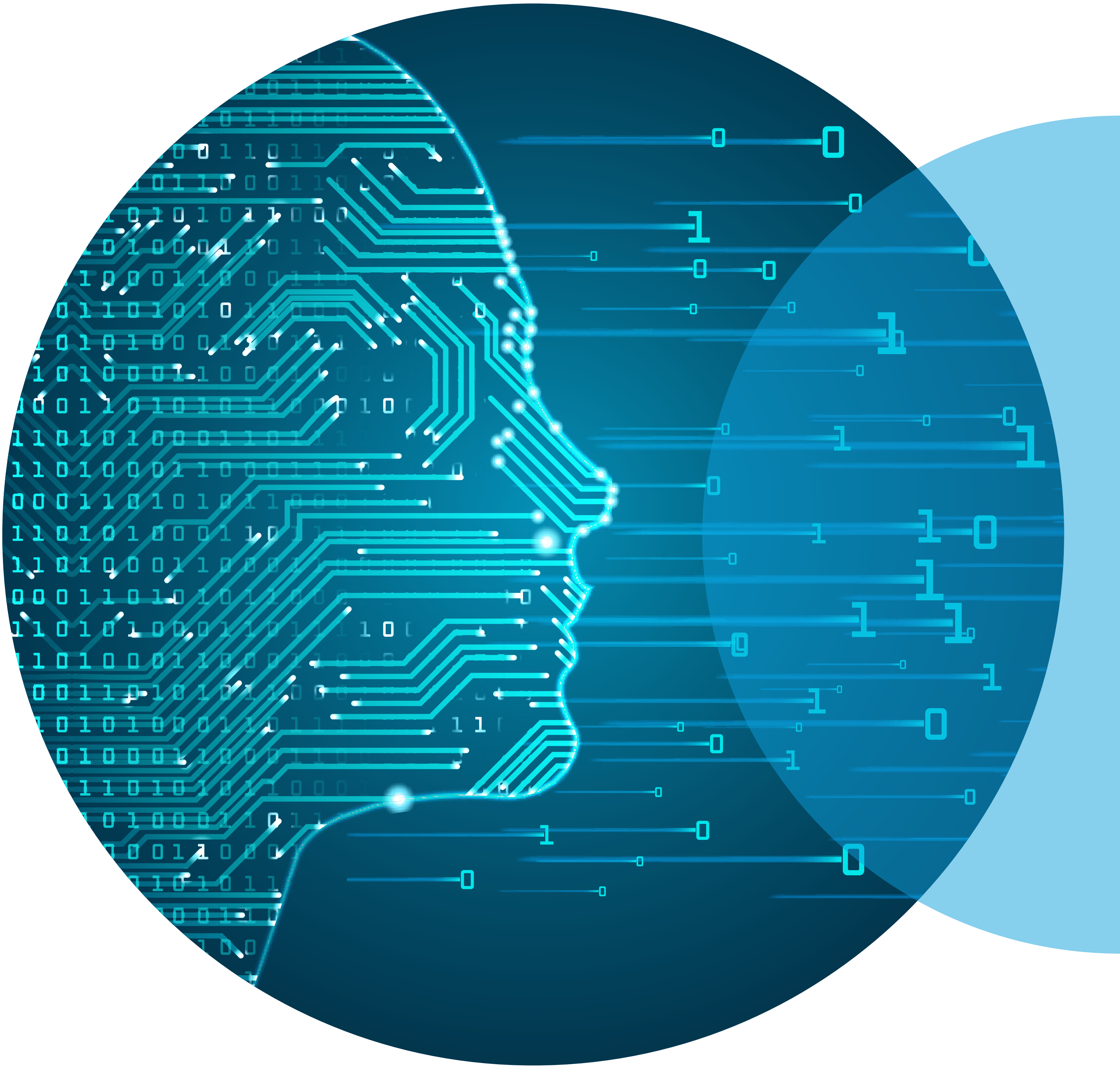
Hybrid CoE Paper 14

AI-based technologies in hybrid conflict: The future of influence operations



Hybrid CoE

Nicolas Mazzucchi – June 2022

Hybrid CoE Papers are finalized pieces of analysis on a topic related to hybrid threats, based on one or several research questions. They may be either conceptual analyses or based on a concrete case study with empirical data.

