastika and Klan hood in Anti-Defamation League’s hate symbol database. *CBC News*. Available on the Internet.

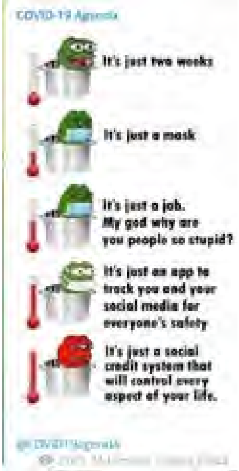


Figure 34: Meme using the misappropriated image of Pepe the Frog shared in an online far-right extremist group channel.
Source: Telegram, channel “COVID-19 Agenda”, posted in 2022.

Using “internet memes” decontextualized from their original context is another technique for spreading online disinformation.[55] An Internet meme consists of a phrase, image, or video that spreads rapidly from person to person via social media channels and messaging applications.[56] Memes often intend to elicit humour to facilitate their spread and have become a social phenomenon to promote ideas, behaviour, or style. Unfortunately, violent extremists also use memes to disseminate hate messages and disinformation.

[55] 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.

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

In particular, some far-right extremists, in efforts to recruit younger members, approach young people with memes that begin as edgy humour, but gradually turn overtly violent and discriminatory.[57] Figure 35 is an example of a meme that manipulates a popular movie (*The Matrix*) to make false accusations against the government of Ukraine. These groups use memes to appear relevant and approachable to younger audiences and create a sense of community. Similarly, the groups also offer 'friendship' to people talking online about being lonely, depressed, or chronically ill, with the ultimate aim of recruiting them by playing on their desire for community.

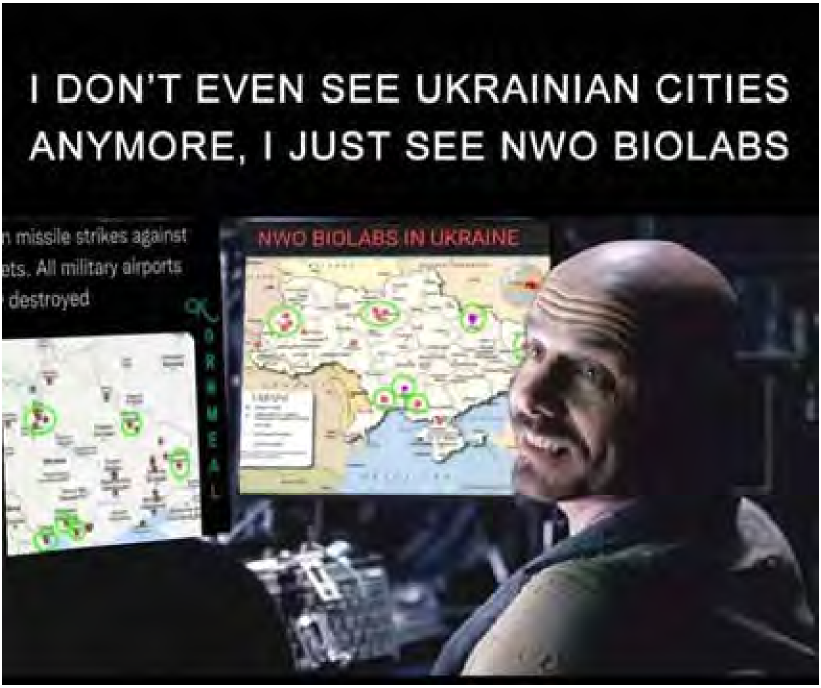

I DON'T EVEN SEE UKRAINIAN CITIES ANYMORE, I JUST SEE NWO BIOLABS
n missile strikes against
ets. All military airports
destroyed
KORNMEAL
NWO BIOLABS IN UKRAINE

Figure 35: Example of the manipulation of a popular movie (*The Matrix*).
Source: Telegram, channel "Holocaust II", posted on 10 March 2022.

[57] Fisher, M. (2021). From memes to race war: How extremists use popular culture to lure recruits. *The Washington Post*. Available on the Internet.

2.2.7 Echo-chambers

An effective technique to spread disinformation is to create an online echo-chamber. An echo-chamber is a virtual environment where a group of individuals participate in online discussions and find their opinions constantly echoed back to them, without exposure to alternative ideas or opinions. Echo-chambers can be used to generate misinformation or disseminate disinformation, distorting the perspective of the individuals and restraining their ability to consider opposing perspectives.[58]

In an online echo-chamber, users search for, interpret and recall information confirming their prior beliefs or ideas. This phenomenon is called confirmation bias.[59] These channels can become instruments for manipulating opinions and radicalizing individuals during a CBRN emergency.

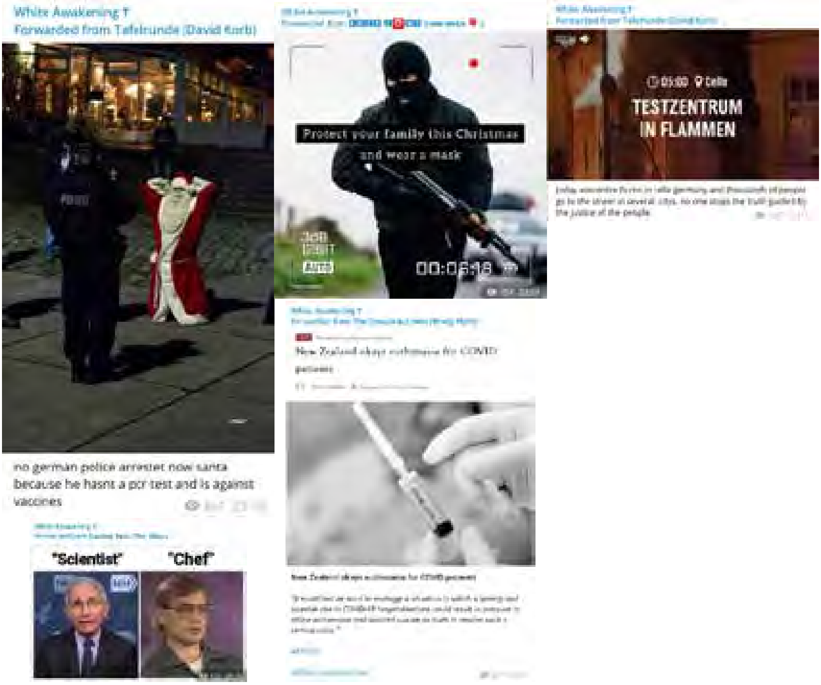
Figure 36 shows an example of a far-right channel (called “White Awakening”) on social media where users’ extremist views are constantly echoed back to them. The selected messages, posted on 25 December 2021, express the same views and, in a combination of irony and false information, criticize the COVID-19 restriction measures adopted by different governments worldwide: the police are “arresting” Santa Claus because he does not have a negative COVID-19 PCR test; a person is wearing a balaclava while holding a gun to “comply” with COVID-19 measures; Dr Anthony Fauci, the director of the National Institute of Allergy and Infectious Diseases (NIAID) at the United States National Institutes of Health, is compared to Jeffrey Dahmer, a notorious serial killer; New Zealand’s Ministry of Health has passed a law to specifically allow doctors to decide whether to euthanize COVID-19 patients (which is not

[58] GCF Global. (n.d.). Digital Media Literacy: What is an echo chamber? *GCFGlobal.org*. Available on the Internet.

[59] Casad, B. J. (2019). Confirmation bias. *Encyclopedia Britannica*. Available on the Internet.

true); thousands of German citizens are burning vaccination centres (which is not true). This channel has no space for an alternative opinion about COVID-19.

Some of the messages are linked to other channels (e.g., channel “Free zone” or “Gazing into The Abyss”) but when the users move to these other channels, they find a similar extremist narrative. In this way, this echo-chamber has an “invisible” mechanism that hooks users by providing information that constantly repeats and confirms the same extremist ideas. As it will be shown in section 2.2.10 (Role of algorithms), algorithms can also facilitate the creation of echo-chambers by suggesting content or connections with other users that have similar ideas or interests.



White Awakening ⚡
Forwarded from Tafelrunde (David Korb)

no german police arrestet now santa because he hasnt a pcr test and is against vaccines

today vaccine centre burns in celle germany and thousands of people go to the street in several citys. no one stops the truth guided by the justice of the people

New Zealand okays euthanasia for COVID patients

Figure 36: Sample of messages shared on Christmas Day 2021 in the far-right channel White Awakening
Source: Telegram, channel “White Awakening”, posted on 25 December 2021.

2.2.8 Attack the credibility of the target of disinformation

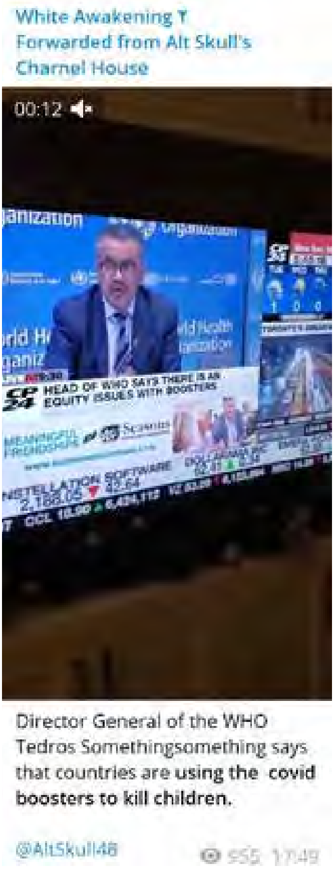
A common online technique to make disinformation appear more legitimate is to attack the credibility of the target of disinformation. The main purpose of this is to damage the target's image, reduce trust in them and, potentially, reduce the target's effectiveness when responding to a CBRN emergency.

For example, during the pandemic, the World Health Organization (WHO) and its Director General Tedros Adhanom Ghebreyesus have been frequent targets of disinformation and conspiracy theories. Disinformation has attempted to jeopardize efforts made by the organization; for example, a popular conspiracy theory falsely claimed that WHO was planning to implement a 'pandemic treaty' that would strip Member States of sovereignty as part of the "World Economic Forum's Great Reset goal". These false claims have been addressed by diverse fact checkers.[60]

Figure 37 shows an example of how the WHO and its Director General have become the targets of attacks, in this case, by a far-right Telegram channel. The false claim manipulates the content of the original speech given by the Director General.

An online defamation campaign could lead to online harassment, especially when a substantial number of people believe the disinformation. A possible consequence of this technique is that any response to the false claims will be disregarded since the target's credibility has been compromised.

[60] Reuters Fact Check. (2022). Fact check-the WHO is not planning to implement a 'pandemic treaty' that would strip Member States of sovereignty. *Reuters*. Available on the Internet.



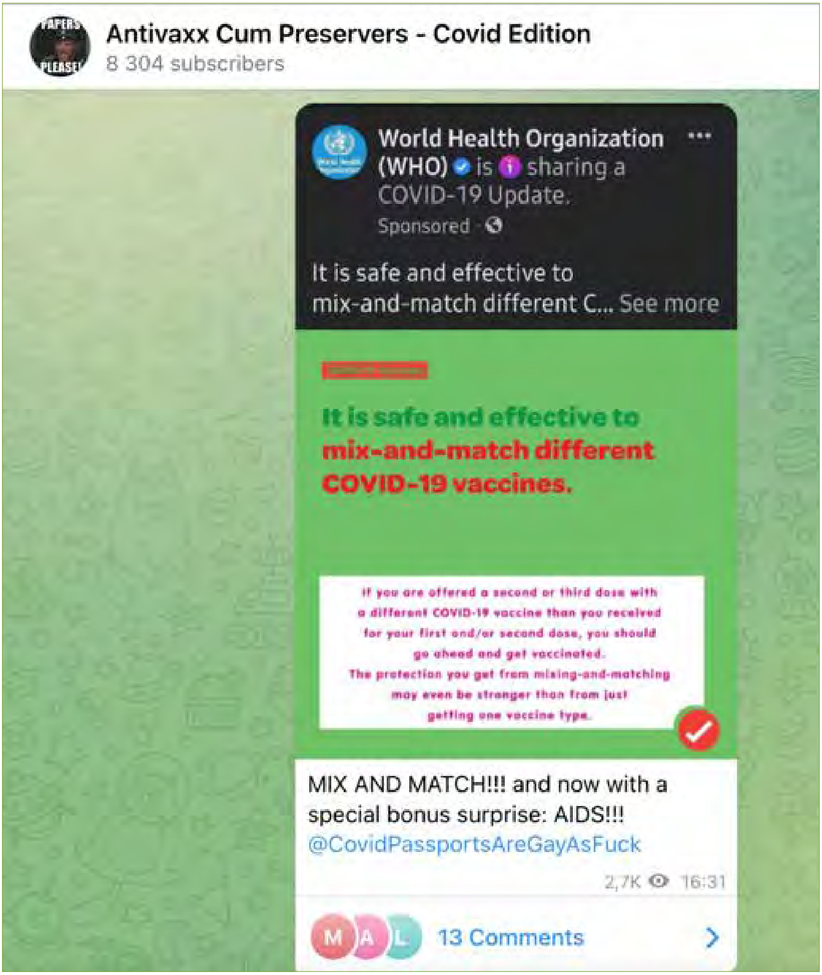
White Awakening
Forwarded from Alt Skull's Charnel House

Director General of the WHO Tedros Somethingsomething says that countries are using the covid boosters to kill children.

