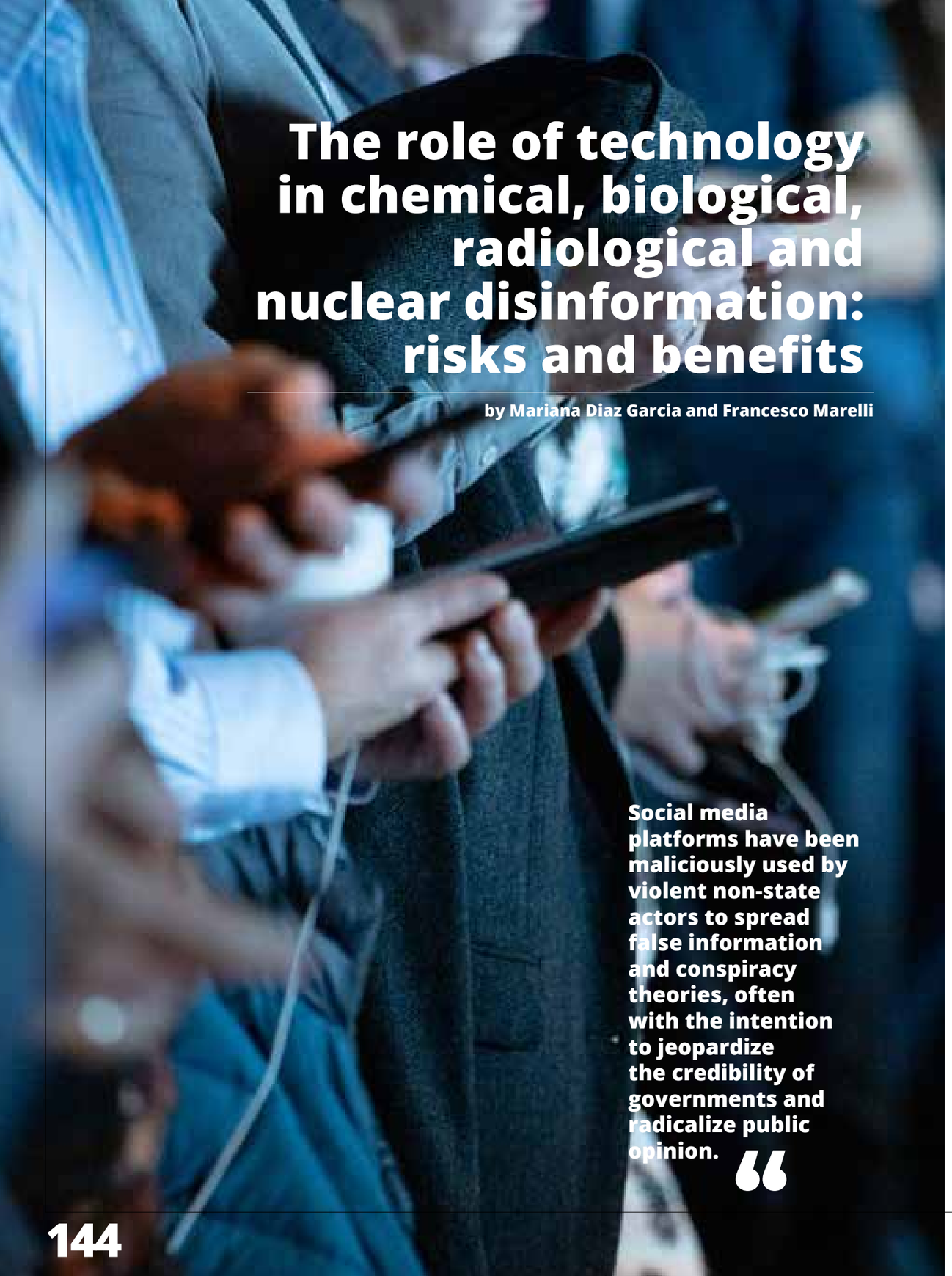
The role of technology in chemical, biological, radiological and nuclear disinformation: risks and benefits

by Mariana Diaz Garcia and Francesco Marelli



Social media platforms have been maliciously used by violent non-state actors to spread false information and conspiracy theories, often with the intention to jeopardize the credibility of governments and radicalize public opinion.

