Key Aspects of Spreading and Creating Disinformation Using Artificial Intelligence

Oleksandr Kin

Candidate of Technical Sciences, Senior Research Fellow, Associate Professor, e-mail: cubalibre1972@ukr.net,
ORCID ID: https://orcid.org/0009-0001-2196-7515

Hennadiy Udovenko Diplomatic Academy of Ukraine under the Ministry of Foreign Affairs, Kyiv, Ukraine

Received: February 4, 2026 | Revised: March 20, 2026 | Accepted: March 31, 2026

UDC 004.8:32.019.5

DOI: https://doi.org/10.33445/psssj.2026.7.1.1

Abstract

The article examines the transformation of mechanisms for the creation and dissemination of disinformation under conditions of the active integration of artificial intelligence technologies into the information and communication space. It substantiates that the development of generative models, particularly large language models and deep learning systems, significantly increases the scale, speed, and persuasiveness of information and psychological influence. It is established that artificial intelligence not only automates the production of fake content but also enables its personalization, adaptation to the characteristics of target audiences, and integration into social media through bots and algorithmic systems.

The study analyzes key artificial intelligence tools used for disinformation (language models, deepfake technologies, and systems for voice and image synthesis), as well as their functional capabilities in the context of cognitive warfare. It is determined that the critical factors intensifying disinformation include the accessibility of technologies, reduced costs of information operations, and the phenomenon of "truth decay," which erodes trust in all sources of information.

Based on the analysis of empirical studies, it is demonstrated that artificial intelligence - generated content can match or even surpass traditional propaganda in terms of persuasiveness. At the same time, a potential negative impact of AI on human cognitive abilities is identified, particularly a decline in critical thinking.

It is concluded that the use of artificial intelligence in disinformation constitutes a systemic threat to information security and requires the development of comprehensive interdisciplinary countermeasures, including legal regulation, technological solutions, and the enhancement of media literacy.

Key words: Hybrid Warfare, Cognitive Warfare, Information Warfare, Cognitive Domain Operations, Artificial Intelligence.

Introduction

Today, defence industries of the world's leading countries demonstrate a clear trend towards the development of advanced weapons systems based on the principles and functional characteristics of the Fourth Industrial Revolution, force-multiplier concepts. These systems significantly exceed, or are expected to exceed, the capabilities of existing weapons (in some countries, such weapons are already being adopted). Such weapons systems will be comparable in their combat characteristics and effects to high-precision and even nuclear weapons. Accordingly, views on the content, tactics and strategy of armed struggle are changing – they are no longer seen as clashes between combat units that strike each other in the course of firefights and capture enemy territory, but as clashes between multifunctional combat systems, the main purpose of which is to deprive the opposing system of its ability to act.

New trends significantly affect various elements and components of defence, requiring military institutions to adjust their strategies and strengthen their capabilities to effectively address the complex challenges of modern warfare.

The use of artificial intelligence (AI) technologies in new weapons control systems will enable a revolution in combat operations. Such systems will accelerate decision-making by taking into account the maximum number of factors and significantly reduce the command-and-control cycle, as well as increase combat capabilities. Currently, American specialists have achieved the greatest results in the field of deep machine learning. It has become possible due to rapid progress in the development of multi-layer artificial neural networks. Pattern recognition and machine learning technologies used in modern AI can provide a high degree of automation in command-and-control decision-making during the analysis and processing of information from multiple sources.

Literature Review

In contemporary scholarly discourse, the issue of disinformation and cognitive influence in the context of digitalization is considered an interdisciplinary field that integrates approaches from political science, security studies, information technology, and cognitive psychology. This topic has gained particular relevance due to the rapid development of AI technologies, which are transforming both the tools and the scale of informational influence.

Classical approaches to the analysis of information operations and propaganda are reflected in the works of H. Pocheptsov, who considers informational influence as a systemic element of hybrid warfare, as well as in studies by A. Oslund, L. Gudkov, and K. Rogov, which emphasize the political and social mechanisms of manipulating public consciousness. A significant contribution to the development of the concept of cognitive warfare has been made by Ukrainian scholars such as H. Perepelytsya, T. Berezovets, O. Semenenko, and V. Horbatyuk, who substantiate the role of information and psychological influence as a key instrument for achieving strategic objectives in modern conflicts.

