Animation, territoires et pratiques socioculturelles
Sociocultural community development and practices
Animación, territorios y prácticas socioculturales

Démasquer la propagande pour saisir les idéologies
Unmask propaganda to grasp ideologies
Desenmascarar la propaganda para captar las ideologías

Manufacturing "realities". The impact of AI in the age of disinformation

Ioan-Claudiu Farcaș
Technical University of Cluj-Napoca, Romania
ioan.farcas@ssuta.utcluj.ro

Cet article examine l'effet de l'intelligence artificielle générative (IA) sur la propagande, induisant un changement significatif par rapport aux méthodes traditionnelles. La propagande du 20e siècle était basée sur des récits simplifiés, destinés à un public de masse. L'IA élargit l'horizon, permettant la création d'une désinformation hyper-personnalisée à très grande échelle. L'analyse se concentre sur la relation entre les réseaux de bots basés sur l'IA et les algorithmes derrière les plateformes des médias sociaux. Elle montre comment ces outils sont déployés pour supprimer les votes, discréditer les opposants, renforcer les mouvements extrémistes et alimenter la polarisation sociale. Pour faire face à cette menace, une contre-stratégie impliquant plusieurs parties prenantes (gouvernements, entreprises technologiques, société civile) est proposée, favorisant le passage d'une posture réactive à une stratégie proactive axée sur la construction d'une résilience sociétale à long terme.

Mots-clés : écosystème autoentretenu, dissémination, guerre cognitive, propagande pilotée par l'IA.

This paper examines the effect of generative artificial intelligence (AI) on propaganda, which has produced a significant shift from traditional methods. Twentieth-century propaganda was based on simplified driven narratives for mass audiences. AI expands the horizon, allowing for the creation of hyper-personalized disinformation at a very large scale. The analysis focuses on the relationship between AI-based bot networks and the algorithms behind social media platforms. It shows how these tools are actively deployed to suppress votes, discredit opponents, strengthen extremist movements, and fuel social polarization. To address this complex threat, a counter-strategy involving multiple stakeholders (governments, tech companies, civil society) is proposed as a way to shift from a reactive posture to a proactive strategy focused on shaping long-term societal resilience.

Keywords: self-perpetuating ecosystem, dissemination, cognitive warfare, AI-driven propaganda.

Este artículo examina el efecto de la inteligencia artificial generativa (IA) en la propaganda, induciendo un cambio significativo con respecto a los métodos tradicionales. La propaganda del siglo XX se basaba en narrativas simplificadas, destinadas a un público de masas. La IA amplía el horizonte, permitiendo la creación de una desinformación hiper-personalizada a muy gran escala. El análisis se centra en la relación entre las redes de bots basadas en IA y los algoritmos detrás de las plataformas de medios sociales. Muestra cómo estas herramientas se utilizan para suprimir los votos, desacreditar a los opositores, fortalecer los movimientos extremistas y alimentar la polarización social. Para hacer frente a esta amenaza, una contra-estrategia que implica varias partes interesadas (gobiernos, empresas tecnológicas, sociedad civil) se propone, favoreciendo la transición de una postura reactiva a una estrategia proactiva centrada en la construcción de una resiliencia social a largo plazo.

Palabras clave : ecosistema autosostenido, diseminación, guerra cognitiva, propaganda impulsada por la IA.

ISSN : 1923-8541

Introduction

In the contemporary digital landscape, the line between fact and fiction is becoming increasingly blurred, especially now that a new, very potent technological force, generative AI, has entered the scene. We are in the midst a paradigm shift in which propaganda is no longer just a simple distortion of the truth. Now it is much more than that: it is a factory of alternative realities, produced on an industrial scale, and then personalized and adapted to an individual level. AI's capacity to generate complex, nuanced, and highly persuasive content significantly transforms the mechanisms of disinformation. Thus, the conflict now resides on cognitive grounds, impacting the public's trust in information, in institutions, and, even more so, in its own perception of reality.

This study consists of an in-depth analysis of how AI is reconfiguring propaganda. The goal is to identify new manipulation techniques, to see what impact they have on social cohesion, and to identify some potential countermeasures. To achieve this objective, a qualitative research method was used, based on a synthesis of the specialized literature on the topic of AI and propaganda, including academic studies, journalistic investigations, reports from non-governmental organizations, and analyses by technology and public policy experts. By correlating these sources, this paper provides a comprehensive perspective on this phenomenon currently in full swing.