Technology plays a central role in the area of chemical, biological, radiological and nuclear (CBRN) disinformation.[1] Social media platforms have been maliciously used by violent non-state actors to spread false information and conspiracy theories, often with the intention to jeopardize the credibility of governments and radicalize public opinion. It could even be said that social media platforms have changed the "rules of the game" in the history of disinformation. 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 started 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 and their ability to manipulate information has introduced a new variable that significantly amplifies the spread of disinformation.

in recent years, terrorists, violent extremists, and organized criminal groups have started to exploit vulnerabilities in the social media ecosystem

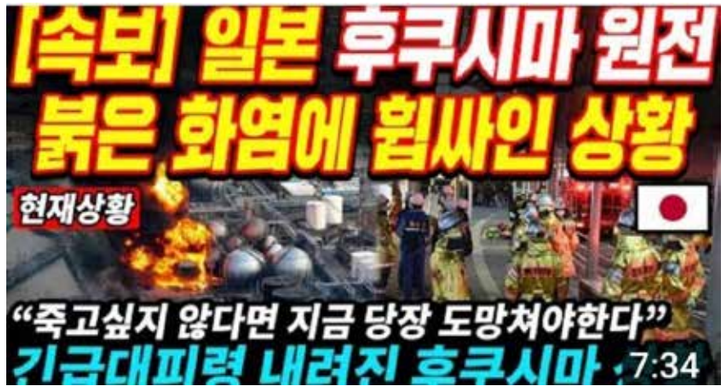
At the same time, technology offers innovative solutions to detect and respond to disinformation. There are a variety of technology solutions that can be used easily and often at no cost to ascertain whether the information or claim is correct, to analyse a website and determine if and to what extent the source is reliable, to verify the contents of photos and videos and so on.

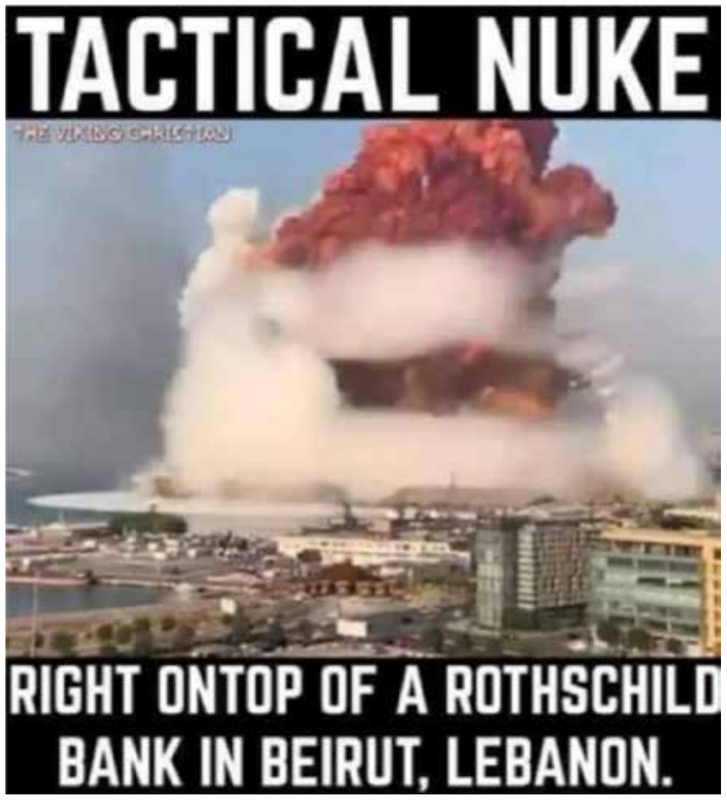
The global exchange of content in real-time on social media has modified citizens' behaviours regarding news consumption.

This article analyses some of the technology risks and advantages related to disinformation. Let us start with the risks by clarifying that technology is not inherently bad or harmful. The development of new digital platforms has 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, including the proliferation of false information and conspiracy theories.