In the international academic discourse, a distinct research direction has emerged focusing on the phenomenon of cognitive warfare, understood as a struggle for control over perception, thinking, and human behavior. In particular, NATO analytical documents highlight that the cognitive domain is becoming a new operational space alongside traditional domains (land, maritime, air, and cyberspace). In this context, information operations are aimed not only at spreading disinformation but also at shaping stable cognitive patterns that determine the behavior of individuals and groups.

A separate body of research examines the impact of digital platforms and social media on the spread of disinformation. Studies by R. Ortung, D. Zolotukhin, H. Yavorska, and O. Bondarenko focus on algorithmic mechanisms of content dissemination that contribute to the formation of information “bubbles” and increased societal polarization. These works demonstrate that the structure of the digital environment itself acts as a catalyst for disinformation, even without the application of advanced AI technologies.

The most recent stage of research is associated with the study of generative AI as a tool for creating disinformation. In particular, Hanley and Durumeric (2023) empirically demonstrate a rapid increase in machine-generated news and its dissemination across both mainstream and disinformation-oriented platforms. Tomz (2024) shows that AI-generated propaganda can be comparable to, or even more effective than, human-produced propaganda, especially after editorial refinement.

An important research direction also concerns the impact of AI on human cognitive abilities. A study conducted by Microsoft and Carnegie Mellon University (2025) identified a tendency toward a decline in critical thinking under conditions of high trust in AI tools. This finding correlates with the concept of “truth decay,” which describes the erosion of trust in facts and expert knowledge due to information overload and the widespread dissemination of misleading content.

At the same time, the analysis of scientific sources reveals several unresolved issues. First, there is no unified methodology for quantitatively measuring the effectiveness of disinformation, particularly AI-generated content. Second, institutional mechanisms for countering such threats remain insufficiently explored, especially regarding coordination among governments, the private sector, and civil society. Third, most studies remain descriptive in nature and lack sufficient operationalization of key concepts.

Thus, the current state of research indicates the emergence of a new paradigm in the study of disinformation, with artificial intelligence at its core as a factor fundamentally transforming the information environment. At the same time, there is a clear need to deepen empirical research, develop interdisciplinary approaches, and design applied countermeasure models to ensure the transition from theoretical understanding to effective practical implementation.

Results

In recent years, there has been a significant increase in the use of AI to create disinformation and propaganda. In particular, since the end of 2023, when AI-based image generators became widely available, the number of fake images created with their help has grown significantly. Until early 2023, such images were a minor part of the manipulated content, but their number began to grow rapidly. A 2023 Arxiv study showed that the number of fake news reports increased by 57.3% on traditional websites and by 474% on disinformation resources between 2022 and 2023 (Hanley, H. W. A., & Durumeric, Z., 2023).

AI-generated content looks very authentic. Some of the strategies currently used to detect and counter manipulative AI-generated content and the accounts that spread it are becoming ineffective. How does disinformation generated by AI differ from traditional, human-generated disinformation? AI can significantly worsen the problem of spreading and creating disinformation in several keyways:

Scale and speed of dissemination – AI makes possible the mass creation of content for disinformation campaigns, which can lead to an increase in the number of fake stories, multiple variations of the same narrative, its generation in different languages, automated dialogues, etc. Compared to manual content creation, AI technologies allow disinformation to be produced in a matter of seconds. These two factors – scale and speed – pose a serious challenge for fact-checkers, who are faced with a continuous stream of disinformation. It can be assumed that in the near future, the media space will be flooded with such information, as factchecking will take a considerable amount of time.

Accessibility – the active distribution of AI tools eliminates barriers for conducting information operations. People of different wealth, status, and social standing will be able to create realistic fake images and videos without professional skills or complex editing. This could lead to the "democratisation" of troll farms.

Persuasiveness of content – deep neural networks can generate realistic texts, images and videos (e.g. deepfakes) that are difficult to distinguish from the real ones. In the age of clip thinking, information is perceived in a fragmented, rapid and superficial manner. High-impact data visualization using AI tools, while may not be perceived as absolute truth, can leave a certain psychological impact, which in turn influences the perception and interpretation of subsequent information on the same topic.