Beginning with the historical context, this study then goes from traditional propaganda to the revolution brought by AI, an instrument so precise that it transforms complexity from a shield of protection into an efficient means of persuasion. Following this arc, this study then addresses the technical tools underlying the new propaganda, going from techniques of hyper-personalization at an individual level to dissemination mechanisms, comprised of botnets and algorithmic amplification on social media platforms.

Thereafter this analysis focuses on exploring these theoretical concepts in practice, using concrete case studies on the use of AI for the purpose of interfering in the election process, causing and maintaining geopolitical conflicts and social instability. It also analyzes AI propaganda's extensive psychological impact, on individuals and on a societal level. Finally, the present paper proposes solutions, focusing on a framework of countermeasures involving governments, tech companies, and civil society. Indeed, this study tries to outline a long-term strategic perspective for a society that is ready to deal with these new faces of disinformation.

Propaganda refers to the deliberate manipulation of the beliefs, attitudes, and actions of a group of people through various techniques and means intended to achieve the manipulators' objectives. Today's electronic media gives propaganda an entirely new dimension. The traditional model of propaganda operated relatively clumsily because it relied on the ability to overwhelm the audience with a large volume of information from an authoritative source. As a result, this model could not easily be adapted to or implemented in new situations, rather it required significant time and resources to rethink and reimplement new strategies. However, these limitations are now being easily overcome by tools based on artificial intelligence.

A novel digital fiend

AI is no longer Science Fiction. AI disinformation is not a matter of projecting dystopian scenarios for potential future threats, it's a present-day reality, as observed in so many recent cases. The World Economic Forum's 2025 Global Risks Report, which surveyed over 900 global leaders,

found that disinformation is the most severe short-term risk, globally, for the second year in a row. (Elsner Atkinson, Zahidi, 2025: 4) So, this is an immediate challenge that brings instability and undermines trust in governance on a global scale. Also, law enforcement agencies are increasingly concerned about the new threats to international security. For instance, Europol's 2025 threat assessment sees AI as a very potent accelerator for organized crime (Europol, 2025), which has a destabilizing effect on society by significantly amplifying current threats. These warnings, coming from key sectors, underline the immediate need for analysis and action.

This urgent need is also fueled by the exponential proliferation of fabricated content, which makes it almost impossible for defensive systems to keep up. The number of deepfake files shared online have grown wildly from 500.000 in 2023 to more than 8 million in 2025, a growth rate of 900% annually. (Khalil, 2025) This has been made possible by the collapse of entry barriers. Recent AI technologies allow for the creation of cheap, convincing cloned voices from just three seconds of audio. For instance, the "deepfake robocall of President Joe Biden used to disrupt the 2024 New Hampshire primary cost just $1 to create and took less than 20 minutes." (Khalil, 2025) The malicious applications of AI are growing at such an alarming rate that it's nearly impossible to create tools to efficiently counteract them in time. A single event, apparently unimportant and insignificant, can create turmoil. A fake AI-generated image of an explosion near the Pentagon in 2023 caused a dip in the U.S. stock market, even if just for a moment. But it was noticeable. (Clayton, 2023) There is, undeniably, an erosion of public trust at an unprecedented level. According to a 2025 Gallup poll, public trust in mass media in the U.S., for instance, was at an historic low of 28% (quite a dramatic fall, when compared to 68% in 1972). (Brenan, 2025)

The confluence of these factors – the global urgency, the exponential growth of fake digital content created with sophisticated intelligent media, the damage to economic stability, and the alteration of social cohesion – outline the central problem of this research. The issue is not the existence of false information, but the development of AI applications capable of creating and disseminating personalized and plausible false information at a scale that endangers our collective ability to discern truth from falsehood. This study addresses the urgent need to understand what lies beneath the current technological developments, what impact they have, and what efficient and potential countermeasures are available to address them.

A paradigm shift

Generative AI brings a paradigm shift in the creation and dissemination of propagandistic messages. The propaganda of the last century emphasized creating slogans accessible to the masses, starting from the simplification of complex realities. AI allows for the creation of complex, context-adapted content, seemingly supported by solid arguments. AI applications can easily combine large amounts of information from various fields, giving messages a substrate that appears well-constructed, thus making the resulting message very convincing (Saab, 2024, 2). "We can say that AI […] can be used to curb mis- and disinformation, to an extent. What this means is that we can create stepwise instructions (or algorithms) or predictive language models that analyze textual data as artifacts of human behavior and then render automated decisions to the system's users." (Moser, 2022, p. 17) Moser's observation is thus framing the issue as a technological double-edged sword. The same systems that create disinformation are also the tools that are supposed to detect and neutralize it.