@AltSkull48 555 17:49

Figure 37: Screenshot showing a far-right Telegram channel sharing disinformation about the WHO.
Source: Telegram, channel “White Awakening”, posted in 2022.

This technique is often applied against fact-checkers and, in general, efforts to stop disinformation by misquoting, misreading or arbitrarily selecting part of the information shared by a fact-checker. For example, figure 38 shows an attempt to confuse users about a fact-checking post from WHO. The post by WHO (“it is safe and effective to mix and match different COVID-19 vaccines”) is labelled as “false” by changing the colour of the circle (from a green circle with a white check to a red circle with a white check). In this way, by decontextualizing the fact-checking service, correct information is used to spread disinformation.



Antivaxx Cum Preservers - Covid Edition

8 304 subscribers

World Health Organization (WHO) is sharing a COVID-19 Update.

Sponsored ·

It is safe and effective to mix-and-match different C... See more

It is safe and effective to mix-and-match different COVID-19 vaccines.

If you are offered a second or third dose with a different COVID-19 vaccine than you received for your first and/or second dose, you should go ahead and get vaccinated.
The protection you get from mixing-and-matching may even be stronger than from just getting one vaccine type.

MIX AND MATCH!!! and now with a special bonus surprise: AIDS!!!
@CovidPassportsAreGayAsFuck

2,7K 16:31

Figure 38: Screenshot showing how a far-right Telegram channel shares disinformation by using a debunking post originally published by WHO.
Source: Telegram, channel "Antivaxx", posted in 2022.

Figure 39 shows how far-right groups have also targeted fact-checkers in their posts by mocking the efforts made by these organizations and individuals.

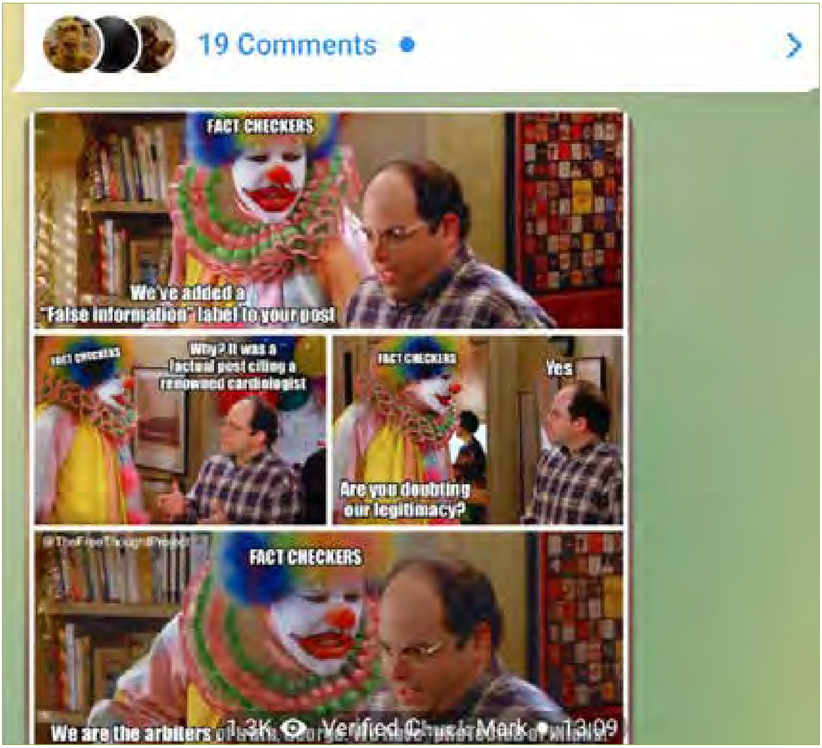

19 Comments
FACT CHECKERS
We've added a "False information" label to your post
FACT CHECKERS
Why? It was a factual post citing a renowned cardiologist
FACT CHECKERS
Yes
Are you doubting our legitimacy?
@TheFreeThoughtProject
FACT CHECKERS
We are the arbiters
1.3K
Verified Check Mark
13:09

Figure 39: Screenshot of a far-right Telegram channel that mocks the work done by fact-checkers.
Source: Telegram, channel "Voluntarist Memes", posted in 2022.

2.2.9 Evading detection and control of authorities

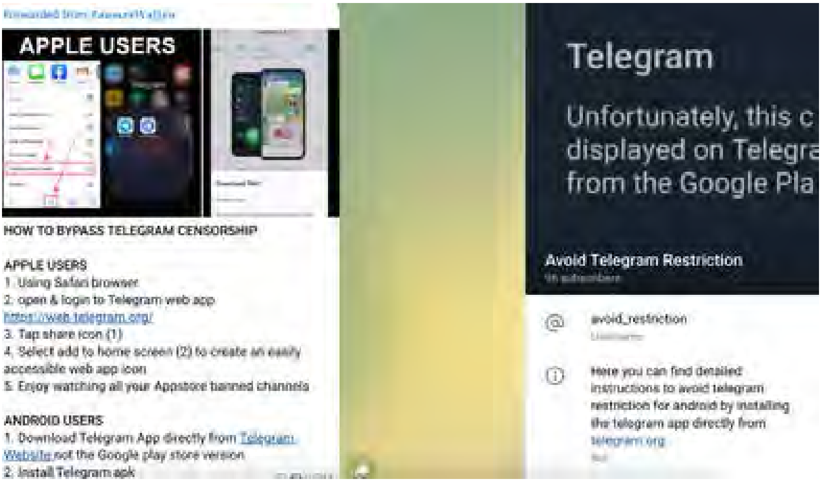
There are different techniques to avoid being tracked online by law enforcement authorities and social media companies while spreading disinformation. For example, ISIL has developed different techniques such as toning-down propaganda language or using *emojis*[61] as substitutes for words that would otherwise be detected (e.g., "weapon", "explosion" and "rocket"). During a ten-day battle to free terrorists from a prison in Hasakah, Syria, at the end of January 2022, ISIL used a hashtag from a fake media outlet to coordinate an online propaganda campaign in which followers from different social media promoted ISIL's prison break.[62]

Another common technique violent extremists use to reduce the risk of censorship is to create a backup account. If some authority identifies and closes their primary account, the violent extremists can immediately move the content to a new account and redirect their audiences. Figure 40 shows how a group of far-right extremists instructed their followers on how to access censored content.

Sometimes, violent extremists redirect their followers to social media platforms and messaging applications where their content is less likely to be detected.

[61] An emoji is "any of various small images, symbols, or icons used in text fields in electronic communication (as in text messages, email, and social media) to express the emotional attitude of the writer, convey information succinctly, communicate a message playfully without using words, etc." Merriam-Webster (n.d.). Emoji. *Merriam-Webster.com dictionary*. Available on the Internet.

[62] Scott, M. (2022). Islamic State evolves 'emoji' tactics to peddle propaganda online. *POLITICO*. Available on the Internet.



HOW TO BYPASS TELEGRAM CENSORSHIP

APPLE USERS
1. Using Safari browser
2. open & login to Telegram web app
https://web.telegram.org/
3. Tap share icon (1)
4. Select add to home screen (2) to create an easily accessible web app icon
5. Enjoy watching all your Appstore banned channels

ANDROID USERS
1. Download Telegram App directly from Telegram Website not the Google play store version
2. Install Telegram apk

Telegram

Unfortunately, this c
displayed on Telegra
from the Google Pla

Avoid Telegram Restriction

avoid_restriction
Username

Here you can find detailed instructions to avoid telegram restriction for android by installing the telegram app directly from telegram.org
Bio

Figure 40: Screenshot of a far-right channel giving instructions on how to access censored content.
Source: Telegram, channel "White Awakening", posted in 2022.

2.2.10 Role of algorithms

A social media algorithm is a set of rules that determines how users see data on the platform. Social media algorithms help to visualize content, such as posts or videos, based on their relevance instead of publishing time. In other words, the algorithms prioritize the content a user sees on the platform according to the likelihood that this user will engage with the content, regardless of publication date.[63] For example, algorithms determine which posts are recommended to users when they scroll through their feed (e.g., Facebook or Instagram).

[63] Golino, M. A. (2021). Algorithms in social media platforms. *Institute for Internet and the Just Society*. Available on the Internet.

In this way, algorithms are designed to support users to identify what can be more interesting and avoid potentially irrelevant or low-quality content. However, social media algorithms can also contribute to disseminating messages with false information and content. For example, a video of the police chasing a boat with criminals went viral in the United States in 2020, when users (mostly teenagers) visualized the video on their For You page (figure 41). Millions of users liked and shared the clip and, as a result, the algorithm began suggesting similar clips, including videos posted by some drug cartels in Mexico that promoted the "potential benefits" of joining the drug trade, such as endless cash, expensive cars, beautiful women and exotic pets (figure 42).[64]

Figure 41: Screenshot of the TikTok video of the boat chase.
Source: TikTok, posted by a user in 2020.

[64] Lopez, O. (2020). Guns, drugs and viral content: Welcome to cartel TikTok. *The New York Times*. Available on the Internet. See also Proceso (2020). "Narcomarketing" La Nueva Estrategia de Cárteles Mexicanos en TikTok: NYT. *Proceso*. Available on the Internet.

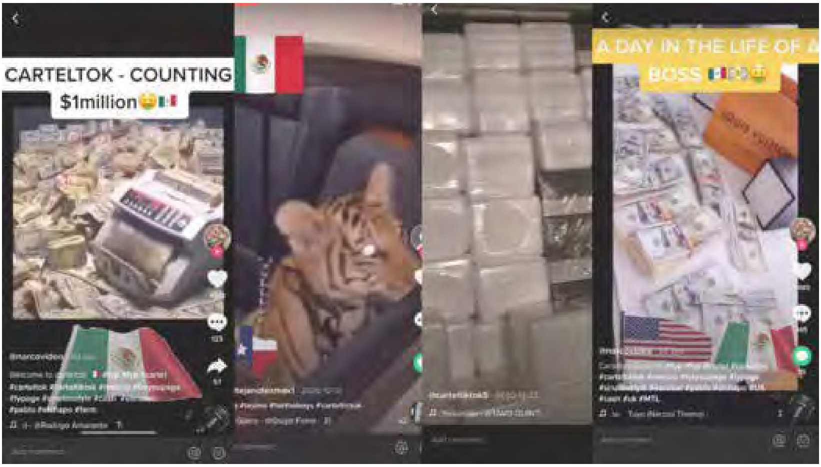
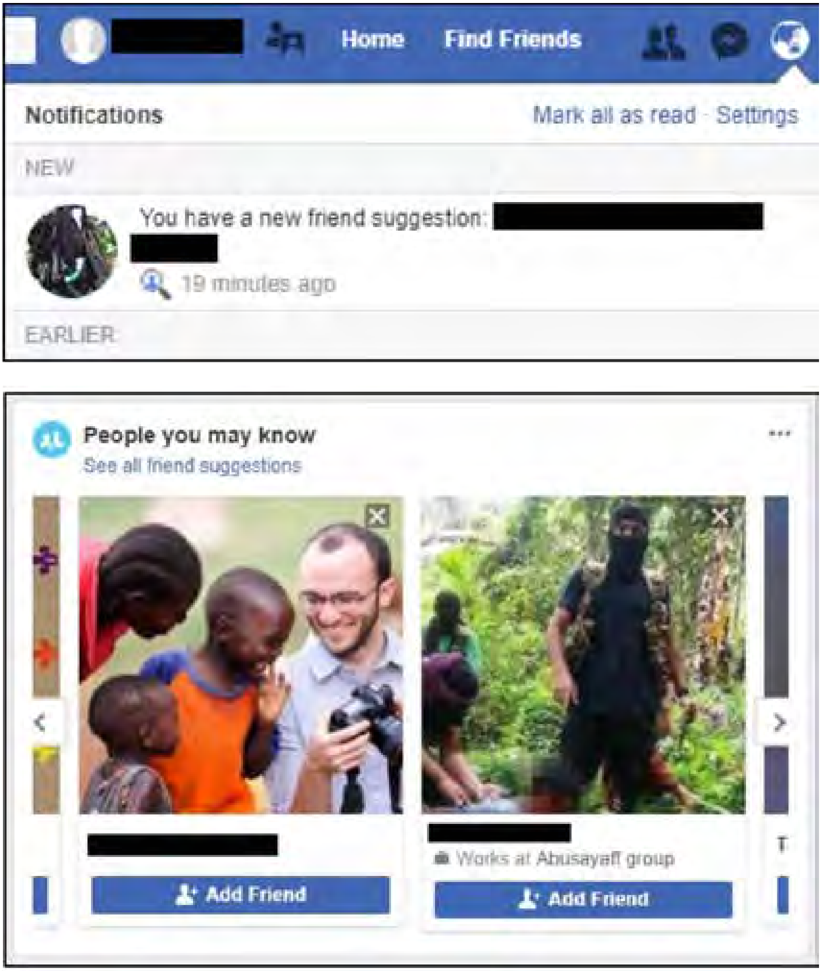


Figure 42: Screenshots of the #CartelTikTok hashtag shared on TikTok.
Source: Morris, E. (2021). Cartel TikTok: Mexican Drug Lords' newest marketing strategy? Glimpse from the Globe. Available on the Internet.

