IRSEM
INSTITUT DE RECHERCHE STRATÉGIQUE
DE L'ÉCOLE MILITAIRE

www.irsem.fr

Strategic Brief - 62

July 20, 2023

Information Warfare in the Age of Artificial Intelligence

Dusan Bozalka

Artificial intelligence (AI) not only enhances the offensive capabilities of malicious actors in the digital realm but also further weakens the defensive capacities of our societies against information manipulation. Based on discussions held at the annual conference of the NATO Strategic Communications Centre of Excellence in Riga, Latvia, as well as on current events, this strategic brief aims to summarize the dangers presented by tools utilizing this technology.

While the democratization of the Internet was initially accompanied by an emancipatory vision, there are now numerous concerns regarding the increasing incursion of AI into our societies. It is evident that **this technology is becoming a permanent fixture in our daily lives**, as indicated by the availability of freely accessible tools such as image generators and conversational agents. The most notorious prototypes in this class of tools are the latest versions of pre-trained generative transformers housed in the conversational agent developed by the American company OpenAI, ChatGPT. Employing a complex architecture known as a "neural network," this technology uses a model dedicated to processing word sequences in order to predict the best possible response based on available data. Image generation tools, such as Midjourney's, utilize a diffusion model where random noise is gradually inverted to generate high-quality images based on user instructions. In doing so, these tools are **strictly shaped by human input**, tempering the apocalyptic scenarios associated with full automation of AI. However, they still represent a factor that intensifies the ambitions pursued by the actors who employ them, including various malicious uses already observed in the digital realm.

This realization is increasingly prompting international institutions and their member states to question the harmful effects of AI, particularly in a geopolitical context marked by a resurgence of information warfare. One of the most recent examples is the annual conference of the NATO Strategic Communications Centre of Excellence, held in June 2023 in Riga. Drawing on exchanges between researchers and high-ranking military personnel, this strategic brief aims to present two lessons regarding the informational threats posed by this technology.

AI firstly represents an **advantage for the offensive capabilities of states and malicious private actors engaged in information warfare**. This is due to the increasing mobilization of tools derived from the technology to mass-produce false information based on formulated requests and **contribute to the intensification of information fog**. As demonstrated in a recent investigation by *NewsGuard*, **the responses generated by ChatGPT models frequently serve as conduits for Chinese and Russian government propaganda**. Once utilized by authoritarian state actors, the provided responses are used to support propaganda narratives and disinformation operations. While safeguards exist to prevent these tools from accessing requests that are considered to be malicious, **there are certain loopholes in their moderation**: it is possible to "jailbreak" ChatGPT, meaning to bypass the limitations imposed by OpenAI, by preceding a query with a hypothetical scenario or introducing it in the form of a role-playing game (prompt hacking). Consequently, some experts warn about the **cost reduction associated with the use of AI in information manipulation**. Indeed, these tools are capable of **offering narratives tailored to specific cultural and linguistic contexts**, targeting diverse sociological groups to ensure better reception. It thus becomes possible to produce high-quality propaganda narratives at a relatively low cost, with AI replacing the need to hire individuals who are aware of national issues. Moreover, the ability of ChatGPT to code in different programming languages also facilitates the creation of online sites and networks of automated accounts (botnets) intended to artificially amplify the formulated narratives, a practice known as astroturfing.

Beyond state actors, this trend is likely to increase malicious activities from private actors in two different ways. Firstly, there may be a **significant increase in the creation of botnets available on the dark web and active on social networks**, precisely when platforms such as Twitter are experiencing massive lay-offs in their moderation capacities; secondly, there is a **proliferation of cybercriminal operations, such as phishing**, where ChatGPT enables the rapid formulation of fraudulent emails with a convincing level of realism capable of bypassing anti-spam filters.

From this first point **ensue systemic limitations inherent to the defensive posture of democracies**, primarily rooted in technical and economic reasons. **Tools derived from AI are not capable of combating information manipulation**, since they are unable to detect it. To date, only certain **recurring errors** allow for the identification of artificial production, such as distorted hands or blurry backgrounds in the case of image generators. In the case of conversational agents, this phenomenon manifests through redundant phrases, the presence of improbable statistics within the body of texts, and error messages published by automated accounts (bots) on digital platforms. In terms of defense, the utility of AI ultimately lies in optimizing the working time available to specialists in information manipulation. They can fully dedicate themselves to identifying these manipulations and automate their categorization.

While AI models are accessible to everyone through open access, **their development and training require hundreds of millions of euros of investment, for the preparation of exploitable databases and the use of powerful computer servers**. This limitation reveals a second economic challenge: similar to digital platforms, the data exchanged between users and **AI tools remain the exclusive domain of an oligopoly of companies that cannot be requisitioned as they would in authoritarian regimes**. Due to lack of transparency and resources these technical and financial shortcomings prevent policymakers and the scientific community from better understanding the threats posed by AI. Additionally, there could be **geopolitical implications for efforts to regulate this technology**. While calls for the creation of international centers to monitor developments related to AI may seem commendable, this process remains arduous and compromised by the realpolitik of states.

The informational dangers that this technology poses to the resilience of Western societies are tangible, but **several responses can mitigate their effects**. One of these involves a **substantial increase in investment dedicated to academic research**. More research would improve the recognition of messages generated by AI tools, particularly through a model based on a combination of stylometry (the recognition of redundant linguistic patterns) and machine learning. Another response involves making the data collected by private companies available, which would amplify efforts towards the defensive use of their tools. A recent study highlights the beneficial potential of such efforts, including by ChatGPT, in enhancing the efficiency and speed of fact-checking processes.

As AI threatens the cohesion of our democracies as a whole, these **responses require rethinking legal frameworks**. Clear rules are necessary to hold companies accountable when their tools are used in disinformation operations. **The academic consensus seems to advocate for a rapid adaptation of national and even European legal frameworks to address AI-related provisions** such as copyright and user data protection. In parallel, preventive measures can be taken to respond to the threats posed by the authoritarian use of artificial intelligence, similar to the decision by the United States to control the export of processors to China. Added to which is a need to **broaden the methods used by government actors in their strategic communication**, supported by the positive effects of preventive exposure to misinformation (pre-bunking). ■

Dusan Bozalka *is a resident PhD candidate at IRSEM and affiliated with the Center for Interdisciplinary Media Research and Analysis at the University of Paris-Panthéon-Assas.*

Contact: dusan.bozalka@irsem.fr

Disclaimer: One of IRSEM's missions is to contribute to public debate on issues relating to defence and security. The views expressed in IRSEM's publications are the authors' alone and are in no way representative of an official Ministry for Armed Forces stance.