Generative AI, especially Large Language Models (like ChatGPT), is trained on vast amounts of information, which allows it to learn and develop patterns and structures of human language, thereby generating coherent informational content that is difficult to distinguish from that produced by a human (Sprenkamp *et al.*, 2023). Thus, AI can create, with little human intervention, propagandistic content that is almost as convincing, or perhaps even more convincing, than that produced by humans. AI easily manages to imitate specific elements related to the speaker's tone and official, institutional style, which increases its legitimacy. It perfectly replicates cues humans associate with such traits as authority, or with a specific identity. "Traditional AI models are adept at identifying patterns in large-scale collections of text, audio, and visual data. Generative models extrapolate from those patterns to generate new content using nothing more than simple, natural-language text prompts, image, video, or audio snippets, setting the stage for an even more complex and vexing information landscape." (Saab, 2024, p. 3) As Saab (2024, p. 3) points out, this generative capability stems from generative AI's ability to go beyond simple pattern recognition.

Speed adaptation to ideologies on a large scale

Two important assets of generative AI are speed and scale, as AI programs can generate enormous amounts of information in a very short time. Relatively difficult tasks, which would require a team of trained people, a well-established strategy, and a relatively long period, are solved almost instantly by an AI agent. Thus, a disinformation campaign can take effect in a few hours. "Recent advances in artificial intelligence offer the potential to exacerbate the volume, velocity, variety, and virality of disinformation, automating the process of content creation and the conduct of disinformation campaigns." (Sedova *et al.*, 2021, p. 6) The advantages offered by artificial intelligence include streamlining the process, reducing costs, shortening time, and providing other benefits. On the other hand, artificial intelligence is now available to anyone, so malicious individuals or groups can use its tools to prepare propagandistic content, plan actions, etc., without significant impediments. For example, the Russian-backed propaganda site, DC Weekly, doubled its daily article production after the advent of generative AI, which shows that mass disinformation can now be done with much greater ease. "To the casual user happening upon DC Weekly through an internet search or social media post, it is likely DC Weekly appears genuine. It was involved in successfully laundering over a dozen carefully crafted and entirely fictional narratives, largely about Ukrainian corruption." (Wack *et al.*, 2025, p. 2) This increased production was effective because the content was convincing enough to deceive the public.

Beyond the considerable amount of information it can produce, AI can also approach subjects comprehensively, both in breadth and depth, by expanding its range of resources to cover related fields and varied topics. This can help create a wider variety of content, covering more areas, which can help a propaganda site like the one mentioned above appear more legitimate and credible. The writing style and the nuance of AI-generated texts can be adapted quickly according to specifications. This allows for the creation of content well-molded to ideologies and, more worryingly, its subtle adaptation as the public reacts. (Wack *et al.*, 2025, p. 4)

A new manner of persuading through complexity

The credibility of AI-generated texts is comparable to that of human-written texts. Many readers consider the materials, and the publications that support them, to be equally credible, and make no distinction between them. This fact raises an alarm, indicating the possibility that

propaganda materials that seem authentic to the average reader could be circulated through media channels. This manner of convincing through apparently well-documented, structured, and complex informational content is different from previous propaganda models, which emphasized simplification.

In contrast, AI-generated content persuades through complexity. Propaganda through simple and repetitive messages was most effective, which is why creators of such content did not resort to complex discourses, which would risk reducing the transmission efficiency and credibility of the message. "If in the past attempts at persuasion were made in writing or through direct, face-to-face dialogue, the development of communication technology has made visual persuasion, based on images, possible. [...] Visual persuasion calls on powerful images, which can influence the viewer, and which can convey the right message to persuade the audience to act or think in a certain way. Unlike manipulation, which has a negative connotation, persuasion is understood as a neutral term, relating to the choice between several options." (Farcaș, 2022: 413) AI tools, or tools aided by them, can create complex, nuanced and well-anchored messages in context, which are easily convincing and appear truthful, based on facts and in-depth analysis of situations, but which are, in fact, entirely false. AI has the ability to generate so-called "slopaganda", (Klincewicz *et al.*, 2025) that is information that seems plausible enough to justify a pre-existing belief or desire. The term slopaganda "refers to a combination of a relatively familiar phenomenon (propaganda) and a relatively recent but prominent one (generative AI slop). […] we understand propaganda to be the intentional manipulation of beliefs to achieve political ends. 'Slop' […] refer[s] to unwanted AI-generated content. Putting these together, we understand slopaganda to be unwanted AI- generated content that is spread in order to manipulate beliefs to achieve political ends." (Klincewicz *et al.*, 2025, 137)