Algorithms can also facilitate interaction between users with similar extremist views and help radicalize and recruit new members into terrorist or violent extremist groups. For example, functions such as "people you may know" or "suggested friends" can facilitate the creation of connections between extremists.

Figure 43 shows an example of a platform's suggested friends' algorithm that recommended ISIL members as suggested friends, connecting extremist profiles and expanding ISIL networks.[65]

[65] Waters, G., & Postings, R. (2018). *Spiders of the Caliphate: Mapping the Islamic State's Global Support Network on Facebook*. Available on the Internet.



Notifications

Mark all as read · Settings

NEW

You have a new friend suggestion: [redacted]

19 minutes ago

EARLIER

People you may know

See all friend suggestions

[redacted]

[redacted]

Works at Abusayaff group

Figure 43: Screenshots of friend suggestions on Facebook that involved members of ISIL.
Source: Waters, G., & Postings, R. (2018). Spiders of the Caliphate: Mapping the Islamic State's Global Support Network on Facebook. Available on the Internet.

3

Debunking disinformation

This chapter describes techniques to effectively debunk disinformation on social media platforms. Debunking is a response measure adopted when disinformation has already taken place. Debunking can be defined as the process of showing that something, such as a belief or theory, is not true, or to show the falseness of a story, idea, statement, etc.[66]

Debunking is not the only strategy to combat disinformation. Another strategy, called pre-bunking, consists in anticipating disinformation before false claims are spread in order to reduce their negative effects. Pre-bunking is based on the inoculation theory, in which a small amount of a virus in the body can promote the generation of antibodies against future exposure to that virus. Similarly, exposure to disinformation (a weak exposure) can help build resistance to future exposure to disinformation.

[66] Encyclopædia Britannica. (n.d.). Debunk. *Encyclopædia Britannica*. Available on the Internet.

Although there are many similarities between debunking and pre-bunking, this Handbook focuses only on debunking techniques. A list of some tools to track and detect disinformation efficiently has been included in the Annex of this Handbook.

Debunking can be seen as a three-step process: the first step is to analyse the disinformation; the second is to decide whether it is worth investing time and resources to debunk the false claim; if the decision is taken to act against disinformation, the third step is to plan and execute the debunking.

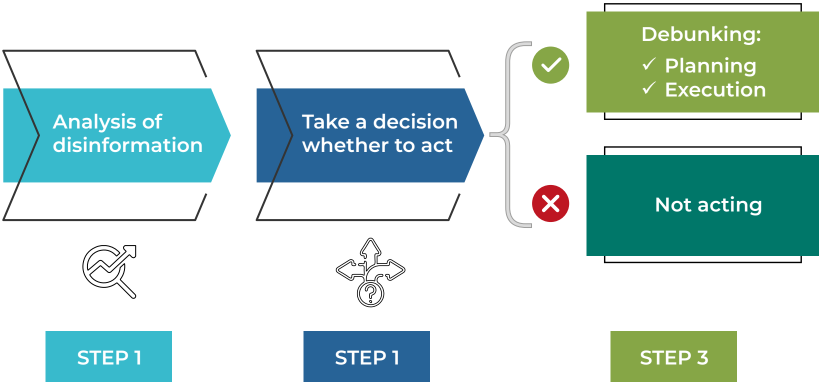


Figure 44: The Debunking three-step process.
Source: UNICRI.

3.1 First step: Analysis of disinformation on social media

When an individual or an organization is targeted by disinformation, the first step is to analyse the content of the false claims. Analysing disinformation can provide essential information about who did it and for what purpose. Some of the different techniques that can be applied are looked at in more detail below (figure 45).

Verification of content in social media

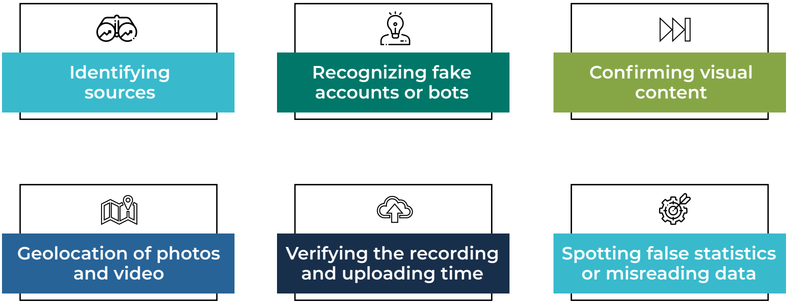

Identifying sources
Recognizing fake accounts or bots
Confirming visual content
Geolocation of photos and video
Verifying the recording and uploading time
Spotting false statistics or misreading data

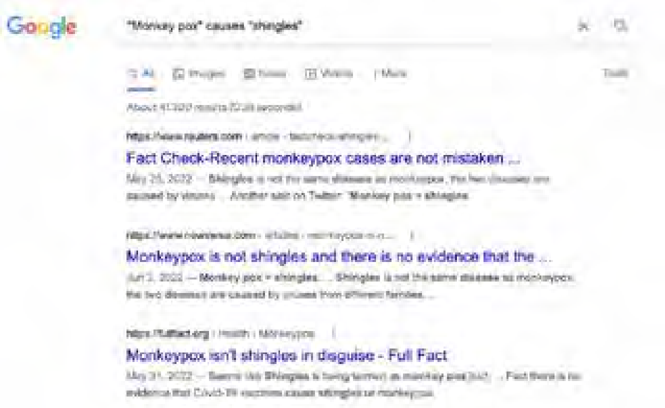
Figure 45: Elements to consider when analysing false information that has been spread through social media.
Source: UNICRI.

a) Identifying and analysing sources

When analysing disinformation, it is important to know the source of the false claim (e.g., an article or a post) and verify if the source contains the same information or if the original content has been modified or manipulated. Equally, checking the date that the false claim was published and who wrote it (what is the background of this person, what other posts/

articles have they published on the Internet/social media) can provide valuable insight.

A possible technique that can be used to identify the source and credibility of a piece of information is to perform a keyword search using a web browser, such as Mozilla Firefox or Google Chrome, to ascertain whether the information or claim has been published on other websites. Figure 46 shows an example of this, where a topic (monkeypox causes shingles) was searched. The terms "monkeypox" and "shingles" are written between quotation marks to find websites that include the exact term in their content. The result shows multiple fact-checking websites where the disinformation is debunked. Other useful symbols that can be used to obtain more specific results in the search are the minus or hyphen (-) to remove specific words from Google search results, the number symbol or hashtag (#) to find the term used in social media platforms, and the tilde symbol (~) before the search keyword to include similar keywords and synonyms, etc.[67]



"Monkey pox" causes "shingles"

About 41,300 results (0.28 seconds)

https://www.reuters.com › article › factcheck-shingles ...

Fact Check-Recent monkeypox cases are not mistaken ...

May 25, 2022 — Shingles is not the same disease as monkeypox. The two diseases are caused by viruses ... Another said on Twitter: "Monkey pox = shingles

https://www.newsweek.com › ... › monkeypox-is-n...

Monkeypox is not shingles and there is no evidence that the ...

Jun 3, 2022 — Monkey pox = shingles. ... Shingles is not the same disease as monkeypox. The two diseases are caused by viruses from different families. ...

https://fullfact.org › Health › Monkeypox

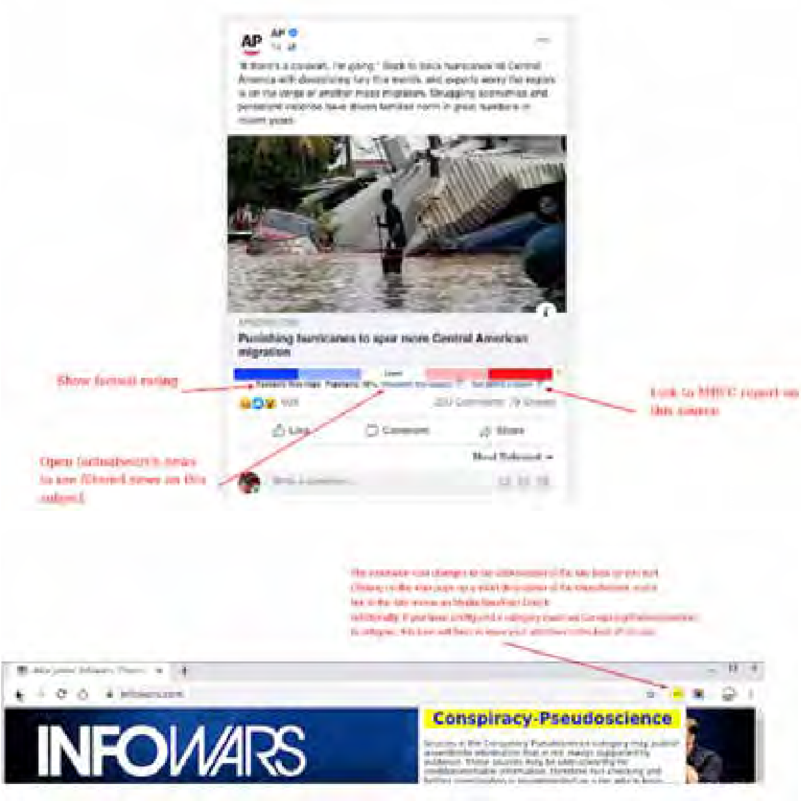
Monkeypox isn't shingles in disguise - Full Fact

May 31, 2022 — Seems like Shingles is being termed as monkey pox ... Fact there is no evidence that Covid-19 vaccines cause shingles or monkeypox.

Figure 46: Screenshot of a search in Google using quotation marks in order to find results that include the exact terms "monkeypox" and "shingles".
Source: Google search, 2022.

[67] See more: Google Search Help (n.d.). Refine web searches. *Google*. Available on the Internet.

Technology tools, such as browser extensions, can analyse a website and determine if and to what extent the source is reliable (see Annex 2). For example, figure 47 shows how a browser extension works when evaluating news and websites online. In the example, the Media Bias/Fact Check Extension displays an icon showing the accuracy or bias on each page that is visited and uses a rating system to evaluate if the source is reliable.



Punishing hurricanes to spur more Central American migration

Figure 47: Screenshots showing how the browser extension Media Bias/Fact Check Extension works.
Source: Media Bias, Fact Check Extension, 2022.

b) Recognizing and excluding fake accounts or bots

In some cases, it is possible to determine if the source of the false claim comes from a fake account (representation on social media of a person, organization or company that does not truly exist) or a bot.[68] A very simple technique is to check the profile of the individual who posted the false information[69], including the profile photo, any potential links to other accounts, a biography, the year in which the account was created, any activity on the social media platform, or any other information that could indicate that the account belongs to a real person. When this type of information is missing, it is likely a bot account. Bots can be also detected because they do not mimic human language perfectly, so the syntax and behaviour used will probably be unusual.

For example, figure 48 shows examples of bot accounts, which have no personal information or images. Figure 49 shows an example of bot activity, where three accounts share the same post with the same headline.

[68] "A 'bot' – short for robot – is a software program that performs automated, repetitive, pre-defined tasks. Bots typically imitate or replace human user behavior. Because they are automated, they operate much faster than human users. They carry out useful functions, such as customer service or indexing search engines, but they can also come in the form of malware – used to gain total control over a computer." Kaspersky. (n.d.). What are bots? – definition and explanation. *Kaspersky*. Available on the Internet.