Manipulating data and misleading the public on social media is rather simple. For example, forging an official letter by a United Nations organization is relatively simple considering that, most likely, the logo and even the signature of the Director General of that organization can be found and downloaded through a search engine. Manipulating

1 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. CBRN disinformation has become a significant problem in the last few years. False information and conspiracy theories on CBRN risks, such as exposure to toxic chemicals, infectious disease outbreaks or theft of radioactive material, can cause confusion and mistrust in governments and even jeopardize the public health response in case of emergency. To find out more see UNICRI *Handbook to combat CBRN disinformation* available at https://unicri.it/Publication/Hanbook-to-combat-disinformation.





photos and videos is also relatively simple. For example, someone can mislead users of social media by posting an original picture of a fire at a Japanese oil refinery next to the Fukushima Daiichi Nuclear Plant (after the terrible earthquake of 11 March 2011) and adding the sentence "[Breaking News] Japan's Fukushima nuclear power plant swept up in a red blaze" as if the event were taking place now.[2]

By mastering complex techniques, it is possible to break into a web server and replace a hosted website with a completely different one (website defacement). It is even possible to generate videos or photographs that misrepresent people by generating images that are nearly indistinguishable from the original (deepfake video). Artificial intelligence techniques can also produce fake news reports, including realistic video and audio, to influence public opinion, impact political campaigns and erode trust in government (e.g., in the area of vaccines).

Social media platforms can be also used maliciously to create echo-chambers, that are virtual environments 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. Virtual echo-chambers have been used, for example, by far-right groups to spread conspiracy theories related to the origin of COVID-19 and the immunization campaigns.

Another example is represented by social media algorithms. Although designed to support users to identify what might be more interesting for them and avoid potentially irrelevant or low-quality content, social media algorithms can also be used to link a video with extremist views to other similar videos or facilitate interaction between users with extremist views.

> **Viral online and sometimes physical attacks have been conducted against almost every stakeholder operating in the area of CBRN risk mitigation**

Today, the malicious use of social media platforms 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 policy-makers, managers of institutions, researchers from universities and research centres, spokespersons from different government departments, in particular the public health sector, journalists and representatives from international organizations. The Russian Federation's invasion of Ukraine has further reinforced this trend, with a massive disinformation campaign that, together with cyberattacks, has targeted infrastructures of Black Sea countries, starting with Ukraine. In this respect, the misuse of social media has become a serious new challenge.

Having said that, we can now analyse the other side of the coin: the advantages brought by technology to combat disinformation. To begin with, there are simple tools that can support the analysis of sources. These include the use of web browser extensions (e.g., WeVerify, CrowdTangle, NewsGuard) that automatically analyse information and images in websites (e.g., Google Image Search or Tineye) and determine if and to what extent the source is reliable. Other tools can confirm that the visual content is correctly attributed to the original source or understand the context in which the image has been used.

There are also tools that help users to verify the recording and upload time of video content (e.g., YouTube Data Viewer). Other tools can be used for the geolocation of photos and videos (e.g., Wikimapia or Google Earth). Some software has also been developed with a gamification approach to practise pre-bunking and debunking skills (e.g., Fake It to Make It, Bad News, Harmony Square, WHO Myth Busters Quiz, Captain Fact).

> **Artificial intelligence has allowed for the creation of tools that address specific issues in disinformation, like the spread of deepfakes,**

Emerging technologies have also been used in the efforts to combat disinformation. Artificial intelligence has allowed for the creation of tools that