Forms of propaganda carried out through AI are characterized by their unprecedented efficiency in personalizing informational content with great precision. This turns these means into what could be called a high-performance instrument. This is the point of departure from established promotion models, whose target was much more imprecisely determined. The latter operated through generalized messages, designed to address the widest possible audience segmented by criteria such as age group, geographical position, or social status, without considering the particular and individual preferences of the audience, on which an overall perspective on a certain product could be built. (Walch, 2024) However, this approach will most often not match the consumer's personal preferences on a more intimate level, thus it will not produce a lasting impression and having no real personal relevance.

On the other hand, AI can operate with large amounts of information, being thus able to compose a detailed and dynamic (updatable, even in real-time) profile of each user. The information on which the user's virtual profile is built comes, in large part, from their interactions in the online environment and in virtual social space, that is, from their very online behavior. Based on this, an instrument as effective as AI can easily generate a psychological profile of the user based on their attitudes and value system, so that, starting from here, it can also create possible scenarios regarding their most plausible future actions. (Garde, 2024) Based on these action variants, creating content molded to the user's preferences, be they commercial or political, is simply an algorithmic matter.

The techniques involved are very versatile, refined, and specific, not limited to a simple demographic classification of users, but involving a much broader depiction of their profiles, resulting from their actions, monitored in real-time, and their interests determined based on the

complex evaluation of multiple factors. AI algorithms allow for the dynamic generation of unique and personal visual content for the user. Content personalization goes as far as prediction does, anticipating the user's preferences even before they consciously manifest them. In the context of AI use, propaganda transforms from a public slogan into a suggestion, a personal recommendation, meant to respond as directly as possible to personal interests. This transformation is all the more effective as it provides the message recipient with an illusion of discovery, with propaganda thus turning into a kind of confirmation of their beliefs and preferences. (Masood, 2025) The message thus becomes discreet, "invisible", which shows its great power: as long as the message transmits information that aligns as effectively as possible with the recipient's horizon of expectations, it is no longer seen as an attack, but rather as a validation. In this way, manipulation becomes much more effective.

Networks of AI-supported bots craft powerful feedback loops

The better tailored the messages are to the measurements of each recipient, the more necessary a corresponding information transmission channel becomes. Usually, to disseminate the propaganda thus created, networks of AI-supported bots are used for message distribution, coupled with social media-adapted algorithms that play the essential role of massively amplifying the scale of propaganda dissemination. AI-managed bot networks are essentially accounts created automatically with the aim of imitating human behavior as faithfully as possible, for the purpose of manipulation. These networks are built using various methods, from utilizing cloud services to exploiting the vulnerabilities of internet-connected devices. They use, for example, geolocation spoofing to make bots appear as real local users. Afterward, the bots are integrated with social media platforms through APIs, thereby giving them user status to post, like, share, and comment. To increase their veracity, they are programmed to interact with other bots, as well as with real human users, often reposting content or adding comments, in the idea of building a history and appearing legitimate. (Smith, 2025)

The integration of Large Language Models (LLMs) was the moment that produced the paradigm shift. Until then, bots were programmed to rely on repetitive templates. Those supported by AI have a much more fertile background, as they are able to call upon AI's ability to use natural language, which in turn allows them to generate credible, plausible, original, context-anchored content that is grammatically correct and semantically coherent. A relevant example in this sense is the "Meliorator" case, in which the United States Department of Justice identified and stopped a so-called Russian bot "farm." They used an AI software called "Meliorator" to create over 1,000 fake American social media profiles, which included convincing details such as profile pictures and other biographical details. "The tool is capable of creating convincing personas in large numbers, using those personas to post credible-sounding information, amplifying messages from other bot personas, and formulating their own messages tailored to the apparent interests of the fake human." (Harding, 2024) The more than 1,000 fake profiles were then used to distribute pro-Russian and anti-Ukrainian information, precisely to show how this method is in fact a readily accessible one.

If these fake, robotic accounts initiate the distribution of content, the algorithms behind modern social media platforms are the ones that have the role of potentiating it, being designed to maximize user engagement, even if the ultimate goal is merely monetary profit, namely generating advertising revenue. Their task is to continuously monitor and analyze user data (likes, comments, shares, but also temporal data regarding post access) in order to select content for promotion.