The European Centre of Excellence for Countering Hybrid Threats
tel. +358 400 253800 www.hybridcoe.fi

ISBN (web) 978–952–7472–42–2
ISBN (print) 978–952–7472–43–9
ISSN 2670–2053

June 2022

Cover photo: Ryzhi / shutterstock.com

Hybrid CoE's mission is to strengthen its Participating States' security by providing expertise and training for countering hybrid threats, and by enhancing EU-NATO cooperation in this respect. The Centre is an autonomous hub for practitioners and experts, located in Helsinki, Finland.

The responsibility for the views expressed ultimately rests with the authors.

Contents

Summary 5
Introduction 6
Influence operations in the digital era 7
Coping with technology 7
The gaps in the doctrine 7

The rise of AI-based technologies 10
From entertainment to malign intent 10
Beyond deepfakes 11

A perfect tool for hybrid warfare? 13
The war in Ukraine: the first deepfake-supported conflict? 14

Conclusions 16
Author 19

Summary

Influence operations have been part of modern military operations for decades. The use of information to intoxicate or manipulate the enemy, or even civilian populations, grew during the second half of the 20th century, especially in the context of asymmetric warfare, and proved to be a major contributor to non-state actors' political victories, as in Vietnam. Yet technology also brought about major changes with the acceleration of communication due to cyberspace and the global social networks that ensued. Technology has therefore been at the core of information warfare, allowing a wider audience to be reached in a quasi-immediate timeframe. Emerging information technologies, especially AI-based ones, could therefore initiate a new major evolution in military influence operations. With the possibility to generate fake individuals, fake videos and a false consensus over an issue, hybrid warfare may enter a new era. Moreover, lower ranked military powers and non-state actors could also benefit from increasingly easier access to these technologies, with properly trained personnel remaining the major hurdle to overcome, for the regular military and hybrid threats actors alike.

Introduction

Since the 1960s and 1970s, major armed forces have considered the information domain an important battlefield, complementing the actions in traditional physical domains such as land, sea and air. The importance of influence operations during World War II, including in the planning of Operation Overlord (the invasion of Normandy in 1944), with dedicated psychological and information operations such as *Bodyguard* or *Fortitude*, highlighted the potential major effects of such actions, which proved to be of particular interest, also in terms of their cost effectiveness. The main aim of these operations is to exert a direct impact on the psychology of a targeted audience – be it a homeland audience, an enemy army, a battleground civilian population, and so forth – by disseminating information.

In this regard, several tactics could be used, from the dissemination of forged information (disinformation) and the use of information to encourage a specific behaviour (manipulation and deception) to the paralysis of enemy decision-making systems by waging an intentionally misleading intellectual offensive (intoxication). These tactics are as old as warfare itself, yet they are rejuvenated from time to time with the evolution of information and communication technologies that allow faster and more extensive broadcasting, reaching new audiences and having stronger impacts.

Yet the digital revolution seems to be ushering in a new era in influence operations, with both the ability to create a new breed of forged influence products along with a rapid decrease in the cost of dedicated tools and systems. This situation could allow all kinds of actors, whether state or non-state, to achieve a high level of sophistication, especially in the context of hybrid threats. As information operations have always been a favoured instrument in asymmetric warfare, emerging technologies, notably artificial intelligence (AI), may accelerate their use in future conflicts.

The aim of this Hybrid CoE Paper is to explore the changing landscape of influence operations with the rise of emerging – largely AI-based – digital technologies that could provide both new content and new dissemination capabilities. The first part of the paper will consider the development of digital technologies and their use in influence operations since the end of the 20th century, as armed forces and state actors tried to cope with the use of digital tools by non-state actors, mainly for propaganda and recruitment purposes. In the second part, the paper will look at the new possibilities created by emerging digital technologies in terms of deepfakes, online community penetration, artificial buzz creation, and their possible use in influence operations. The issue of access to technology will also be explored as it is a major hurdle to overcome in terms of hybrid warfare. The last part of the paper will focus on the potential use of these technologies in the context of hybrid warfare, exemplified by the ongoing war in Ukraine.