[69] Social profiles are a description of individuals' social characteristics that identify them on social media sites. A social profile also displays information that helps to understand the type and strength of an individual's relationships with others; for example, their level of participation and contribution in different initiatives, projects, communities, or conversations; their reputation among other participants, and so on. Gartner. (n.d.). Definition of Social Profiles. *Gartner Information Technology Glossary*. Available on the Internet.

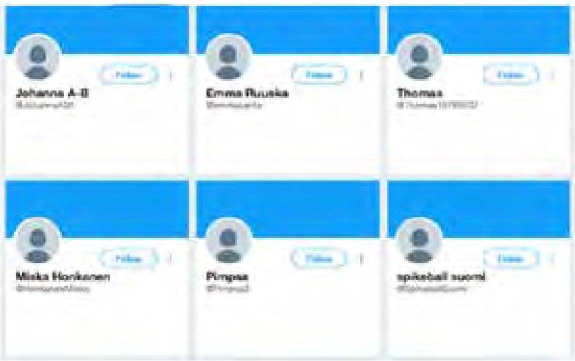


Figure 48: Screenshot of an example of bot accounts.
Source: Atlantic Council's Digital Forensic Research Lab, 2017.

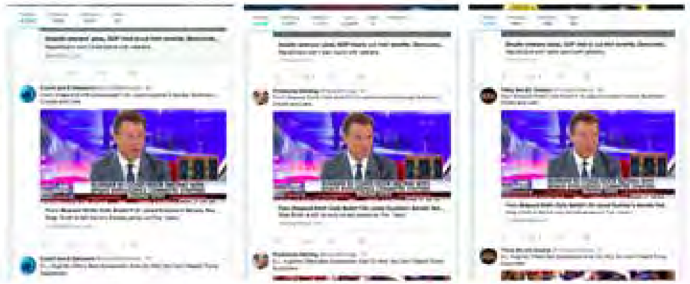
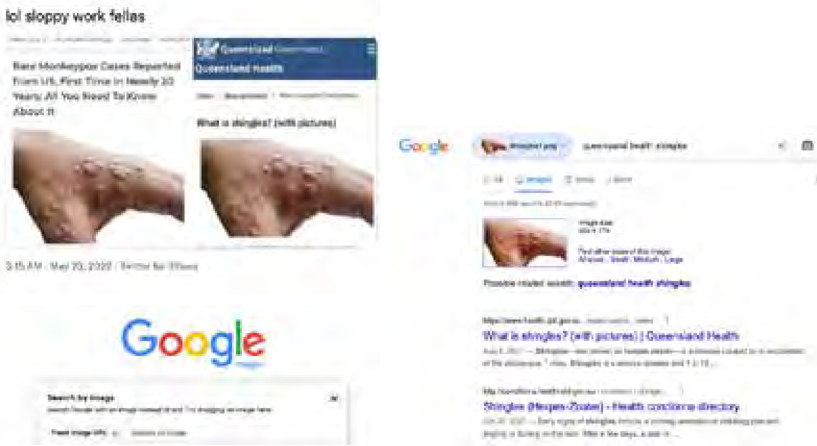


Figure 49: Screenshot of an example of bot activity where different bot accounts share the same post and headline
Source: Atlantic Council's Digital Forensic Research Lab, 2017.

c) Confirming that the visual content is correctly attributed to the original source

The tools Google Image Search or Tineye can be used to find the source of an image. The image can be uploaded directly to the browser from a computer or by pasting the URL of the image. The browser will indicate whether the picture has been shared on other websites. Using this tool, it is possible to identify if the image has been taken out of context and find on which websites and with what information the image has been shared in the past. For example, figure 50 demonstrates how to search for the source of an image in Google Image

Search: first, the image should be uploaded or its URL should be pasted; then simply click the search button to find the websites where the image has been published. The results from the example, which refer to an image that had been mislabelled as monkeypox, show the original source of the image, as well as a website that addresses the false claim.[70]



lol sloppy work fellas

Rare Monkeypox Cases Reported From US, First Time In Nearly 20 Years: All You Need To Know About It

Queensland Health

What is shingles? (with pictures)

3:15 AM · May 20, 2022 · Twitter for iPhone

Search by image

Possible related search: queensland health shingles

What is shingles? (with pictures) | Queensland Health

Shingles (Herpes-Zoster) - Health conditions directory

Figure 50: Example of a search with Google Image Search to verify the origin of a picture that was posted by a user on Twitter in 2022. The Twitter user claimed that monkeypox did not exist and juxtaposed monkeypox and shingles article images to prove that the monkeypox coverage was fake. However, an analysis of the picture posted on Twitter through Google Image Search shows the website where the picture was originally published in which it clarified that monkeypox exists.
Source: Twitter, posted by a user in 2022; Screenshots from a Google Image Search, 2022.

[70] Schirrmacher, S. (2022). Fact check: Four fakes about Monkeypox. *Deutsche Welle*. Available on the Internet.

d) Verifying the recording and upload time of content[71]

If a video is the source of disinformation, it can be useful to verify when the video was uploaded and if it has been manipulated. Confirming the veracity of videos related to public events, such as protests or conferences, is generally straightforward since information about the event can be found in news reports and on other social media sites. In other cases, where the video is about an event that was not public or was not previously planned (such as an accident, an attack, or a natural phenomenon), it might be necessary to seek additional information.

For example, YouTube videos are time-stamped in Pacific Standard Time (PST) from when the upload begins. In order to obtain additional information about a video uploaded to YouTube, Amnesty International offer a tool called the YouTube Data Viewer. This tool makes it possible to reverse image search with images from the video. To use the tool, the user only needs to paste the YouTube URL of the video to be verified. Figure 51 shows an example search performed on the YouTube Data Viewer for a video entitled "Covid-19: Vaccines, Boosters and Outpatient Therapies", which was originally posted on YouTube. The tool displays information about the video, including the upload time and date.

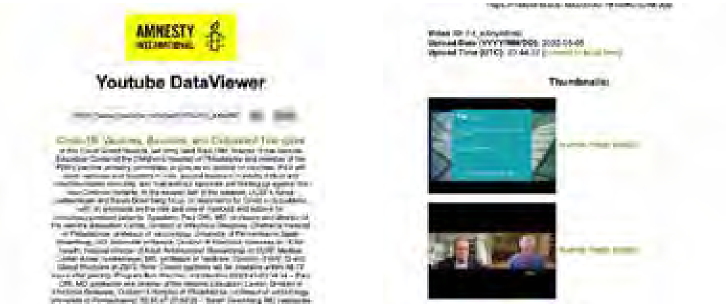


Figure 51: Screenshots of a video search performed in 2022 on the YouTube Data Viewer.
Source: YouTube Data Viewer, 2022

[71] Silverman, C. (Ed.) (2016). *Verification handbook*. Available on the Internet.

e) Geolocation of photos and video[72]

Most of the time an image or a video contains helpful information. This includes a distinctive streetscape, a building, a church, a line of trees, a mountain range, a minaret, or a bridge – all of which can be compared with satellite imagery and geolocated photographs. An image or a video could also provide other relevant details: a business name could be listed in online classifieds (online advertising that includes listings of goods, products or services on Internet sites) or a local directory, and dialogue and accents in the video could be used to determine a specific region, a street sign could signal a precise location, car registration plates or advertising billboards could indicate information about the location, the sunlight or a shadow could help to guess the approximate time of day of the event.

Some tools allow for the automatic detection of the geolocation, even though not all images have a geolocation tag. For example, geolocation can be verified online using Google Maps (including Street View to get a closer image or the "Photos" option to check if geolocated photographs match the video location), Wikimapia, or Google Earth. Figure 52 shows the interface of GeoImgr, which can be used to automatically detect the geolocation of an uploaded image. The user can click on the blue screen (right) and select the file to upload.

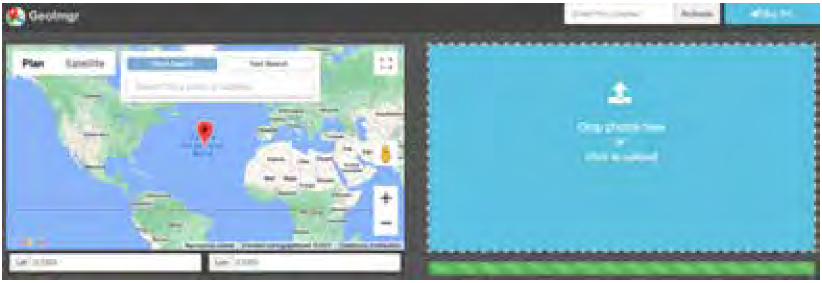


Figure 52: Screenshot of the main page of the GeoImgr website.
Source: GeoImgr, 2022.

[72] Silverman, C. (Ed.). (2016). *Verification handbook*. Available on the Internet.

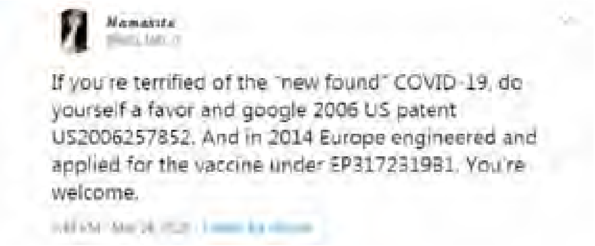
f) Spotting false statistics or misleading data

Poor reading of statistics and data manipulation is also a tactic used to spread disinformation on social media. Accurate data can also be decontextualized to suggest that a false claim is true. This was the case with the official document of May 2021 from the Italian Medicines Agency that provided statistics on deaths that occurred after the first and the second dose of COVID-19 vaccines in Italy (see 2.2.2 Mimicking scientific debate) or the case of the Pirbright Institute patent that was erroneously classified as a COVID-19 patent (see figures 53 and 54).

The examples below show a tweet falsely claiming that there were COVID-19 patents in 2006 and 2014. In this case, it is very simple to demonstrate that this tweet is not true by looking through the online United States Patent and Trademark Office (USPTO) patent database.[73] The United States patent numbers in the tweet do not exist. A Google search for the patent number leads to the page of the United States National Institute of Health's PubChem open chemistry database. However, the PubChem page shows that a similar patent as the one mentioned in the post is linked to an application for a patent published in 2006. It is almost identical to the one mentioned in the social media posts, but with an extra zero. This patent, with an extra zero, refers to nucleic acids and proteins from Severe Acute Respiratory Syndrome (SARS) Virus, which can be used in the preparation and manufacture of vaccine formulations for the treatment or prevention of SARS.[74] The European patent mentioned in the tweet is the Pirbright Institute 2018 patent for a coronavirus that primarily affects chickens and could potentially be used as a vaccine to prevent respiratory diseases in birds (IBV), explained previously in section 2.2.3.

[73] The online United States Patent and Trademark Office (USPTO) patent database is available on the Internet.

[74] Dunlop, W. G. (2020). False claims on patents fuel novel coronavirus conspiracy theories online. *AFP Fact Check*. Available on the Internet.



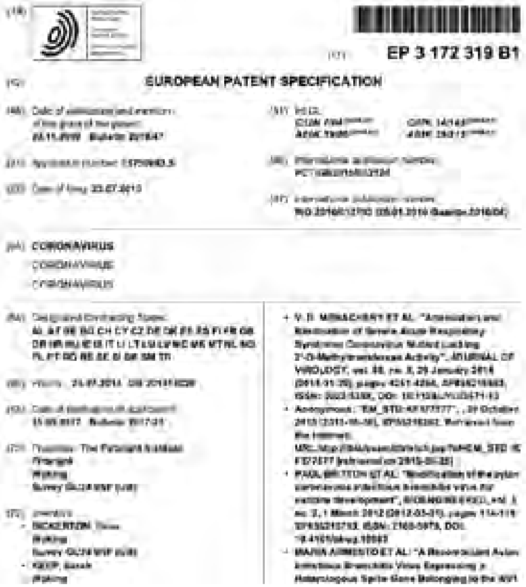
Namasita

[illegible]

If you're terrified of the "new found" COVID-19, do yourself a favor and google 2006 US patent US2006257852. And in 2014 Europe engineered and applied for the vaccine under EP3172319B1. You're welcome.

[illegible]

Figure 53: Screenshot of a tweet falsely claiming that there were COVID-19 patents, one in the United States and one in Europe (referring to the Pirbright Institute patent number). None of the patents refer to COVID-19.
Source: AFP Fact Check.