Targeting and personalisation – AI technologies make possible the launch of personalised disinformation campaigns targeting specific audiences (or even individuals) based on their preferences or beliefs, even without in-depth knowledge of the target group's language or culture. AI is capable of experimenting in real time, observing people's behaviour and quickly identifying the most effective strategies and tactics. This allows AI to adapt its approaches much faster than

humans. For example, such disinformation can be targeted at people of different ages, political views, religious beliefs or personality types (e.g., extroverts or introverts), which increases its efficiency. People who are already socially marginalised or have low media literacy may be particularly vulnerable. In addition, AI can analyse digital activity and financial transactions, creating a psychological profile or target data package for almost every person.

Automatization of manipulations – bots and algorithms can imitate human activity on social networks, creating the illusion of support for certain ideas or pressure on public opinion. In addition, this makes it possible to significantly reduce the financial costs of such information campaigns, in particular by limiting the number of specialists involved in their implementation.

Truth Decay – a large amount of high-quality fake content can cause general distrust of all sources of information. This makes it difficult to find reliable information.

Theoretically, if AI detects that the target has neurotic traits, it will be able to create a manipulative message that will cause severe stress, to which such a person is more susceptible. Although AI is not yet capable of diagnosing disorders, it is rapidly advancing in this direction. For example, if a soldier is identified as depressed, AI could potentially try to provoke suicide.

How can all these AI capabilities be implemented in real-life cases, such as the war in Ukraine? This tactic has been used repeatedly: AI collected photos of killed soldiers, compared their faces with photos on social media, identified them, and sent the photos to the families of the deceased. Of course, deepfakes may also be used. Such actions will have a devastating psychological effect on families, which, in turn, undermines morale and sows fear. This strategy can also overload the command of rear-echelon units and the garrison psychological services. If soldiers on the front line become aware of such incidents, it may have a serious demoralising effect.

Mass content creation can reinforce tactics aimed at distracting the audience and creating the illusion of majority (since the content appears to come from different sources). For example, state and state-affiliated actors, such as russia's "Internet Research Agency", have long used hundreds of accounts to distract attention from uncomfortable topics, and with the advent of generative AI, this practice is becoming even easier. At the same time, creating automated real-time dialogues can help disguise bots and make their social media accounts more credible.

Every year, the number of publicly available AI systems is steadily increasing, and each of them serves as evidence not only of the active development of cyber technologies, but also as a source of threats to the information and communication domain and national security as a whole. Given the diversity of AI systems, they can be typologically divided according to the following criteria: level of development, training methods, operating principle, and functionality (Table).

To understand AI disinformation tools, it is useful to examine several technologies that are most commonly used to generate and spread false information.

1. GPT-3 and GPT-4 (OpenAI). These language models can generate realistic text that is used to create fake news, articles, and to automate comments and posts on social media. They can imitate the style of well-known sources, making them perfect for creating disinformation.

2. DeepAI and other deepfake generation tools. Generating fake videos using deep-fake algorithms is one of the most effective and harmful forms of disinformation. Instruments such as DeepAI, FakeApp, 4Reface, or Deep FaceLab make it possible to create realistic videos with specified parameters featuring any person in the lead role, prescribing scenarios for their actions in such a way that it becomes extremely convincing for the target audience.

3. Artbreeder. This is an AI model that allows users to create realistic images based on genetic algorithms. It is used to create fake profiles of people or manipulated images for disinformation purposes.

4. Synthesia. A tool for creating videos with avatars, which can be used to generate fake news or statements on behalf of public figures.

5. Lyrebird (Descript). This platform allows users to create audio files that accurately imitate human voices using AI. It is actively used to create fake recordings, which are then used to manipulate public opinion or in smear campaigns.

6. BERT (Bidirectional Encoder Representations from Transformers). BERT-based models are actively used for text analysis, creating fake comments or reviews on social networks, as well as for detecting targeted manipulations.