From this, one can see that these systems place more emphasis on content with an emotional charge, that is polarizing content, or content that aligns with the user's beliefs, without regard for its factual accuracy. This results in a powerful feedback loop, which can lead to the situation that researchers call "algorithmic radicalization" (Busch, 2023) whereby users are gradually exposed to increasingly radical opinions.

This process uses the idea of echo chambers, digital environments in which individual users are repeatedly exposed to the same points of view, being shielded from contradictory perspectives (Ross Arguedas, 2022). In time, users will be increasingly more prone to accepting disinformation. The real-world implications of echo chambers are becoming increasingly evident. For example, certain sources (see The Wall Street Journal) indicate that Facebook's recommendation algorithms are responsible for over 60% of registrations to extremist groups on the platform (Hao, 2021). This shows that the system is vulnerable and can be exploited for malicious purposes. In this scheme, bot networks have the role of producing an initial peak of likes and shares, with the aim of tricking the algorithm into amplifying the propagandistic content and directing it towards a wider audience. A complex system is emerging here, capable of self-regulation and self-maintenance, a propaganda apparatus based on artificial intelligence that functions as a closed-loop feedback mechanism, each element being designed to exploit the others. LLMs produce content adapted to the user's profile in such a way that the audience becomes emotionally engaged. Therefore, the role of AI bots is to produce the initial trigger by imitating local engagement, thus causing the algorithms to trigger the dissemination of information to real users. These users, faced with plausible and provocative content, will interact, providing feedback to the algorithms, which will react by amplifying the message even more. Thus, an automatic and extremely effective system of psychological manipulation appears, which far exceeds previous methods, such as spamming.

Falsified content: cloned voices, deepfake images and false scenarios

AI's versatility makes it a very effective means for building political campaigns, in which a diverse range of AI-generated media content is instrumentalized, containing cloned voices and deepfake images on the basis of which entire false scenarios are constructed to influence the electorate, to denigrate political opponents, and thus to endanger the very integrity of the electoral process. The recent incident from the 2024 American elections became notorious (Yan *et al.*, 2025), when a robocall, convincingly cloning President Joe Biden's voice, urged voters not to participate in the election. Another notorious example occurred in Ron DeSantis's campaign, when AI-generated images depicted the former president at the time, Donald Trump, embracing Dr. Anthony Fauci, a controversial figure, with the aim of creating dissent among Trump's supporters.

This type of action, however, is also occurring on a much larger scale, involving other international powers, such as China (through the network called "Spamouflage") and Russia (through "Doppelganger"), which used AI to create networks of fake accounts on social media platforms, pretending to be American citizens, with the aim of distributing content meant to divide the country (Klepper, 2024). The same Russian-backed propaganda site mentioned earlier, DC Weekly, used AI to produce false information about corruption in Ukraine. The action was successful, as the information was later picked up and adopted by members of the U.S. Congress, showing that the action was successful.

A year earlier, in Slovakia, a few days before the parliamentary elections, a deepfake audio recording appeared in which the leader of the liberal party discussed plans to rig the elections and

increase the price of beer, banking on two sensitive issues for voters. A similar situation occurred in the case of the elections in Taiwan, where deepfake videos promoted false information about the incumbent leader. In India, candidates used AI to create campaign messages in multiple languages, to be able to address more segments of the electorate, but also to distribute manipulated images. In Pakistan, a video of a candidate was modified to tell voters to boycott the vote. In Moldova, a deepfake presented the pro-Western president supporting a pro-Russian party. (Kinnard, 2024)

Of course, many of these attempts to use falsified content were quickly discovered and debunked by journalists and specialists. If the chances of success for such disinformation campaigns are small, then why is there an insistence on these tactics? Does this perhaps indicate a more subtle strategic objective? It is possible that the main purpose of AI-based political propaganda is not to have a short-term effect, but to obtain what is known as the "liar's dividend" ("Hypothesis: In the face of scandal, claims of misinformation (fake news or deepfakes) will increase average support for politicians relative to no response, apologizing, or simply denying a scandal." (Jackson Schiff, 2024)) This phenomenon occurs when the information ecosystem becomes so saturated with potential fakes that the public begins to lose trust in all sources of information, including authentic ones. Even the possibility that an audio or video recording might be a deepfake can be a powerful tool, in that it can help propagandists label authentic content as fake. (Jackson Schiff, 2024) The center of gravity thus shifts from persuasion or the acceptance of false information to inducing a certain negative state among the electorate. A false content produced by AI, even if quickly unmasked, still plants a shadow of a doubt in the voter's mind: any audio or video recording of a politician could be a deepfake. Moreover, the voters will begin to lose trust in the media, in institutions, in democracy. Therefore, the emphasis is on strategy (destroying trust) and not on tactics (changing votes).