(19) [illegible]

(11) EP 3 172 319 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent: [illegible] Bulletin [illegible]

(21) Application number: [illegible]

(22) Date of filing: [illegible]

(51) Int Cl.: [illegible]

(86) International application number: PCT/[illegible]

(87) International publication number: WO [illegible]

(54) CORONAVIRUS
CORONAVIRUS
CORONAVIRUS

(84) Designated Contracting States: AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: [illegible]

(43) Date of publication of application: [illegible] Bulletin [illegible]

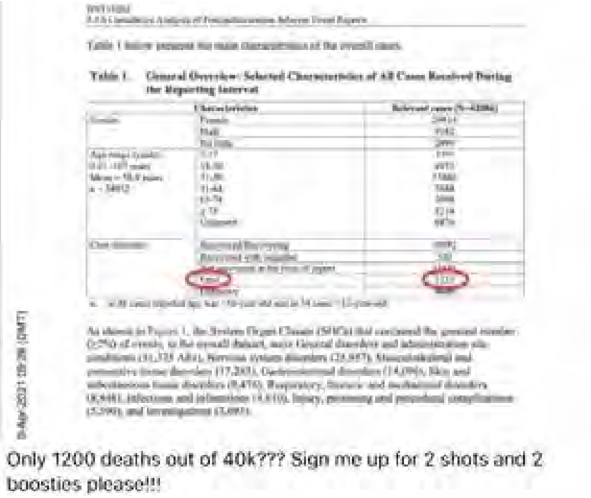
(73) Proprietor: The Pirbright Institute
Pirbright
Woking
Surrey [illegible] (GB)

(72) Inventors:
- BICKERTON, [illegible]
Woking
Surrey [illegible] (GB)
- KEEP, Sarah
Woking

- V. D. MENACHERY ET AL: "Attenuation and Restoration of Severe Acute Respiratory Syndrome Coronavirus Mutant Lacking 2'-O-Methyltransferase Activity", JOURNAL OF VIROLOGY, vol. 88, no. 8, [illegible] pages [illegible], ISSN: [illegible], DOI: [illegible]
- Anonymous: [illegible] Retrieved from the Internet: URL: [illegible]
- PAUL BRITTON ET AL: "Modification of the avian coronavirus infectious bronchitis virus for vaccine development", [illegible] 1 March 2012 (2012-03-01), pages [illegible], ISSN: [illegible], DOI: [illegible]
- MARIA ARMESTO ET AL: "A Recombinant Avian Infectious Bronchitis Virus Expressing a Heterologous Spike Gene Belonging to the [illegible]

Figure 54: Screenshot of the 2018 Pirbright Institute patent of a new approach to vaccine development to make improved infectious bronchitis virus vaccines for poultry. This patent is not related to COVID-19.
Source: European Patent Office.

In figure 55, a post in a far-right channel on Telegram tries to mislead readers by showing a table with the number of patients who died of COVID-19. However, if you read carefully, the table does not mention COVID-19.



[illegible]

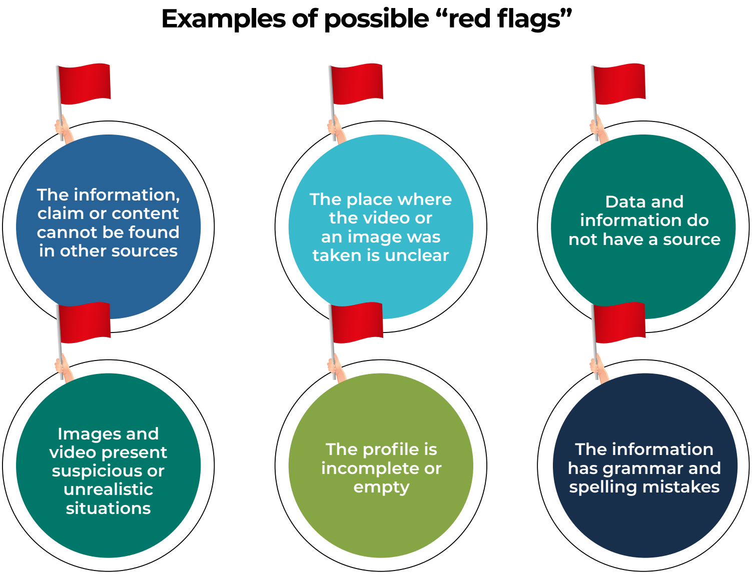
Table 1. General Overview: Selected Characteristics of All Cases Received During the Reporting Interval

[illegible]

Only 1200 deaths out of 40k??? Sign me up for 2 shots and 2 boosties please!!!

Figure 55: Far-right channel misreading statistics that are not related to COVID-19 cases.
Source: Telegram, channel “White Awakening”, posted in 2022.

As explained in this section, during the process of analysing information some elements can raise “red flags” or warnings that something could be false or come from an unreliable source. Figure 56 shows some elements that should be considered when assessing if the information is false.



Examples of possible “red flags”

The information, claim or content cannot be found in other sources

The place where the video or an image was taken is unclear

Data and information do not have a source

Images and video present suspicious or unrealistic situations

The profile is incomplete or empty

The information has grammar and spelling mistakes

Figure 56: Examples of possible “red flags” when trying to corroborate the information found online.
Source: UNICRI.

Analysing disinformation can provide essential information and help us to move to the next two steps.

Summary: Analysing disinformation

- ✓ Identify sources and assess if they are reliable.
- ✓ Recognize and exclude fake accounts and bots.
- ✓ Confirm that images and videos are attributed to the original source.
- ✓ Verify the uploading information of a video.
- ✓ Identify the geolocation of an image or video.
- ✓ Spot false statistics or misleading data.
- ✓ Assess the reliability of a website by checking the sources, URL , phrasing and punctuation of the text, and general content.

Figure 57: Summary of elements to consider when analysing disinformation.
Source: UNICRI.

3.2 Second step: Making a decision

The second step is to decide whether to respond to a false claim. The decision needs to be based on information collected during the previous step and on additional factors that should be considered before investing time and resources in debunking the disinformation.

The following questions can help one to decide:

How widely spread is the false claim?

Debunking may not be necessary when the false claim or conspiracy theory is not spread widely or does not have the potential to cause harm now or in the future.[75] Debunking could be needed when a false or similar claim is widely spread. For example, false information regarding the transmission of a virus or its origin or a disinformation post that receives a considerable amount of attention could all require a response, since a false claim that spreads fast online could make a large portion of a population change their opinion about the immunization campaign or the precautionary measures. Attention can be measured in likes or similar reactions on social media, the number of comments engaging with the post, or the times the post has been shared or reposted. Another possible case where debunking might be required is if the same digital or graphic content is shared widely in multiple posts by different users. This can happen when popular false statistics or manipulated images go viral and become a common topic of discussion among users of social media platforms.[76]

[75] Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. Available on the Internet.

[76] If a video, image, or story goes viral, it spreads quickly and widely on the internet through social media and email.

How relevant is the false claim to your field or expertise?

Debunking disinformation may not be necessary if the false claim does not relate to the area of expertise, positions, or organizations in which the targeted victims work. For example, suppose the false claim just mentions a general conspiracy theory without any particular reference to the victim's work or position, such as "you are probably part of the reptilian elite". In that case, it is probably not worth investing resources to debunk it.

Who is behind it? Is it only an isolated episode or part of orchestrated online harassment?

It is important to understand if the disinformation is an isolated episode or the result of orchestrated online harassment. Online harassment can take the form of Internet trolling (a deliberately offensive or provocative online posting to upset someone or elicit an angry response from them), cyber-bullying (which involves the reception of offensive or malicious messages or can even include websites built with harmful intent towards an individual or certain groups of people)[77] or hate speech (which covers all forms of expression that spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance).[78]

In the case of orchestrated online harassment, the victim could consider some alternative measures before replying, such as ignoring the comments, blocking the user, or reporting malicious content.[79] If the person wants to report the harassment, it is important to take screenshots of the offensive

[77] Council of Europe (2017). Internet – Addressing the challenge. *Internet Literacy Handbook*. Available on the Internet.

[78] Council of Europe (n.d.). Hate speech in *Council of Europe Freedom of Expression*. Available on the Internet.

[79] Endsleigh (n.d.). What is internet trolling? *Endsleigh*. Available on the Internet.

content as evidence. Remembering that a reply could generate further engagement of the aggressor or other online trolls and bullies is also essential. For this reason, when encountering disinformation, it is necessary to control emotions before drafting a response. Most of the time, it is better to take some time to calm down and decide what would be the best option. In case of online harassment, the victim can access online resources and helplines that specialize in providing the required assistance.[80]

Be careful of the backfire effect

It is important to stress that any correction could have unintended consequences, including reinforcing false claims. This is called the backfire effect, which can be defined as a situation where "correction inadvertently increases belief in, or reliance on, misinformation relative to a pre-correction or no correction baseline".[81] This could be due to the nature of repetition, which creates familiarity, and familiar information is frequently perceived to be more truthful than new information. Repeating disinformation can make removing it from person's memory more complex, sometimes making readers recall it as the actual fact. This, however, does not mean that the false claim should not be mentioned while debunking, since repeating it is safe in many circumstances, and can even increase the effectiveness of the correction.[82]

In some instances, it may be preferable to focus on the facts rather than the myth. For example, creating an online post

[80] Some of the organizations that can provide this type of assistance include Access Now and The Cyber Helpline. Available on the Internet.

[81] Ecker, U.K.H., Lewandowsky, S. & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5, 41.

[82] Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. Available on the Internet.

that highlights the benefits and safety of a vaccine may have a greater beneficial effect than starting by directly debunking a vaccine-related false claim. The former would generate a positive set of talking points, whereas the latter could unintentionally redirect the conversation to the false claim, reinforcing its familiarity in the public's consciousness.[83] The possibility of contributing to the backfire effect should therefore be taken into account when deciding whether or not to act.

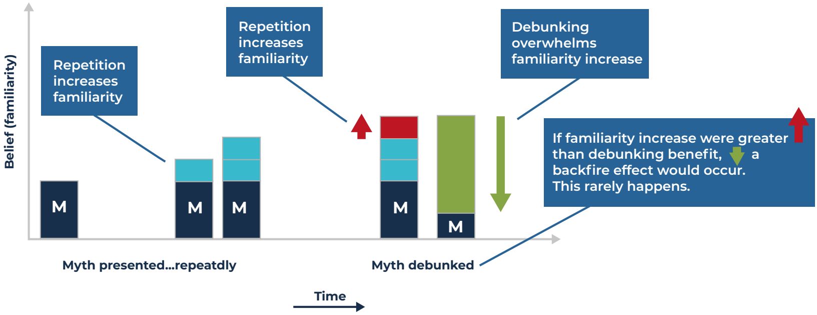

Belief (familiarity)
Repetition increases familiarity
Repetition increases familiarity
Debunking overwhelms familiarity increase
If familiarity increase were greater than debunking benefit, a backfire effect would occur. This rarely happens.
M
M
M
M
M
Myth presented...repeatdly
Myth debunked
Time

Figure 58: Familiarity in the backfire effect.
Source: The Debunking Handbook.

To avoid the backfire effect, one of the most effective approaches is to focus on the facts rather than the myth when communicating, which can be done by using the fact as the headline or the most noticeable part of the debunk.[84]

Nevertheless, recent evidence does not provide reasons to avoid debunking for fear of a backfire effect.[85]

[83] ibid

[84] Cook, J., Lewandowsky, S. (2011). *The Debunking Handbook.*

[85] Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020.* Available on the Internet.

Do you have enough information to debunk a false claim?

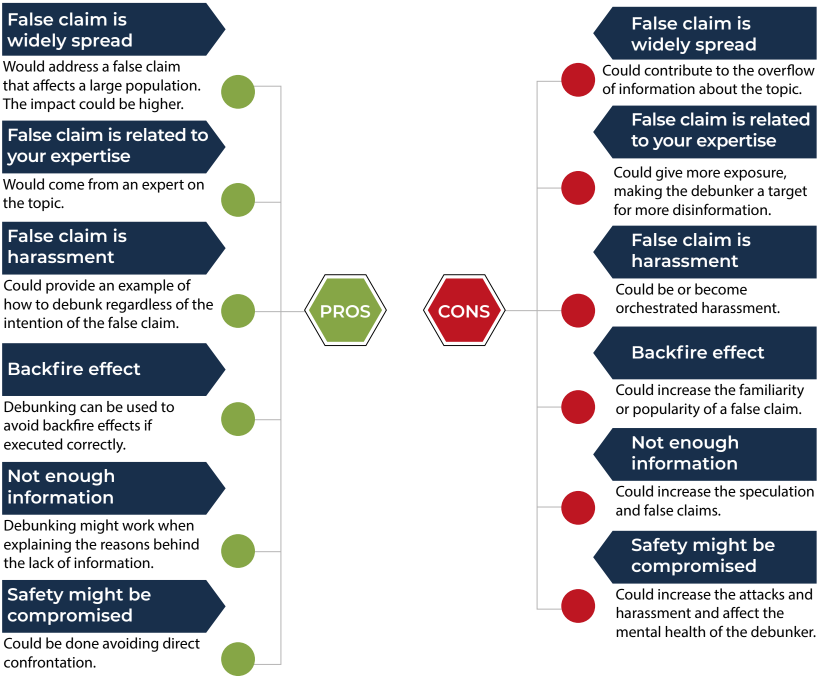
In some cases, the victims of disinformation may not have enough information to debunk a false claim either because it is outside their area of expertise or because there is not enough information or evidence to support a counterargument. In this case, it is possible to communicate to the audience that the information is limited or that updates will probably be provided as the situation evolves. For example, there can be disinformation regarding a new virus where more research is needed to assess the transmission. Since the situation is developing, the information could eventually be updated. In this event, a government can clarify that currently there is not enough knowledge about new viruses and that the risk assessment could change as the situation evolves. By doing this, the audience can be aware that the information shared by the government could be updated if new elements are identified during the research.