Influence operations in the digital era

Armed forces had to adapt their influence doctrine and operations to the rise of digital technologies during the late 1990s to the 2000s. As exemplified by operations in Afghanistan and Iraq, they faced a dual challenge: controlling the communication on their own side with the rise of personal IT devices in the hands of their soldiers, and the fight on the digital battlefield against insurgents or armed terrorist groups that were ambitious to use these digital technologies to open up a new front. Western armed forces soon discovered that the global communication capabilities could become a particularly powerful battlefield that they had to dominate, as the main lesson learned was that digital technologies could have a significant impact on the theatre of war and beyond, at a relatively low cost.

Coping with technology

The need to cope with the evolution of technology is one of the major issues for modern military operations, especially considering the civilian-originated technological developments in the field of information, including massive digital communication enablers. They have already had a significant impact on military operations, as the first key operations of the 21st century, namely *Enduring Freedom* in Afghanistan and *Iraqi Freedom* in Iraq, highlighted the importance of digital technologies. The use of personal digital video recorders and social network accounts by Western soldiers engaged in these operations, and the rapid use of digital media communication by the Taliban and Iraqi insurgents in demonstrating their military capabilities against the Western coalitions, underlined the salience of the digital battlefield. The widespread use of personal IT-based technologies duly ushered in a new era in military communication and propaganda.

At issue here is the difficulty that armed forces have in coping with the acceleration of technology development, and the ease of access to the public regarding some of these technologies. Since the Vietnam War, Western forces have been aware that they need to act both at a local and a global level to win the war, and not only by purely kinetic means. The need to articulate strategic communications (stratcom)[1] to strengthen the homeland population's consensus on the military operations, and to influence operations on the battlefield as well as operational levels with regard to local populations and opposing forces, created an info-ops nexus with multiple levels of interaction. The digital era added a new layer of complexity with the difficulty of articulating not only in relation to space (homeland vs. battleground) but also time, as digital technologies allowed quasi-realtime communications that needed to be carried out along with the information operations planification process, which was sometimes cumbersome.[2]

The gaps in the doctrine

Regarding the evolution of the doctrines of Western armed forces (NATO, US, European) in terms of information operations, there is a traditional separation – and complementarity – between both strategic-operative-tactical levels and homeland-battleground levels.

[1] US DoD, *Principles of Strategic Communication*, August 2008.

[2] N. Mazzucchi, 'L'arme de l'information dans les conflits armés' [The information weapon in armed conflicts], in C. Marangé and M. Quessard (eds.), *Les guerres de l'information* [*Information wars*] (Paris: PUF, 2021).

After the Vietnam War, the lessons learned led to a reorientation of military doctrines during the 1970s and 1980s towards better integration of the mass media and their impact through the embedment of journalists, the use of professional-level strategic communications, and the need to consider information as a military domain in itself. These changes enabled the development of psy-ops (actions on the behaviours and mindset of an enemy or population) and info-ops (action through information and the media) capabilities in support of kinetic operations. From the 1980s on, information has been considered a value chain, articulating the emitter, the message, the medium, and the receiver. The success of any information operation is therefore linked to the ability to create consistency between all value chain elements. These considerations and changes in doctrine contributed to the success of the coalition in the 1991 Gulf War, resulting in global media dominance. Info-ops specialists could therefore construct their own war narrative by controlling both the media, though journalists' embedment in military units, and the message, through images and papers provided directly to the mass media.

This information dominance was short-lived, however. The Gulf War was a dissymmetric war, fought between states and regular armed forces. Hence, there was no need to engage in a deep analysis of the relationship between information warfare and asymmetry. Yet new technological developments gave rise to turmoil in the provision and consumption of global information. The interconnection between communication networks since the early 21st century has blurred the borders between territories by producing a new geography of communication with a theoretical quasi-immediate worldwide reach for any information. There was therefore hardly any possibility of exerting information dominance that could compare to the Gulf War era, with information provision capabilities becoming increasingly accessible to a wide range of actors, including non-state actors. By the mid-2000s, regular armed forces and non-state actors – sometimes used as proxies – were competing for the attention of a global audience in the field of information.