Table: Types of AI systems according to various criteria

TYPES OF AI SYSTEMS	ALGORITHM DESCRIPTION
By level of development	
Artificial Narrow Intelligence (ANI)	specialised AI that performs one specific task (e.g., voice assistants, search algorithms, recommendation systems)
Artificial General Intelligence (AGI)	hypothetical AI capable of thinking and learning like a human being, adapting to new tasks
Artificial Superintelligence (ASI)	theoretical level at which AI surpasses human intelligence in all areas
By training methods	
Machine Learning	AI that learns from data using statistical methods
Deep Learning	uses deep (multi-layer) neural networks to analyse complex patterns
Evolutionary algorithms (EA)	simulate evolutionary processes such as natural selection to find optimal solutions
By operating principle	
Symbolic AI	works on the basis of logical rules and knowledge
Neural Network-based AI	uses artificial neural networks inspired by the human brain
Hybrid AI	combines multiple approaches to improve efficiency
By functionality	
Reactive AI	reacts to input data, has no memory of past experiences (e.g., Deep Blue chess computer)
Limited Memory AI	takes into account previous data (e.g., autopilots in cars)
Theory of Mind AI	capable of understanding human emotions and intentions
Self-aware AI	AI endowed with consciousness; currently theoretical

Source: Compiled by the author.

7. ChatGPT. This chatbot can be used for automated conversations, spreading propaganda, as well as for manipulated discussions and creating fake text dialogues that look authentic.

8. TuringBot. This bot uses AI to generate texts that look and sound like real news or messages. It is capable of mass-producing false information to be spread on the Internet.

9. Grok 3. The latest AI model developed by xAI under the leadership of Elon Musk. It combines the capacity for reasoning with a large amount of knowledge, making it one of the most powerful AI systems today. This AI has repeatedly been in the spotlight due to its involvement in spreading misinformation. For example, in April 2024, the Grok chatbot generated fake news about Iran's attack on Israel. This false information was published before the actual event and subsequently appeared in the official popular news section on platform X, expanding its reach among users (Lee, S., et al., 2025).

With the help of the above-described and other software, it is possible to exert reflexive control over an individual within the framework of cognitive warfare, when emotions, behaviour,

and consciousness of the opponent are influenced using specially prepared information supported by AI tools, which encourages them to make the desired decision. The long-term influence of AI-generated information affects the cognitive abilities of individuals, namely, it suppresses them, destroys critical thinking and abilities for creative self-realisation as well as the ability to make non-standard decisions. This was proven by a study conducted by Microsoft and Carnegie Mellon University (Tomz, M., Weeks, J. L. P., & Yarhi-Milo, K., 2024), involving 319 IT professionals. A direct correlation was found between confidence in AI and a decline in critical thinking. Critical thinking was defined as falling into one of six categories: knowledge (memorising ideas), understanding (understanding ideas), application (implementing ideas in the real world), analysis (contrasting and connecting ideas), synthesis (combining ideas), and evaluation (evaluating ideas). Participants demonstrated "the potential for excessive trust in technology without proper verification", leaving critical thinking to ChatGPT instead of doing it themselves and strengthening their cognitive abilities. As a result, in an era of clip-like thinking and "smart feeds" on social media pages, the flow of information and microtargeting create an unalterable artificial reality in which the individual does not have time to filter data and build cause-and-effect relationships. Such tools for protection against disinformation as debunking (refutation) and prebunking (pre-refutation) cease to be effective. Debunking means refuting something that has already happened, while prebunking is a proactive measure where refutation is made in advance. Both of these strategies require the creation of an effective roadmap for presenting information to counteract the other, and if AI technologies are not used mirror-like in this process, all efforts will be drowned out in a sea of generated false or distorted information, the spread of which is dozens of times faster than the release of information by traditional media methods.

Based on the experiment conducted by Microsoft and Carnegie Mellon University, it can be assumed that since the use of AI reduces human cognitive and creative abilities, such generated information will have less potential impact, as it will be more primitive and less creatively presented. The answer to this assumption was given by Michael Tomz, professor of political science at Stanford School of Humanities and Sciences and a research fellow at the Stanford Institute for Human-Centred Artificial Intelligence (HAI) (Tidy, J., 2024, April). Based on the experiment, the professor determined how disinformation generated in three different ways affects respondents.

The first group, the control group, read a series of statements that people must be made to believe in. For example: "Most US drone strikes in the Middle East were directed against the civilian population, not terrorists" or "Western sanctions have caused a shortage of medical supplies in Syria". Since this group only read these statements without the accompanying propaganda, they became the starting point for assessing how many people already believe these statements. The second group of participants read propaganda created by people, which was written based on these thesis statements and later exposed by journalists-investigators or researchers. The third group received propaganda materials on the same topics as the first two groups, but created by the large language model GPT-3.