AI-generated propaganda is not only applied to the electoral political scene but also transitions into other more sensitive areas, such as geopolitical conflicts. AI tools are very effective in creating and distributing misleading, emotionally charged materials, designed to escalate tensions and support extremist ideologies. Artificial intelligence has become the new front in information warfare. AI-based disinformation was ubiquitous, for example, in the Israel-Gaza conflict. Both the pro-Israeli and pro-Palestinian camps used AI tools and bot farms to spread graphic and emotional propaganda, with the aim of dehumanizing the opponent, producing dissent among international supporters, and exerting pressure on politicians. (Klepper, 2023) An example that shows the complexity of this new environment involved a real photograph of the charred corpse of a baby, distributed by an Israeli source. The image was quickly attacked by online accounts that claimed it was AI-generated, a claim that was itself a form of disinformation. The intention was to discredit authentic evidence by falsely accusing the party in question of using AI. Another situation of this kind occurred in the early period of the war in Ukraine, when President Volodymyr Zelenskyy appeared in a deepfake video asking his troops to surrender. (Wakefield, 2022) Then, a series of attempts were identified from mechanisms supported from the shadows by the Russian state to generate content for fake news sites, with the role of causing the West to reduce the intensity of its efforts to support Ukraine.

AI tools were also used in an attempt to amplify and channel internal social unrest. Such a case occurred in Pakistan, in the moments after the arrest of former Prime Minister Imran Khan, when a multitude of AI-generated images presented various horrors to the public which though quickly discredited, offered a disturbing scenario of widespread government oppression. Another situation occurred in Kenya, on the occasion of mass protests regarding the adoption of

a controversial finance bill. AI was used here in two senses, on the one hand constructively, by activists, to create chatbots and databases with the role of exposing corruption and organizing an opposition movement. (Sami, 2024) On the other hand, maliciously, with bots being implemented to distribute disinformation materials. AI can, therefore, have an ambivalent potential in social movements and can cause palpable changes with great ease. For instance, AI managed to cause a short-term, but significant, drop in the stock market in the United States, starting from a fake image that showed an explosion at the Pentagon. (Bond, 2023)

The most worrying aspect is that AI offers complex and accessible propaganda tools to extremists and terrorist organizations. Online groups on platforms like Telegram have used a music generation tool, Suno AI, to create songs with lyrics promoting antisemitic ideologies. (Theobald, 2025) Members of these groups managed to trick the vigilance of platform moderators by using coded language and intentional misspellings in the instructions given to the system. A Telegram channel affiliated with ISIS used generative AI to produce several fake news bulletins, in which AI-generated news presenters presented false events imitating the style of legitimate media institutions, such as CNN. Being so advanced, the system allowed for the generation of characters, synthesized voice, and synchronized lip movement to create convincing propaganda, instantly translatable into multiple languages. It is clear; therefore, how psychological warfare is being waged on completely different coordinates today. Whereas not long ago sophisticated, large-scale propaganda campaigns were accessible only to state powers with considerable financial and technological resources, modern generative AI tools are accessible and low-cost, being available to anyone. A single user can now easily generate extremist content in minutes, with minimal expense. (Theobald, 2025)

Conclusion

The use of AI algorithms for propaganda is not just a trend that uses the latest technologies, but rather a paradigm shift that will have profound, long-term implications. AI is crafting a new manner of persuading by detecting patterns from data provided by users' virtual communication thereby creating:

- complex solid arguments and coherent informational context adapted to the psychological profile of the user based on their attitudes, beliefs, and value system;
- falsified content and false scenarios by imitating human behavior, institutional style, or speaker's tone through cloned voices and deepfake images;
- polarizing content through emotional charged materials designed to escalate tensions and support ideologies without regard for their factual accuracy;
- feedback loops crafted by networks of AI used for message distribution that offer speed adaptation and credibility to ideologies on a large scale.

This personalized and precise informational content is difficult to distinguish from that produced by a human. The fictional narratives are applied to the electoral political, geopolitical conflicts, etc. has unprecedented manipulation's efficiency.

The accelerated pace at which these practices are becoming more effective is worrying, raising challenges to democratic governance through the ease with which they allow public trust to be eroded. (Sedova *et al.*, 2024) In identifying solutions to overcome the problem, it is not enough to adopt a position of isolating the problem, but rather to take up a strategy of developing

and maintaining, on a social scale, a proactive counterstrategy. AI has become the new front in information warfare.