To this end, armed forces have to act simultaneously on two fronts – the homeland with the need to maintain the population's support for military operations, and the battleground, with the need to influence local populations and enemies – comprising different audiences and perceptions, sometimes with paradoxical orientations. The 2006 Second Lebanon War demonstrated the importance of information operations, especially in the second phase of the war, with the Lebanese Hezbollah acting in both physical domains and in the information domain to exert an impact on the whole Middle-Eastern audience as well as Western audiences, including through Lebanese diasporas. The creation of digital content included forged images and videos, particularly regarding the attack on the Israeli Sa'ar 5-class corvette *INS Hanit*, which were used to emphasize the ability of Hezbollah operatives to strike anywhere – including at sea – balancing the asymmetry between Hezbollah's "low-cost fighters" and a major force at the forefront of military technology.

Yet in recent years, major Western armed forces – following the Israeli post-2006 example – have engaged in a deep re-evaluation of doctrine and operational guidelines to deal with information-based threats, including in the context of growing hybrid threats that blur the distinction between state and non-state actors. This change also has to be considered in the context of a global doctrine evolution in Western armed forces, along with multidomain integration in which the information domain is a key element, also when it comes to enabling joint capabilities[3] There is therefore a deep link between technological developments in the IT sector and military information and influence operations, with changes in the former enabling new opportunities or threats for the latter.

In recent years, another major technological change has occurred with the fast-paced development of AI-based technologies. The shift from Information 1.0 to Information 2.0 in the mid-2000s had a major impact on military operations; the transition from Information 2.0 to Information 3.0 could also be a game-changer for operations, especially in the context of growing hybrid threats.

[3] UK Ministry of Defence, *Joint Concept Note 1/20, Multi-Domain Integration*, London, November 2020.

The rise of AI-based technologies

From entertainment to malign intent

After the emergence of cyberspace in the 1980s–1990s and social networks in the 2000s–2010s, the new emerging major information technology is centred around the use of artificial intelligence in communication, and across IT-related sectors generally. Artificial intelligence technologies seem to be of particular interest in the field of military influence, notably to elaborate dedicated messages. Creating relevant content for a specific targeted audience is one of the most critical phases of information operations. Yet AI-related technologies may have other uses beyond content creation, including the creation of fake profiles.

Creating a credible avatar to disseminate information is often one of the first – if not the most common – pitfalls of influence operations. In this respect, having an identity that is both trustworthy and that does not endanger the operators responsible for carrying out these manoeuvres is a major issue. Artificial intelligence technologies can provide viable solutions for this issue, with some of the machine learning-based technologies having the capability to improve the resolution in pictures using dedicated algorithms, and allowing videos to be created from fixed images, among other applications.

This kind of technology, based on neural networks, and useful for the analysis of satellite imagery for intelligence purposes, can also be diverted to create completely fictitious images. These deepfakes, created using generative adversarial networks (GAN), have been documented since 2014[4] with a notable growth in their use in social networks for setting up fake profiles, usually with a social engineering objective.[5] In this vein, AI-based technologies in deepfakes could be used for influence operations in the following orientations:

- **Avatar creation:** the aim is to create a totally fictitious person, using a picture or even animated images that could be used to penetrate a specific community, incorporating physical or moral traits identifiable by the target audience. To this end, it is necessary to train the AI in charge of creation using as much data as possible on the target community in order to acquire the maximum number of specified characteristics. The resulting fake persona has several uses, especially in terms of intelligence: mapping interpersonal networks, insertion into discussions, phishing, and so on. GAN-created avatars are now used on certain television channels – in China in particular – becoming the first quasi-human virtual presenters. Similarly, fake individuals created using GAN are even more present in social networks.[6]
- **Fake speech using a real person:** Here, the deepfake aims to utilize the image of a public personality – such as a political leader or military commander – to put them in an embarrassing situation or make them say things that they never said. This "2.0 forgery system" is particularly useful due to the tendency to

[4] I. Goodfellow et al., 'Generative Adversarial Networks', *Advances in Neural Information Processing Systems* 27, 2014.

[5] K. Giles, K. Hartmann and M. Mustaffa, *The Role of Deepfakes in Malign Influence Campaigns* (Riga: STRATCOM COE, 2019).

[6] See e.g., https://thispersondoesnotexist.com/, which publishes documented examples of these fake individuals.