2 See UNICRI *Handbook to combat CBRN disinformation*.

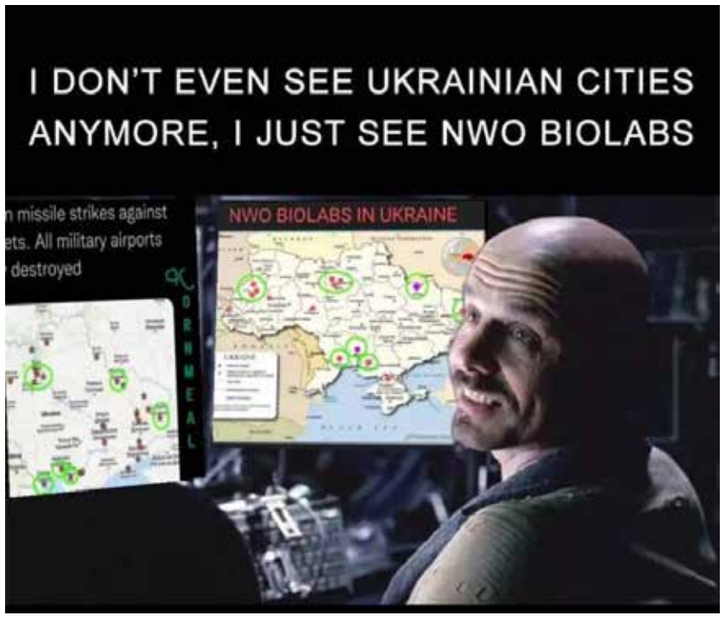
address specific issues in disinformation, like the spread of deepfakes, which include manipulated photos, videos, or audio files. Some tools (e.g., Microsoft Video Authenticator) can analyse a still photo or video frame and provide a percentage chance that the content is artificially manipulated. In some cases, artificial intelligence and big data visualization have been used to automatically detect if a Twitter account is a bot (e.g., Bot Sentinel).

Social listening with artificial intelligence has also been used by the World Health Organisation to more accurately target their campaigns against misinformation. Social listening can be used to identify which topics are currently being discussed by the most people. By frequently obtaining an updated analysis of what the popular topics are, researchers can identify relevant health-related topics and focus their efforts on these. This is done by using machine learning to analyse pieces of information on various social media platforms.

Machine learning can also obtain insights into the kind of emotions users are experiencing

Machine learning can also obtain insights into the kind of emotions users are experiencing. 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.

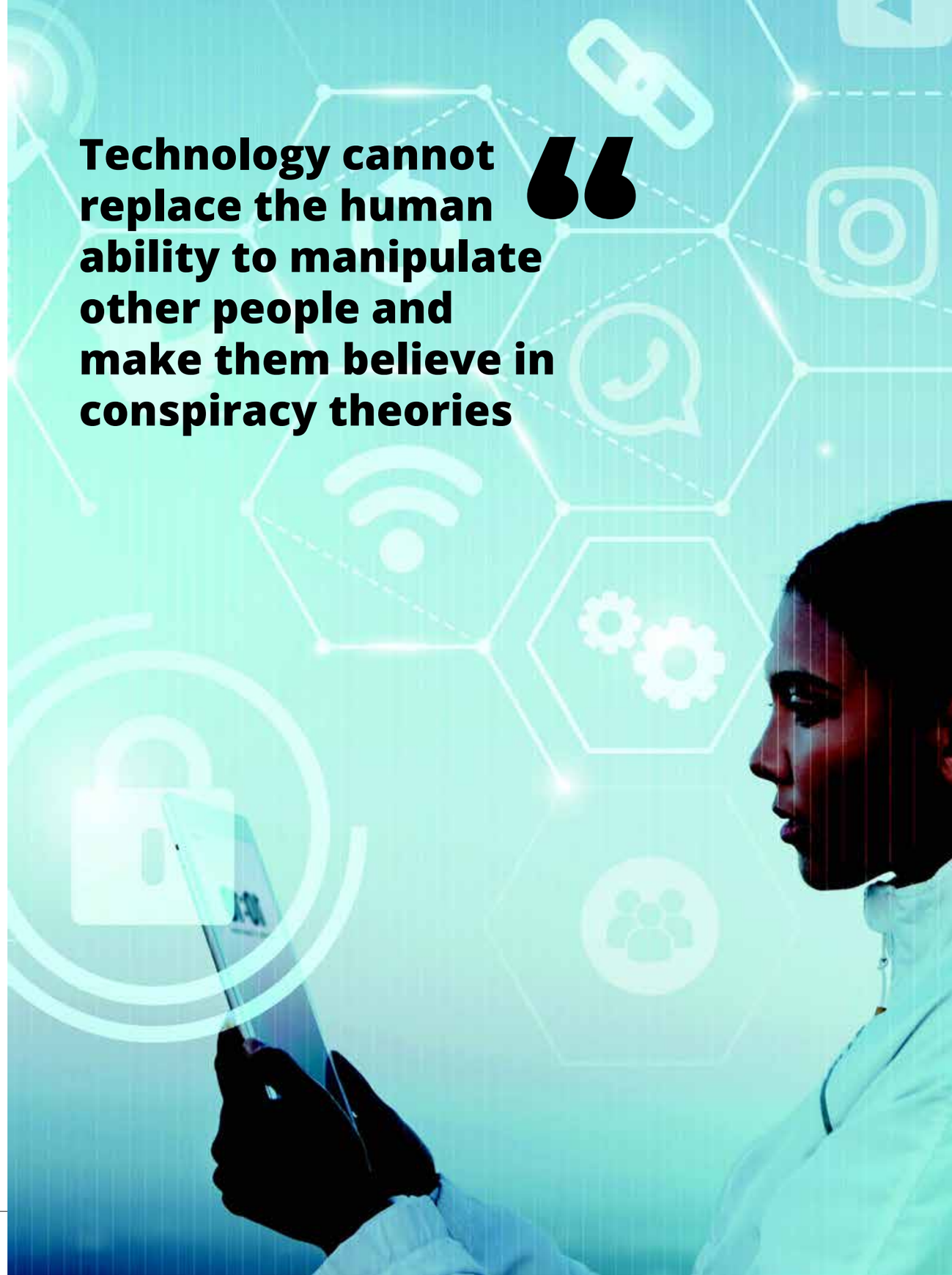
To summarize, the process of designing and employing new technologies for good or bad purposes has become so fast that almost every new technology, such as ChatGPT (launched in November 2022), can be used by opponent parts to spread or combat disinformation in a sort of technological race.