As the capability of generative AI advances exponentially, so too will the means of propaganda evolve and intensify. Currently, the technology is so advanced that it allows for the creation of realistic, high-quality videos just from text instructions. These AI applications are becoming more accessible, and current countermeasures, such as digital watermarking, may prove insufficient, as they can be altered or removed. (Sedova *et al.*, 2024) AI-based campaigns will likely become a standard element of information warfare. Groups with significant resources such as nation-states will seek to refine these tools and technologies, but AI also produces cheap and handy tools for creating multimedia propagandistic content for extremist groups, terrorists, and even isolated individuals. As the information environment becomes increasingly saturated with convincing fakes, the public's ability to distinguish truth from fiction decreases, and this leads to the establishment of a perpetual state of uncertainty. Perhaps the greatest danger is not an uninformed electorate, but a disengaged society that loses faith in the democratic process as a whole. If trust in institutions, in the media, and in each other is systematically eroded, the social cohesion necessary for the existence of a functional democracy begins to disappear.

To counter the challenges posed by AI technologies, a purely defensive strategy, based on detecting and blocking disinformation, is difficult, perhaps impossible to achieve, given that the scale and speed of content generation will always outpace the abilities of the best human moderator, or even the best content analysis software applications. Therefore, the focus should be on implementing, at the level of the entire society, a multi-level strategy that accepts the inevitability of a polluted information environment and seeks to prepare the public against the effects of this environment. The most effective solutions will be those that encourage critical thinking and that rely on man's ability to discern the truth from a plethora of fakes and deceptions.

Bibliography

Bond, Shannon. "Fake viral images of an explosion at the Pentagon were probably created by AI", [onlone], *NPR*, https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai, published 22 May 2023, accessed 28 Jun 2025.

Brenan, Megan. "Trust in Media at New Low of 28% in U.S.", [online], https://news.gallup.com/poll/695762/trust-media-new-low.aspx, published 2 Oct 2025, accessed 27 Oct 2025.

Busch, Kristen E. "Social Media Algorithms: Content Recommendation, Moderation, and Congressional Considerations", [pdf], *Congressional Research Service*, published 27 Jul 2023.

Clayton, Abené. "Fake AI-generated image of explosion near Pentagon spreads on social media", [online], https://www.theguardian.com/technology/2023/may/22/ pentagon-ai-generated-image-explosion, published 23 May 2023, accessed 27 Oct 2025.

Elsner, Mark, Grace Atkinson, Saadia Zahidi. *Global Risks Report 2025*, published 15 Jan 2025, [pdf], https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025. pdf, accessed 27 Oct 2025.

Europol, *European Union Serious and Organised Crime Threat Assessment – The changing DNA of serious and organised crime*, Publications Office of the European Union, Luxembourg, 2025.

Farcaș, Ana-Daniela. „Imagine și cultură: comunicarea vizuală persuasivă și publicitatea" (Image and Culture: Persuasive Visual Communication and Advertising", in *Buletin științific,* Fascicula Filologie, Seria A, vol. XXXI, 2022, pp. 409-419.

Garde, Sameer. "Driving Performance with Content Hyper-Personalization Through AI And LLMs", [online], *Forbes*, published 23 Feb 2024, accessed 28 Jun 2025.

Hao, Karen. "The Facebook whistleblower says its algorithms are dangerous. Here's why", [online], *MIT Technology Review*, https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/, published 5 Oct 2021, accessed 18 Jun 2025.

Harding, Emily. "A Russian Bot Farm Used AI to Lie to Americans. What Now?", [online], *CSIS* https://www.csis.org/analysis/russian-bot-farm-used-ai-lie-americans-what-now, published 16 Jul 2024, accessed 27 Jun 2025.

Jackson Schiff, Kaylyn, Daniel S. Schiff, Natália S. Bueno. "The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?", [online], *Cambridge University Press*, https://www.cambridge.org/core/journals/american-political-science-review/article/liars-dividend-can-politicians-claim-misinformation-to-evade-accountability/687FEE54DBD7ED0C96D72B26606AA073, published 20 Feb 2024, accessed 15 Jun 2025.

Khalil, Mohammed. "Deepfake Statistics 2025: AI Fraud Data & Trends", [online], https://deepstrike.io/blog/deepfake-statistics-2025, publisher 8 Sep 2025, accessed 27 Oct 2025.