Consider safety!

Disinformation can come from violent groups. It is important to consider carefully whether it is safe to engage in a conversation with potentially violent individuals. For example, suppose members of a neo-Nazi group target an individual or an organization. In that case, the victims should consider if it is safe to reply, since the neo-Nazis could organize other attacks, such as hacking emails or online harassment. In these cases, debunking can be carried out by addressing the false claims of these groups without directly replying to their comments or mentioning their name (e.g., a far-right group could falsely claim that a research institute is making a bioweapon; the institute could debunk the false claim and explain the scope of their activities without directly answering to the far-right group).

Figure 59 summarizes the main elements to consider when deciding whether to debunk a false claim or not. The image lists some arguments in favour of debunking on the left, while the right side shows the cons.

Making a decision of whether to act



PROS

False claim is widely spread
Would address a false claim that affects a large population. The impact could be higher.

False claim is related to your expertise
Would come from an expert on the topic.

False claim is harassment
Could provide an example of how to debunk regardless of the intention of the false claim.

Backfire effect
Debunking can be used to avoid backfire effects if executed correctly.

Not enough information
Debunking might work when explaining the reasons behind the lack of information.

Safety might be compromised
Could be done avoiding direct confrontation.

CONS

False claim is widely spread
Could contribute to the overflow of information about the topic.

False claim is related to your expertise
Could give more exposure, making the debunker a target for more disinformation.

False claim is harassment
Could be or become orchestrated harassment.

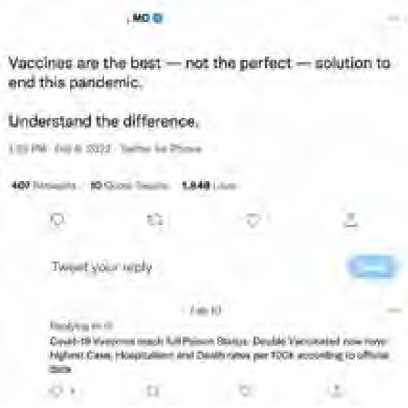
Backfire effect
Could increase the familiarity or popularity of a false claim.

Not enough information
Could increase the speculation and false claims.

Safety might be compromised
Could increase the attacks and harassment and affect the mental health of the debunker.

Figure 59: Points to consider when making a decision.
Source: UNICRI.

An example of deciding whether or not to act when encountering disinformation can be seen below. Figure 60 shows the reply of a Twitter user to a statement made by a doctor. In this case, the doctor decided not to debunk the false claim since the user did not provide any sources, and the comment did not generate any reactions from other users.



, MD

Vaccines are the best — not the perfect — solution to end this pandemic.

Understand the difference.

1:33 PM · Feb 8, 2022 · Twitter for iPhone

407 Retweets 10 Quote Tweets 1,848 Likes

Tweet your reply

Reply

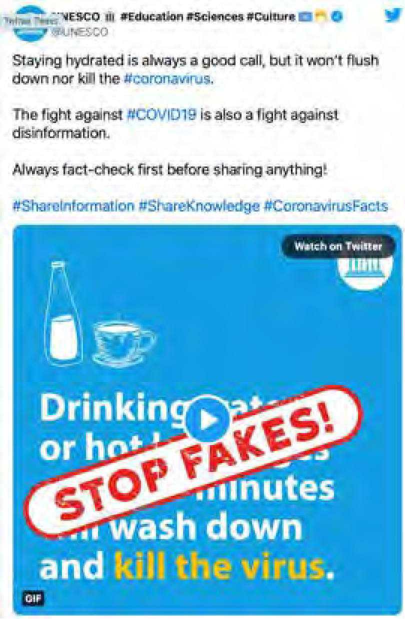
· Feb 10

Replying to @

Covid-19 Vaccines reach full Poison Status: Double Vaccinated now have Highest Case, Hospitalisation and Death rates per 100k according to official Data

Figure 60: Screenshot showing user's response, based on false information, to a doctor's Twitter post.
Source: Twitter, user, posted on 8 February 2022.

Figure 61 shows an example of debunking a widely spread false claim about an alleged treatment for COVID-19, rather than debunking a false claim made in a single post or comment.



UNESCO #Education #Sciences #Culture
@UNESCO

Staying hydrated is always a good call, but it won't flush down nor kill the #coronavirus.

The fight against #COVID19 is also a fight against disinformation.

Always fact-check first before sharing anything!

#ShareInformation #ShareKnowledge #CoronavirusFacts

Watch on Twitter
Drinking water or hot ... minutes ... wash down and kill the virus.
STOP FAKES!
GIF

Figure 61: Example of debunking a widely spread false claim about possible treatments for COVID-19.
Source: Twitter, UNESCO, posted in 2022.

3.3 Step 3: Debunking disinformation

If a person or an organization decides to respond, the third step is debunking the false claim. Debunking can be divided into two phases: planning how to respond and executing the debunking.

Planning phase

Three elements need to be clearly defined before replying to a false claim: the target audience of the message, the topic or false claim that will be debunked, and the means or platforms in which the message will be transmitted.

Planning
Topic
Target audience
Means

Figure 62: Elements to consider when planning a debunking strategy.
Source: UNICRI.

Defining the target audience

Debunking should be tailored to a specific audience. The target audience will most likely be the specific group(s) among which the disinformation has been spread. If it is not evident who the audience is, the platform or means used to spread the disinformation, or defining characteristics of the group may provide more insight (e.g., if disinformation is being spread on TikTok, the message will probably need to be made considering teenagers as the target group).

Defining the target audience

- Who has been mainly affected by the disinformation?
- Which group(s) are more susceptible to believing the false claim? Consider the relevant characteristics of the target group(s) depending on the topic (age, gender, education, religion, values, beliefs, interests, etc)

Figure 63: Guiding questions to define the target audience when debunking disinformation.
Source: UNICRI.

Defining the topic

It is important to select which false claims to debunk. When selecting the disinformation that will be debunked, it is important to consider the type of false claims being spread so that the debunking efforts can centre on topics that would have a larger impact. For example, some false claims and myths are more widely spread than others. Some false claims could have a larger impact on society if they are not debunked, increasing their priority in terms of debunking. In addition, some general questions can help to focus the debunking efforts on the right topic, such as: What false claims could have the most impactful negative consequences? What information do people need (identify trending questions or disinformation issues)? What are the main topics addressed when sharing disinformation? Are there any questions or topics you can anticipate and debunk considering the current disinformation being shared?

Examples during the COVID-19 pandemic

How to anticipate/pre-bunk disinformation about...

- The origin of the virus?
- The means of transmission?
- The possible containment measures?
- The immunization policy?

Figure 64: Example of guiding questions to define the false claim to be debunked.
Source: UNICRI.

Defining the means

Depending on the target audience, the means of dissemination could vary. The type of content and strategy will depend on the selected platform. For example, if the debunking efforts will be disseminated through TikTok, then the best option is to create a short video with the information. If the debunker is going to use Facebook, then posts with graphics can be considered alongside the videos.

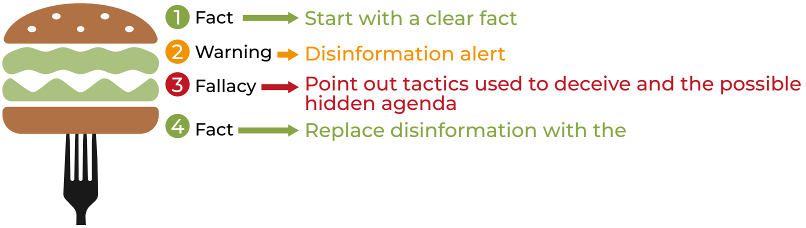
Defining the means

- How are you going to distribute the information?
- Which platforms are used by your target audience?
- What type of content (images, videos, graphs, interactive content) can best fit the platform?
- Can you create links between platforms? How can you design the content to make it more shareable on the platforms (e.g., keeping it simple and making it visually attractive)?

Figure 65: Guiding questions to define the means to debunk false information.
Source: UNICRI.

Execution phase

Different models have been developed to debunk disinformation.[86] One of the most effective techniques, developed by the linguist George Lakoff, is called the “truth sandwich”. The truth sandwich technique is composed of four elements:



1 Fact → Start with a clear fact
2 Warning → Disinformation alert
3 Fallacy → Point out tactics used to deceive and the possible hidden agenda
4 Fact → Replace disinformation with the

Figure 66: Example of the “truth sandwich” strategy for pre-bunking and debunking. Linguist George Lakoff developed the idea.
Source: UNICRI.

1 Fact → Start with a clear fact

Start with the facts that support the verified information: The debunking should begin with the **verified factual information** stated in a simple way. The information should not be complex and should have explanatory relevance. The prominence of the factual information could also be reinforced with a supporting headline or title. The fact should not be

[86] Garcia, L., & Shane, T. (2021). A guide to prebunking: a promising way to inoculate against misinformation in *First Draft*. Available on the Internet. Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. Available on the Internet. Cook, J., Lewandowsky, S. (2011). *The Debunking Handbook*. Available on the Internet.

written in a negative form or rely on a simple retraction (e.g., “this claim is not true”).[87]

2 Warning → Disinformation alert

Warning and addressing the false claim: The second element in the truth sandwich consists of an explicit warning that the information, visual, or audio content that is about to be presented is false.[88] When addressing disinformation, it is important to repeat the false information only once, directly before the correction. Unnecessary repetitions should be avoided to minimize the risk of the backfire effect or generating familiarity with the false claim. In addition, debunking can be more effective if there is an explanation as to why the source of disinformation is not credible and what is the real intent of the disinformation (e.g., undermining trust in a government or making profits).

3 Fallacy → Point out tactics used to deceive and the possible hidden agenda

Explain the fallacy: An explanation of the false claim and why it is wrong is the third element in a truth sandwich. Corrections should be presented in contrast with the mistaken information to ensure a clear rebuttal. It should be virtually impossible for the audience to ignore or overlook the corrective element, even when skimming the information. This can be achieved by using simple and concise language when explaining, as

[87] Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020.* Available on the Internet.

[88] Cook, J., Lewandowsky, S. (2011). *The Debunking Handbook*. Available on the Internet.

well as graphic elements to help highlight the facts, such as a larger font or different colours. Details should also be provided about why it is important to correct the false claim, why it becomes clear after debunking that the claim was false, and why the alternative explanation is correct. Demonstrating inconsistencies in disinformation can protect the audience from returning to the pre-correction beliefs. The credibility and background of the person or organization that is performing the debunking is another important element to highlight when successfully debunking a claim. Using credible sources will also help to generate more persuasive responses. To ensure that the reader pays attention to the sources, the message should be tailored to the specific target groups.[89]

4 Fact → Replace disinformation with the fact

Restate the fact: Replace disinformation with the facts at the end of the process, making sure that there is no gap in the story.[90] In most cases, by providing a fact that fills a "gap" in an explanation, debunking becomes easier because the fact can replace the inaccurate information in an individual's initial understanding with a new version of what happened.[91]

When designing debunking strategies, the following additional aspects can be considered:[92]

- **Start with a headline**: The headline should be the fact, presented in a clear and short way. It should not, however, be a negative sentence.

[89] ibid

[90] ibid

[91] Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020.* Available on the Internet.

[92] ibid

- **Add some detail, but not too much**: Explain why the claim is false without overwhelming the audience. Aim to increase belief and provide people with counterarguments that will allow them to debunk the claim when they encounter it.