Is there still a role for human skills in such a fast-paced technological race?

We think that the answer is yes. To explain why, we can borrow a famous assertion from the philosopher Michael Polanyi of 1958: "we can know more than we can tell"[3]. It means that human beings possess information (tacit knowledge) that is difficult to transfer to others, including

Technology cannot replace the human ability to manipulate other people and make them believe in conspiracy theories



3 Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy,* (1958).

to an artificial intelligence. This includes personal or professional experience or abstract concepts such as intuition. An example is beauty: everybody knows what beauty is, but we often find it difficult to explain.

In this sense, technology, including artificial intelligence and big data, can assist us to design and spread conspiracy theories or, alternatively, to detect disinformation and explain why a conspiracy theory is false, but it cannot replace us in *convincing* another person that a conspiracy theory is true or false. Technology cannot replace the human ability to manipulate other people and make them believe in conspiracy theories. Equally, technology cannot replace human skills when evaluating the veracity of information, judging a situation and convincing a human being why a certain piece of information is wrong or why an echo-chamber does not help you to grow your knowledge and judgment.

If this is true (and we believe it is), we can still empower people using social media so that they can make their own informed decisions about what is true and what is not.

ABOUT THE AUTHORS

Francesco Marelli has been working for the United Nations Interregional Crime and Justice Research Institute (UNICRI) since 2003. As Head of the CBRN Risk Mitigation and Security Governance Unit, he is the coordinator of UNICRI's activities in the area of CBRN risk mitigation, which include the CONTACT programme to strengthen Member States' capacities to prevent illicit trafficking of radiological and nuclear materials, and the implementation of the European Union CBRN Centres of Excellence, a network-based initiative that supports more than 60 countries in strengthening their national CBRN policies and capabilities.

He is also responsible for the Knowledge Centre on Security through Research, Technology and Innovation (SIRIO) in Geneva (Switzerland), which assesses emerging risks and identifies, tests and promotes innovative solutions to reduce the risk of crime, including in the field of technologies such as big data, biotechnology and blockchain.

He received his PhD from the School of History at the University of Leeds in 2002. He is the author of several publications.

Mariana Diaz Garcia is a Fellow at the United Nations Interregional Crime and Justice Research Institute (UNICRI), in the Knowledge Centre on Security through Research, Technology and Innovation (SIRIO) and the CBRN Risk Mitigation and Security Governance Unit. She also collaborates with UNICRI's Centre to Combat Disinformation in Geneva. Her research focuses on the malicious use of social media by extremist groups; and technology strategies to improve chemical, biological, radiological and nuclear defense. Her research also focuses on the ideology, radicalisation and legitimisation of armed groups.