King, Ashley. "Generative Music AI Platform Suno Being Used to Spread Hate", [online], https://www.digitalmusicnews.com/2024/06/20/suno-hateful-music-generated-by-ai/, published 20 Jun 2024, accessed 19 Jun 2025.

Kinnard, Meg. "Election disinformation takes a big leap with AI being used to deceive worldwide", [online], *Quartz*, https://qz.com/election-disinformation-takes-a-big-leap-with-ai-being-1851334182, published 14 Mar 2024, accessed 19 Jun 2025.

Klepper, David. "China-linked 'Spamouflage' network mimics Americans online to sway US political debate", [online], https://apnews.com/article/china-disinformation-network-foreign-influence-us-election-a2b396518bafd8e36635a3796c8271d7, published 3 Sep 2024, accessed 21 Jun 2025.

Klepper, David. "Fake babies, real horror: Deepfakes from the Gaza war increase fears about AI's power to mislead", [online], *AP News*, https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47, published 28 Nov 2023, accessed 22 Jun 2025.

Klincewicz, Michał, Mark Alfano and Amir Fard. "Slopaganda: The interaction between propaganda and generative AI", in *Filosofiska Notiser*, Årgång 12, Nr. 1, 2025, p. 135–162.

Masood, Adnan. "The most dangerous aspect of AI propaganda is its invisibility", [online], UST, https://www.ust.com/en/insights/adnan-masood-the-most-dangerous-aspect-of-ai-propaganda-is-its-invisibility, published 6 May 2025, accessed 28 Jun 2025.

Ross Arguedas, Amy, Craig T. Robertson, Richard Fletcher, Rasmus Kleis Nielsen. "Echo chambers, filter bubbles, and polarisation: a literature review", [onlone], https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literature-review, published 19 Jan 2022, accessed 18 Jun 2025.

Rubin, Victoria L. *Misinformation and Disinformation: Detecting Fakes with the Eye and AI*, Springer, 2022.

Saab, Beatrice. "Manufacturing Deceit. How Generative AI supercharges Information Manipulation", [pdf], 2024.

Sami, Waleed. „The Perilous Role of Artificial Intelligence and Social Media in Mass Protests", [online], *Modern Diplomacy*, https://moderndiplomacy.eu/2024/12/07/ the-perilous-role-of-artificial-intelligence-and-social-media-in-mass-protests/, published 7 Dec 2024, accessed 20 Jun 2025.

Sedova, Katerina, Christine McNeill, Aurora Johnson, Aditi Joshi and Ido Wulkan. "AI and the Future of Disinformation Campaigns" [pdf], published by Center for Security and Emerging Technology, Dec 2021.

Smith, Michael. "AI Plus Social Media Bots = Large-Scale Disinformation Campaigns", [online], *Vercara*, https://vercara.digicert.com/resources/ai-plus-social-media-bots-large-scale-disinformation-campaigns, published 15 Jul 2024, accessed 26 Jun 2025.

Sprenkamp, Kilian, Daniel Gordon Jones and Liudmila Zavolokina. "Large Language Models for Propaganda Detection", în arXiv:2310.06422 [cs.CL], 27 Nov 2023.

Theobald, Emma; Alexis d'Amato, Joey Welles, Jack Rygg, Joel Elson and Samuel Hunter. "Examining the Malign Use of AI: A Case Study Report", [online], *Reports, Projects, and Research*. 126, https://digitalcommons.unomaha.edu/nciterepor tsresearch/ 126, published Apr 2025, accessed 28 Jun 2025.

Wack, Morgan, Carl Ehrett, Darren Linvill and Patrick Warren. "Generative propaganda: Evidence of AI's impact from a state-backed disinformation campaign" [pdf] *PNAS Nexus*, Volume 4, Issue 4, April 2025.

Wakefield, Jane. "Deepfake presidents used in Russia-Ukraine war", [online], *BBC*, https://www.bbc.com/news/technology-60780142, published 18 Mar 2022, accessed 22 Jun 2025.

Walch, Kathleen. "How Generative AI Is Driving Hyperpersonalization", [online], *Forbes*, published 15 Jul 2024, accessed 28 Jun 2025.

Yaojun Yan, Harry, Garrett Morrow, Kai-Cheng Yang, John Wihbey. "The origin of public concerns over AI supercharging misinformation in the 2024 U.S. presidential election", [online], *Misinformation Review*, https://misinforeview.hks.harvard.edu/ article/the-origin-of-public-concerns-over-ai-supercharging-misinformation-in-the-2024-u-s-presidential-election/, published 30 Jan 2025, accessed 25 Jun 2025.