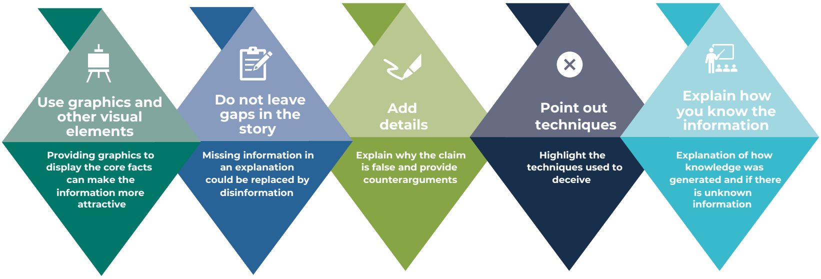
- **Briefly point out the techniques and tactics that are being used**: When correcting a false claim, remind the audience that this tactic is not exclusive to this example.

- **Explain how the debunkers know the information they are sharing and what is still unknown**: Provide an explanation of how knowledge was acquired helps build trust. This process can give the audience more clarity when analysing the conflict between fact and myth and allows them to gain the tools needed to reject the claim easily in the future. In addition, by explaining that there is currently unknown information, the audience will be prewarned that the facts might change as the situation develops. In this case, when providing information, it might be useful to centre on the current consensus, reminding people what experts agree on and why.

- **Do not leave gaps in the story:** Missing information in an explanation could be replaced by disinformation. This could lead to investing resources into debunking the same false claim multiple times to clarify the information, possibly leading to confusion since the false claim would have been addressed multiple times with different facts in the debunking efforts.

- **Keep the information short and simple (KISS principle)**: When debunking it is important to prioritize clarity, conciseness, and consistency. This can decrease the probability of the backfire effect occurring or the target audience becoming confused.[93]

- **Use graphics and other visual elements**: It might be useful to provide graphics to display the core facts since the information will look more attractive to both read and reshare on social media platforms. This can also make it easier to understand the information as it would need to be kept short and simple.



Use graphics and other visual elements
Providing graphics to display the core facts can make the information more attractive
Do not leave gaps in the story
Missing information in an explanation could be replaced by disinformation
Add details
Explain why the claim is false and provide counterarguments
Point out techniques
Highlight the techniques used to deceive
Explain how you know the information
Explanation of how knowledge was generated and if there is unknown information

Figure 67: Aspects to consider when pre-bunking and debunking a false claim.
Source: UNICRI.

Figures 68 and 69 show two examples of how to apply the truth sandwich when addressing disinformation. In the examples, debunking begins with stating the fact rather than the myth, it is followed by a warning that clarifies that the information that is about to be presented is false, and it continues by mentioning the fallacy and explaining why it is wrong. Finally, the examples end with a fact. Each example of the truth sandwich structure is followed by examples of possible posts that show how the debunking information can be presented in a graphic manner.

[93] Changing Minds. (n.d.). Keep it short and simple (KISS principle). *Changing Minds Organisation.* Available on the Internet.

1 **Fact**

The American Red Cross accepts blood donations from people vaccinated against COVID-19.

Start with the facts that support the verified information

2 **Warning**

A **false claim** has been circulating in several online posts in social media.

Add explicit warnings that the content about to be presented is false

3 **Fallacy**

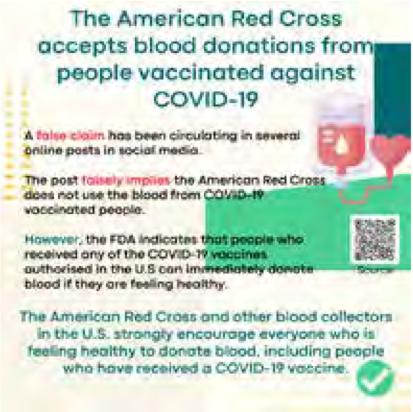
The **post falsely implies** the American Red Cross does not use the blood from COVID-19 vaccinated people. **However**, the FDA indicates that people who received any of the COVID-19 vaccines authorized in the U.S can immediately donate blood if they are feeling healthy.

Explain what is the false claim and why it is wrong

4 **Fact**

The American Red Cross and other blood collectors in the U.S. strongly encourage everyone who is feeling healthy to donate blood, including people who have received a COVID-19 vaccine.

Replace disinformation with facts at the end of the process



The American Red Cross accepts blood donations from people vaccinated against COVID-19

A false claim has been circulating in several online posts in social media.

The post falsely implies the American Red Cross does not use the blood from COVID-19 vaccinated people.

However, the FDA indicates that people who received any of the COVID-19 vaccines authorised in the U.S can immediately donate blood if they are feeling healthy.

The American Red Cross and other blood collectors in the U.S. strongly encourage everyone who is feeling healthy to donate blood, including people who have received a COVID-19 vaccine.

Figure 68: Example of the debunking of a false claim related to the role of an organization (American Red Cross) during the COVID-19 pandemic. The example is based on a false claim that fact-checkers debunked.[94]
Source: UNICRI.

[94] Jaramillo, C. (2022). Red Cross accepts blood donations from people vaccinated against COVID-19. *FactCheck.org*. Available on the Internet.

1 Fact

Monkeypox is an animal-to-human (zoonotic) transmitted disease.

Start with the facts that support the verified information

2 Warning

A **false claim** has been circulating in a video disseminated in different social media platforms.

Add explicit warnings that the content about to be presented is false

3 Fallacy

The **post falsely implies** that monkeypox is biological warfare being unleashed onto the public by the WHO, IMF and Bill Gates. **However**, the monkeypox outbreak was <u>not</u> caused by biological warfare.

Explain what is the false claim and why it is wrong

4 Fact

The disease was discovered in 1958 in monkeys, and the first human case was recorded in 1970. There have been several outbreaks in humans, none related to biological warfare.

Replace disinformation with facts at the end of the process

Monkeypox is an animal-to-human (zoonotic) transmitted disease

A false claim has been circulating in a video disseminated in different social media platforms.

The post falsely implies that monkeypox is biological warfare being unleashed onto the public by the WHO, IMF and Bill Gates.

The monkeypox outbreak was not caused by biological warfare. There have been several outbreaks in humans, none related to biological warfare.

Figure 69: Example of the debunking of a false claim related to the role of different actors during the 2022 outbreak of monkeypox. The example is based on a false claim that fact-checkers debunked.[95]

Source: UNICRI.

[95] Rahman, G. (2022). No evidence monkeypox is an agent of biological warfare. *Full Fact*. Available on the Internet.

Annexes

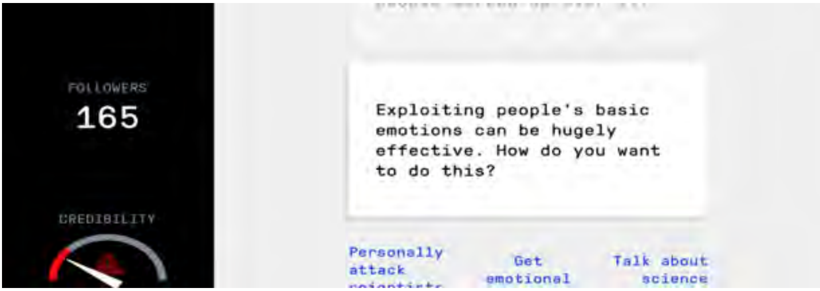
Several technology tools have been developed to support the efforts of pre-bunking and debunking. These tools are based on different approaches, such as gamification or enhancing media literacy skills. The list of tools presented is not comprehensive. However, it aims to provide an overview of some of the existing technologies that can facilitate debunking or the understanding of the phenomena of disinformation.

Annex 1: Technology tools for pre-bunking

Different technology tools help to develop media literacy skills, including familiarity with pre-bunking and debunking. Since pre-bunking is done before the disinformation is spread, the tools centre on improving the skills of the person, particularly identifying disinformation techniques and the strategies of actors that adopt the techniques.

Some technology tools use a gamification[96] approach to encourage people to actively practice their pre-bunking skills:

Bad News: This is an online game that seeks to inoculate players against fake news across different cultures by focusing on pre-bunking misinformation and disinformation techniques. Users build an understanding of the techniques used to disseminate disinformation. The game exposes players to common fake news tactics by making them a fake news tycoon. Players win by publishing headlines that attract the most followers.[97]



Exploiting people's basic emotions can be hugely effective. How do you want to do this?

Personally attack

Get emotional

Talk about science

Go Viral: This is a game based on Bad News but focuses on COVID-19 misinformation.[98]

Nope, Miguel's whale fact is actually true! Distinguishing fact from fiction on social media can be hard. Aside from fact-checking, you can spot dodgy content by checking whether it uses manipulation tactics like fearmongering, fake experts or conspiracy theories. Which Miguel's post did not.

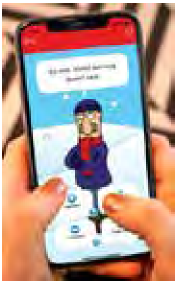
SWEET, THANKS, BACK TO MY FEED

[96] "The process of adding games or gamelike elements to something (such as a task) so as to encourage participation". Definition obtained from: Merriam-Webster. (n.d.). Gamification definition & meaning. *Merriam-Webster*. Available on the Internet.

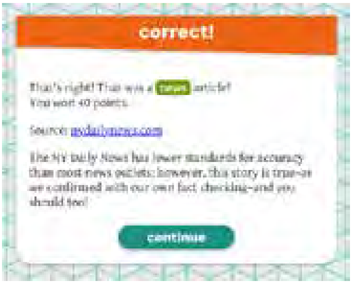
[97] https://www.getbadnews.com/#intro

[98] https://www.goviralgame.com/en/play

Cranky Uncle: This game uses humour and critical thinking to expose the misleading techniques of science denial and build public resilience against misinformation. In the game, players are mentored by a cartoon personification of a climate science denier, who explains 14 science denial techniques (including fake experts, cherry picking, and different logical fallacies).[99]



Factitious-Pandemic Edition: This is a game designed to sharpen skills for spotting fake news. It includes the source, which helps individuals practise reliable source identification skills.[100]



correct!

That's right! That was a news article!
You won 40 points.

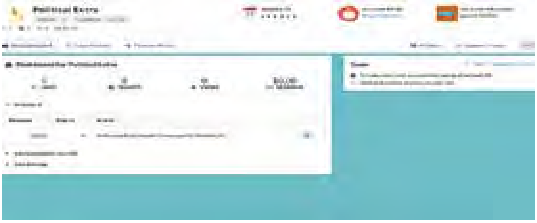
Source: nydailynews.com

The NY Daily News has lower standards for accuracy than most news outlets; however, this story is true–as we confirmed with our own fact checking–and you should too!

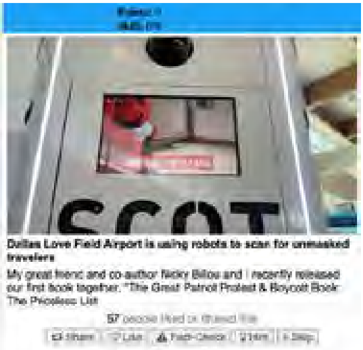
[99] https://crankyuncle.com/game/

[100] http://factitious-pandemic.augamestudio.com/#/

Fake It to Make It: The user has to create a character, goal and website, and to achieve the goal set, the person must create fake news. The game is based on the premise that by making players more aware of the process of creating and distributing fake news, they will be more sceptical of the information they encounter in the future.[101]



Fakey: This is a simulated news feed where the user has to analyse the various stories, with the end goal of teaching media literacy and studying how people interact with misinformation.[102]



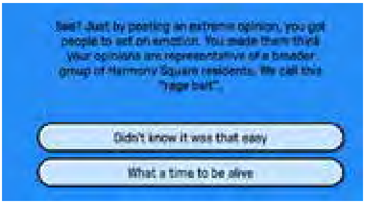
Dallas Love Field Airport is using robots to scan for unmasked travelers

My great friend and co-author Nicky Billou and I recently released our first book together, "The Great Patriot Protest & Boycott Book: The Priceless List

57 people liked or shared this

[101] https://www.fakeittomakeitgame.com/
[102] https://fakey.osome.iu.edu

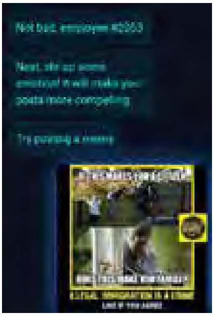
Harmony Square: This game is designed to expose the multiple tactics and manipulation techniques that are used to mislead people, build up a following, or exploit societal tensions for political purposes. It intends to work as a psychological "vaccine" against disinformation by building cognitive resistance. The game is set in a neighbourhood obsessed with democracy, where the player is hired as Chief Disinformation Officer whose job is to foment internal divisions and pit its residents against each other.[103]



See? Just by posting an extreme opinion, you got people to act on emotion. You made them think your opinions are representative of a broader group of Harmony Square residents. We call this "rage bait".