> impersonate political and military communication. It is a classic subversion manoeuvre whereby the opposing leader is portrayed to his disadvantage as cruel/weak/a liar, and so forth. The tactic being deployed cannot be considered particularly innovative, yet in recent years there has been a spectacular increase in videos using famous people and created with GAN. Many examples are circulating in social networks, featuring Barack Obama or Donald Trump in speeches that are most often conspiratorial. The availability of many images or videos for specific people such as the US President allows extremely realistic fake videos to be created by melting numerous multi-angle image samples and voiceprints.

In this area, it is also relevant to consider military counter-influence, specifically the detection of deepfakes and GAN-generated images and videos. This capability should become critical for the armed forces in a few years, given the increase in the digital influence knowledge of hybrid threat actors. Considering the current situation in Mali and the Central African Republic with the implication of the Wagner Group in disinformation and influence operations against the French armed forces, it is clear that Russian-supported groups and individuals are already engaged in digital and non-digital influence operations, using social networks and electronic communications. If the technical level of these operations is still low to medium, the rapid increase in the use of digital technologies in sub-Saharan Africa could lead to the utilization of AI-based systems in the years to come.

Beyond this issue of the fight against a state or a state proxy such as the Wagner Group, it is also important to consider the hypothesis of non-state actors using these technologies. ISIS, for example, succeeded in setting up a professional-level communication system with multilingual entities and products (e.g., *Dabiq*, *Dar al-Islam* magazines). The ISIS propaganda machine relied in particular on the use of professional or semi-professional technologies such as drones for image and video recording, or computer graphics software. In this context, it is foreseeable that major non-state actors – including some state-supported ones such as Hezbollah – may use AI-based technologies in the short to medium term for digital influence purposes.

Beyond deepfakes

Moreover, one of the major challenges in information operations is the creation of "sound boxes" which, after the creation of a message, provide amplitude in dissemination to obtain the desired effect.

Beyond the creation of isolated individuals or a single deepfake video, the danger lies more in complex immersive systems called fictitious algorithmic projections.[7] Using AI-based technologies, they create a massification of false data to enable multi-level interactions, and could lead to the ability to create a mass of fake persons interacting to give the impression, for example, of a consensus within a group. With trend and popularity being at the core of social networks – as highlighted by concepts such as Google Ranking and Top Tweets – having the

[7] T. Berthier and B. Teboul, *From Digital Traces to Algorithmic Projections* (Elsevier Science Direct, 2018).

possibility to emulate a conversation between multiple accounts may lead to the spread of information using a snowball effect.

Beyond deepfakes and content creation, the issue of broadcasting is a major hurdle for influence operations. Reaching a large audience has always been an important issue, especially in asymmetric warfare in regions with low access to digital media. In Afghanistan and sub-Saharan Africa during the 2000s, there was barely any use of digital influence operations, as the local populations relied on traditional media, mostly radio or television, for information. Thus, influence through digital tools was mostly oriented towards audiences in Europe and the US, with significant use of social media by armed groups such as the Taliban, Iraqi insurgents or Somalian al-Shabab[8] militants to disinform Western audiences, including through agit-prop (agitation propaganda) tactics.

Yet the situation is evolving with the global deployment of broadband mobile networks, especially 5G- based. The 5G bandwidth allows high-quality video content to be disseminated and a large number of targets to be reached, enabling mass audience influence strategies in a growing number of territories beyond Europe and the US. More than broadband terrestrial networks, the coupling between space and Earth with the development of satellite constellations in Low Earth Orbit (LEO), supported by major US and Chinese companies, may also strengthen the massive digital communication systems by covering new territories and enhancing the bandwidth and resilience of terrestrial networks. Combining 5G, beyond 5G and LEO communication networks and protocols with instant messaging systems, such as Telegram, Signal or WhatsApp, should make the dissemination of deepfakes and influence messages much simpler, potentially with a very significant snowballing effect.

[8] F-B. Huyghe, O. Kempf and N. Mazzucchi, *Gagner les cyberconflits* (Paris: Economica, 2015).

A perfect tool for hybrid warfare?