Researchers found that approximately a quarter of the control group agreed with the statements without reading any propaganda. Propaganda written by humans increased this indicator to 47% and propaganda generated by AI to 43%. When the materials generated by AI were edited by specialists and articles that did not convey the desired message were rejected, it was found that a full 53% of participants agreed with the thesis statements after reading the propaganda, which proved to be more effective than even propaganda written exclusively by humans.

This experiment showed how influential the proper use of AI can be in disinformation campaigns, even though the immediate consequences of such persuasive propaganda are not obvious. When we consider the political battlefield of cognitive warfare, especially foreign policy issues, it is important to remember that people may not have in-depth knowledge or formed

opinions about them. Precisely this target group becomes the main audience for the influence of AI disinformation. The efficiency of automated propaganda writing also frees up human resources for other tasks within the campaign, such as creating fake accounts on social media.

Discussion

The obtained results indicate a qualitative transformation in the nature of disinformation under the influence of artificial intelligence technologies; however, several provisions require critical clarification.

First, the claim of a universal increase in the effectiveness of AI-driven disinformation appears partially oversimplified. Although empirical evidence demonstrates a high level of persuasiveness, it also reveals significant variability in effects depending on context, audience type, and the level of information saturation. Therefore, the causal relationship between AI use and audience behavior change is not linear.

Second, the assumption regarding the decline in cognitive abilities due to AI use should be formulated more cautiously. The observed effects may be associated not only with AI technologies but also with broader processes of digitalization, including information overload and shifts in information consumption patterns. This suggests a risk of false causality.

Third, the institutional dimension of countering disinformation is insufficiently elaborated. While the need for cross-sectoral cooperation is emphasized, specific mechanisms of coordination among governments, technology companies, and the scientific community are not detailed. This creates a methodological gap between problem diagnosis and practical solutions.

An alternative interpretation suggests that AI functions less as an autonomous source of threat and more as a multiplier of existing information practices. In this context, the key factor is not the technology itself but the socio-political conditions of its use.

To strengthen the scientific validity of the study, it is advisable to:

expand the empirical base through cross-country comparative studies;

apply quantitative methods of impact assessment (experimental designs, A/B testing);

integrate approaches from behavioral economics and cognitive psychology;

elaborate institutional countermeasure models (at NATO, EU, and national security levels).

Thus, the study provides an important theoretical foundation but requires further operationalization of its conclusions to move from a descriptive to an applied analytical level.

Conclusions

Based on the above, it can be argued that new technologies are developing much faster than government policy and often undermine existing legal and political frameworks. To ensure the responsible use of AI and develop appropriate responses to its potential misuse at an early stage, it is necessary to establish stronger ties, partnerships and open dialogue between legislators, engineers and researchers.

Directions for future research are related to scientific examination of several factors that are becoming decisive in military activities:

first, changes in the security environment;

second, the development of existing cutting-edge technologies and their use in the cognitive domain;

third, the emergence of new breakthrough technologies that will be used (adapted) in one way or another to ensure the cognitive advantages of NATO and its allies in understanding both themselves and their enemies, which will have significant practical implications.

Funding

This study received no specific financial support.

Competing interests

The authors declare that they have no competing interests.

References

Hanley, H. W. A., & Durumeric, Z. (2023). *Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites*. arXiv. https://doi.org/10.48550/arXiv.2305.09820

Lee, S., et al. (2025). *The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*. Microsoft Research. https://www.microsoft.com/en-us/research/uploads/prod/2025/01/lee_2025_ai_critical_thinking_survey.pdf

Tomz, M., Weeks, J. L. P., & Yarhi-Milo, K. (2024). How persuasive is AI-generated propaganda? *PNAS Nexus, 3*(2), pgae034. https://doi.org/10.1093/pnasnexus/pgae034

Tidy, J. (2024, April). Elon Musk's X pushed a fake headline about Iran attacking Israel. X's AI chatbot Grok made it up. *Mashable*. https://mashable.com/article/elon-musk-x-twitter-ai-chatbot-grok-fake-news-trending-explore

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. https://rm.coe.int/information-disorder-report/168076277c