Troll Factory: This game demonstrates how information operations work on social media with the objective of illustrating how fake news, emotive content and bot armies are utilized to affect moods, opinions, and decision-making by using examples of authentic social media content.[104]

Other approaches include the use of courses, guides, infographics, and other material to provide individuals with the tools to pre-bunk false claims.



Not bad, employee #2363

Next, stir up some emotion! It will make your posts more compelling.

Try posting a meme

[103] https://harmonysquare.game/en

[104] https://trollfactory.yle.fi/

First Draft - Learn how to spot vaccine misinformation and myths: This course consists of 20-minute sessions that focus on skills for researching, monitoring, verification, and more.[105] The videos include study companion guides. The organization offers other tools, such as an online challenge to verify content,[106] multiple guides for better online journalism,[107] and a dashboard with tools for verification and responsible reporting.[108]

How to use CrowdTangle's public Coronavirus Live Displays

How to use Google Trends' Coronavirus Dashboard

Covid-19 CrowdTangle Live Display

How to find Twitter Lists using advanced Google search

The International Center for Journalists (ICFJ) Resources for Journalists: This includes 'Covering COVID-19: Resources for Journalists', a resource which aims to provide tools for learning from health professionals and other experts. It gives access to webinars on different topics, including health disinformation, newsroom leadership, and the latest research on COVID-19. This tool also includes new resources on covering COVID-19, journalism tips, trends and opportunities, and the possibility to collaborate with others.[109]

[105] https://firstdraftnews.org/vaccine-insights-flexible-learning-course/

[106] https://ftp.firstdraftnews.org/articulate/2020/en/OVC/story_html5.html

[107] https://firstdraftnews.org/long-form-article/first-drafts-essential-guide-to/

[108] https://start.me/p/vjv80b/first-draft-basic-toolkit

[109] https://www.icfj.org/resources

Annex 2: Technology tools for debunking

As mentioned in the case of pre-bunking, technology can provide useful tools to make debunking more effective. These tools can be used during the three steps previously described for debunking (analysis of disinformation, making a decision and executing the strategy). Some of these tools include:

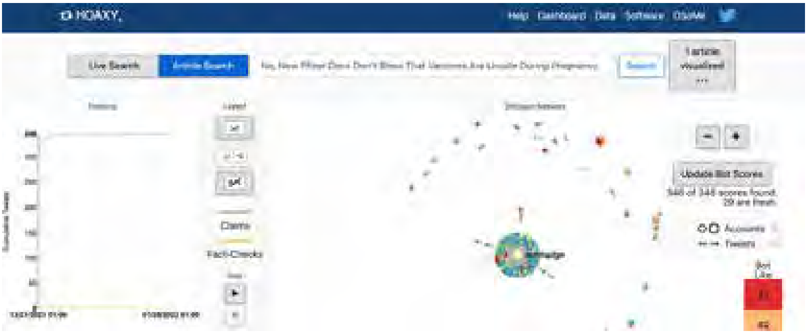
Analysis of disinformation

Detection of bots: In most cases, tools can automatically detect if the account is a bot. For example, Bot Sentinel[110] uses Artificial Intelligence to examine Twitter accounts and classify them as trustworthy or untrustworthy in an effort to help users identify bots. Identified bot accounts are kept in a database to track their daily activity. The data collected can be used to explore the effect of bot propaganda on discourse and to find ways to counter the spread of the disinformation they disseminate.

Another similar tool is Botometer,[111] which uses machine learning to classify how close Twitter accounts are to a bot. It analyses features of a profile such as friends, social network structure, temporal activity, language, and sentiment. Then, it gives the account an overall score that provides a measure of the likelihood that it is a bot.

Hoaxy[112] is a tool that searches for claims, tracking the sharing of links to stories from low-credibility sources and independent fact-checking organizations. It also calculates a bot score, which is a measure of the likely level of automation.

[110] https://botsentinel.com/info/about
[111] https://botometer.osome.iu.edu
[112] https://hoaxy.osome.iu.edu



Interface of Hoaxy

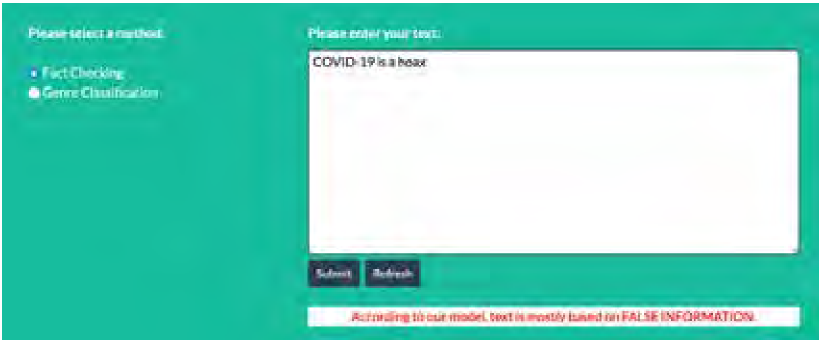
Fact-checking websites and tools: ClaimBuster[113] is a web-based, live fact-checking tool that uses Artificial Intelligence (natural language processing and other supervised learning techniques) to identify factual and false information. Another tool is SciCheck,[114] which is part of FactCheck.org and focuses on false and misleading scientific claims.

Other examples include Lead Stories,[115] a web-based fact-checking platform that can point out false or misleading stories, rumours, and conspiracies. Its Trendolizer™ engine indexes links from different Internet sources and then measures their engagement rate to identify trending content. This content is then fact-checked by their team of journalists.

Fake News Detection[116] is a tool that collected news articles with veracity labels from fact-checking websites and used them to train text classification systems. You can paste a piece of text and examine its similarity to their collection of true versus false news articles.

[113] https://idir.uta.edu/claimbuster/
[114] https://www.factcheck.org/scicheck/
[115] https://leadstories.com/how-we-work.html
[116] http://fakenews.research.sfu.ca/#footer

Some other examples include The Vaccines Insights Hub,[117] which provides a dashboard with live insights, intelligence and reporting guidance on emerging health and vaccine misinformation, and the Google Fact Check Tools[118] (Fact Check Explorer and Fact Check Markup Tool) that aim to facilitate the work of fact checkers, journalists, and researchers.



Please select a method:

● Fact Checking
○ Genre Classification

Please enter your text:

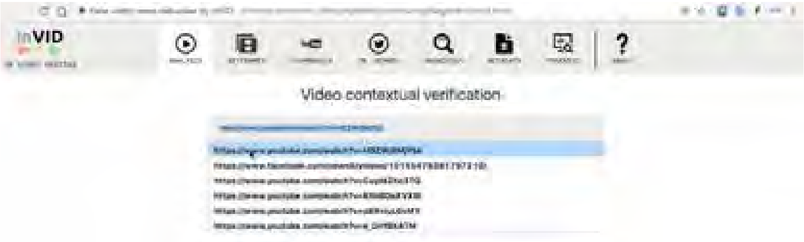
COVID-19 is a hoax

According to our model, text is mostly based on FALSE INFORMATION.

Fake News Detection interface

Website and source ratings: Other tools include browser extensions that evaluate the veracity of content while the individual is looking at the information online, such as NewsGuard's HealthGuard[119] or Our.News,[120] which use a label to provide information about the reliability of the sources. Media Bias/Fact Check, FakerFact, TrustServista, Check, and TrustedNews are other options.[121] Another example is the Video Verification Plugin (InVid)[122] that obtains contextual information and verifies content on social networks.

[117] https://firstdraftnews.org/vaccineinsights/
[118] https://toolbox.google.com/factcheck/explorer
[119] https://www.newsguardtech.com/solutions/healthguard/
[120] https://our.news/how-it-works/?main
[121] https://thetrustedweb.org/browser-extensions-to-detect-and-avoid-fake-news/
[122] https://www.invid-project.eu/tools-and-services/invid-verification-plugin/



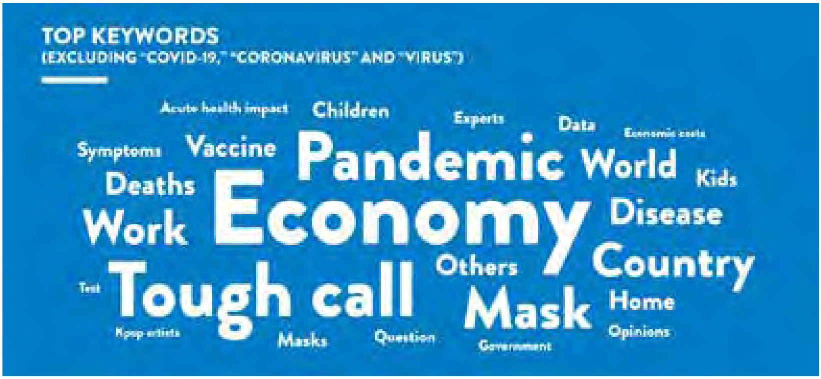
InVid interface

Selecting a topic

Social listening with Artificial Intelligence: When selecting a topic to debunk during the execution phase, social listening can be used to identify which topics are currently being discussed by the most people. This can help to focus efforts on popular topics, while frequently obtaining an updated analysis of what the relevant topics are.

Technology is now being used by WHO to adopt a social media listening strategy to identify the main health-related topics that are being discussed online. This is done by using machine learning to analyse pieces of information on various social media platforms. Searches are then conducted based on taxonomy so that information can be categorized into topics.[123] Machine learning can also obtain insights into the kind of emotions users are experiencing. Rather than just dividing the data by type of sentiment (positive, neutral, negative), language analytics can analyse anxiety, sadness, denial, acceptance, and other emotions expressed in social media posts. This information can develop an effective offensive strategy and assuage the public's concerns before misinformation can gain steam.

[123] WHO (2020). Immunizing the public against misinformation. *World Health Organization*. Available on the Internet.

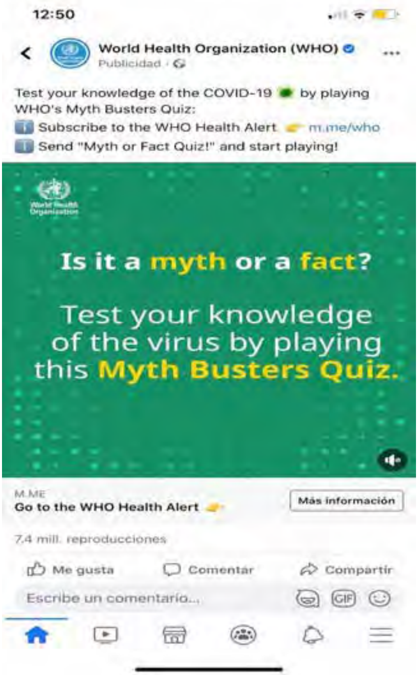


Top public health topics for 9-15 July identified by the WHO tool.
Source: WHO.

Practising and acquiring skills

Gamification approach: As with the tools for pre-bunking, debunking can also be more interactive. Captain Fact[124] is a website based on collective moderation. It employs a gamification approach that works with a video overlay (thanks to a web browser extension) that adds sources and context to videos viewed on the Internet. The text overlay on the video means that it even functions as a debate platform, as users can challenge what is being said, point by point. The WHO Myth Busters Quiz[125] also uses this gamification approach by allowing the users to test their knowledge about popular COVID-19 myths and facts.

[124] https://captainfact.io/
[125] https://www.facebook.com/watch/ref=saved&v=433416304464216



12:50

World Health Organization (WHO)
Publicidad ·

Test your knowledge of the COVID-19 🦠 by playing WHO's Myth Busters Quiz:
ℹ️ Subscribe to the WHO Health Alert 👉 m.me/who
ℹ️ Send "Myth or Fact Quiz!" and start playing!

World Health Organization

Is it a myth or a fact?

Test your knowledge of the virus by playing this Myth Busters Quiz.

M.ME
Go to the WHO Health Alert 👉

Más información

7,4 mill. reproducciones

WHO Myth Busters Quiz interface

Other initiatives: Other efforts have been made to improve debunking efforts, such as the code of principles of the International Fact-Checking Network[126] for organizations that report on the accuracy of statements or other widely circulated claims related to public interest issues. The WHO's framework for managing infodemics[127] proposes a framework based on skillsets to prioritize and problem solve the issue of excessive and inaccurate information about the COVID-19 infodemic.

[126] https://www.poynter.org/ifcn-fact-checkers-code-of-principles/

[127] https://www.who.int/publications/i/item/9789240010314