The importance of info-ops in modern warfare has to be linked with the cost effectiveness of such operations. For militaries, the use of influence-dedicated tools and massive information operation campaigns remains difficult to develop as it is still impossible to precisely assess the outcome. Therefore, some military commanders are still reluctant to use these capabilities as the result is largely unforeseeable and the performance hard to analyze. Often time-consuming, influence operations are also considered difficult to implement when facing a non-state or hybrid threat actor.

On the other hand, non-state and hybrid threat actors are often eager to engage in influence tactics for asymmetric warfare, yet access to sophisticated technologies remains an issue for non-state or proxy actors. AI-based technologies are currently considered to be emerging and they require specific skills and equipment to create convincing avatars and content for their information operations. As technology and technical skills are critical issues, an analysis of their availability to a wide audience is needed to assess the possibility of AI-based technologies being used by hybrid threat actors and non-state actors to deceive or intoxicate an opposing force.

In 2020, Hwang proposed a model for assessing the conditions for AI deepfake-based technology proliferation.[9] Four different bottlenecks were identified: training data, specialized hardware, technical expertise and software.

- **When it comes to training data**, the importance of military and political communication, as already highlighted, helps hybrid threat and non-state actors to access large amounts of data for training AI systems to create deepfakes of political leaders or military commanders with relative ease.
- **Specialized hardware** is a more difficult issue as some of the most powerful dedicated processors – like the x86 category – fall under dual-use export regulations, such as the Export Administration Regulations of the United States, and are therefore under scrutiny. Yet the development of new generations of high-capacity processors in China, with companies such as Huawei, ZTE or Cambricon, could lead in the coming years to significant dissemination alongside a major drop in costs. The analysis of AI technologies embedded in the personal devices market, for example, with the anticipated increasing use of AI chips in smartphones for instance,[10] indicates the potential fast-track democratization of AI-chip use in next generation mainstream smartphones, making them easily accessible for a wide range of actors.
- **Technical expertise** is also a complicated issue for hybrid threat and non-state actors as the human resources for AI-based technologies are considered to be in short supply, also for the wealthiest companies. Hence, attracting people with sufficient knowledge to create algorithms and AI-based systems could pose a major bottleneck for these actors, emphasizing the importance of state support in this particular area.

[9] T. Hwang, *Deepfakes, Primer and Forecast* (Riga: STRATCOM COE, 2020).

[10] See https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2020/ai-chips.html.

- **Software**, conversely, is not a major issue as much AI-dedicated software that is useful for the creation of GAN and deepfakes is open-source. Google TensorFlow, one of the most popular AI-training platforms with a high level of flexibility, is a major example that provides immediate access to medium- to high-quality deepfake creation tools at very low cost.

An intermediate conclusion using the Hwang value chain of AI-based technologies for deepfakes is the importance of human resources for hybrid threat and non-state actors. As most of the technical elements are already available or experiencing rapid cost decrease and availability increase, the issue of personnel with the requisite knowledge to create and maintain dedicated capabilities in deepfakes remains key.

The war in Ukraine: the first deepfake-supported conflict?

In the recent war in Ukraine, deepfakes have been used to support military operations, especially on the Russian side for deception purposes. On 16 March 2022, the Ukrainian TV channel Ukraine 24 appeared to have been hacked by pro-Russian hackers, leading to the broadcasting of a written message supposedly from President Zelensky calling for Ukrainian soldiers to surrender. That same day, deepfake videos using Volodymyr Zelensky's face were broadcast on the instant messaging system Telegram, promoting the same message that Ukrainian soldiers were to surrender to the Russian forces. This fake video was also published on several social media platforms,[11] including the Russian Vkontakte under the indirect supervision of the Kremlin. On the opposing side, deepfakes of Vladimir Putin were also broadcast on social media, underlining the growing use of this technology.

Yet this fake video of the Ukrainian President appears to be quite simplistic, with low-quality samples of Zelensky's voice and technical problems in the animation.[12] As a consequence, the fake was rapidly debunked and had hardly any impact on the Ukrainian population. Nevertheless, the use of deepfakes incorporating a major political leader in wartime is a new feature of influence operations, combined with the simultaneous action in cyberspace to hack the Ukraine 24 TV channel.

In terms of hybrid warfare, the combination of operations in the cyber and information domains is consistent with the Russian habitus of *maskirovka* or deception.[13] Moreover, the use of hybrid threat and non-state actors could also be understood as a way to bypass the decision by major social media platforms such as YouTube or Facebook to ban video channels under direct Russian support. The major issue here remains the digital literacy of populations. Even with medium-quality deepfakes, targeting a population with low level of digital literacy and awareness of media forgery could lead to major real-life impacts, such as protests or even riots.

[11] See https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-war-report-hacked-news-program-and-deepfake-video-spread-false-zelenskyy-claims/.

[12] See https://www.bbc.com/news/technology-60780142.

[13] Deception in a military sense is defined in the US DoD Joint Publication 3–13.4 as "actions executed to deliberately mislead adversary military, paramilitary, or violent extremist organization decision makers, thereby causing the adversary to take specific actions (or inactions) that will contribute to the accomplishment of the friendly mission".

Influence operations to deceive opposing military forces and to force surrender or paralyze the chain of command are fairly traditional in the history of warfare and strategy. With the anticipated development of AI-based technologies and their availability to the public at large, the sophistication of these operations could reach a new level. Fortunately, deepfake detection – also using GAN – is likewise being developed and financed by states and major IT companies, as the risk of large-scale deepfake influence campaigns has been considered for several years.

Conclusions

Acceleration in the development of information technology created a new era in communication at the end of the 20th century, allowing quasi-immediate access to any document or information in the world, via the use of cyberspace. The Web 2.0 that flourished after 2004 deepened this revolution in communication by abolishing the traditional wall between information providers and information consumers, making all users potential information prosumers (producer-consumers). Yet the optimistic view of cyberspace as an enabler of peace and stability, with its capacity to raise awareness and to allow free speech and free information, proved to be a distortion of reality. The first conflicts of the 21st century demonstrated that information technology and cyberspace could be used with malign intent to design and execute influence operations targeting mass audiences and specific communities, especially in the West. Non-state actors such as ISIS, the Taliban or Hezbollah turned out to be significant users of digital communication for propaganda and influence.

Nowadays, emerging digital technologies in content creation as well as in broadcasting could have major impacts on conflicts. **The rise of AI-based technologies offering the possibility to create realistic deepfakes – with easy access to most of the elements needed to create these distorted videos – and the future telecommunications networks able to broadcast high-quality video content, are salient features that armed forces will have to deal with.**

For hybrid threat actors, the ability to access these capabilities could strengthen influence and information operations targeting homeland and battlefield audiences. All major info-ops and psy-ops tactics such as deception or intoxication could therefore benefit from these emerging digital technologies with the potential automation of forgery. **The 2022 war in Ukraine demonstrates that the use of deepfakes is an emerging trend**, and even if the videos that were broadcast appeared quite unsophisticated, it must be remembered that we are only on the eve of the AI-powered influence era.

NATO, the EU and European countries should therefore consider major updates in the doctrine and process for influence and counter-influence operations using the information domain. The need to develop internal capabilities to detect and counter deepfakes using GAN technologies arose with the use of deepfakes in Ukraine. **In the near future, the use of AI-based technologies to create false speech and videos could therefore become a new normal in conventional and unconventional warfare. The training of dedicated staff is essential to achieve the requisite knowledge of how to counter information warfare 3.0.** As most influence training has historically revolved around psychology and the use of traditional media, there is now a need to add a new layer with the increasing importance of AI-based technologies.

As these technologies are not only related to the military and cannot be considered "military equipment" by nature, there will probably be no legal opportunity to restrict their use. Yet the question that needs addressing in the future is whether the use of AI-forged content could be considered ethical even in an open warfare context for armed forces in the democratic world.

Author

Dr Nicolas Mazzucchi is Research Director at the Centre for Strategic Studies of the French Navy (CESM). Before joining the CESM, he pursued a career at the French MoD working in research and in operations, as well as in major French think tanks (FRS, IRIS). He is a specialist in cyber, energy and primary goods issues. A reserve officer (Lt-Col/OF-4) at the French MoD Joint Staff, and a Professor in Geopolitics for the French War College and General Officers Course, he holds a PhD in Economic Geography and is also an alumnus of the French War College.

Hybrid CoE

The European Centre of Excellence
for Countering Hybrid Threats