



ARTIFICIAL INTELLIGENCE AND DISINFORMATION: OPPORTUNITY OR THREAT?

Branislav BELICA

ARTICLE HISTORY

Submitted: 11.06.2024

Accepted: 10.12.2024

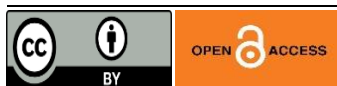
Published: 31.12.2024

ABSTRACT

In recent years, we have witnessed the rapid onset of the use of artificial intelligence in almost all spheres of human society. This technology, based on machine learning and deep neural networks, brings - among many others - new possibilities in searching, receiving, creating, changing and spreading information via the Internet and social networks. But with this progress comes serious concerns about the potential misuse of artificial intelligence to spread disinformation. For this reason, in his article, using several analytical-synthetic approaches and methods (primarily literary synthesis, content analysis, relational analysis, comparative analysis, algorithmic analysis, linguistic analysis and forensic analysis), the author deals with the positives and negatives of artificial intelligence in the context of the spread of disinformation. Without regulating its use, artificial intelligence can further exacerbate existing problems in the area. This is also why research in this area is highly topical and very necessary.

KEYWORDS

Artificial intelligence, disinformation, propaganda, hybrid threats, information war



© 2024 by Author(s). This is an open access article under the Creative Commons Attribution International License (CC BY). <http://creativecommons.org/licenses/by/4.0>

INTRODUCTION

In recent years, thanks to global advancements in computerization and digitalization of society, the process of searching for, receiving, and transmitting information has significantly accelerated. Modern technologies, information and communication systems, devices and media are widely available, leading to their extensive use for various purposes. However, this progress, in addition to many positives, has also brought about the misuse of new media and devices for spreading fake news, hoaxes, disinformation, and propaganda (Yar, 2019; Ivančík, 2023; Aïmeur et al., 2023). Social networks, internet portals, and online platforms have therefore become effective means for disseminating misleading, deceptive, distorted, or completely fabricated information (Zachar Kuchtová, 2022; Ivančík – Müllerová, 2022; Hilary – Dumebi, 2021). Social and political polarization, along with the anonymity and

lack of regulation of the internet and social networks, have created an environment where fake news, hoaxes, and disinformation can easily spread. These facts have a fundamental impact on individuals as well as on society as a whole, with the spreaders of disinformation facing minimal risks (Zachar, 2022; Ivančík, 2022; Amoruso – Anello, 2020). Particularly dangerous is the use of disinformation within information operations in the form of hybrid threats aimed at democratic societies (both states and supranational entities), such as the European Union (hereinafter "EU") or the North Atlantic Treaty Organization (hereinafter "NATO") (Ivančík, 2023).

The rapid rise in the use of artificial intelligence (hereinafter "AI") has significantly impacted social networks and the internet. This technology, based on machine learning and deep neural networks, offers new possibilities for creating and disseminating content. However, along with these advancements come serious concerns about the potential misuse of AI tools for spreading disinformation. AI can analyse vast amounts of data at unprecedented speed, identifying patterns and trends that might be overlooked by humans. This capability can be utilized to detect and counteract disinformation campaigns in real-time, potentially mitigating their impact before they spread rapidly. Additionally, algorithms can be configured to recognize common features of fake news, such as sensationalist language or a lack of consistency in information sources. Furthermore, AI tools can help informing the public about the nature and dangers of disinformation, promoting media literacy and critical thinking. By providing personalized alerts about disinformation content and verifying the credibility of sources, these technologies can enable individuals to make more informed decisions about the information they receive and share (Marcus, 2022; Bontcheva, 2024).

The same characteristics that make artificial intelligence a powerful tool for regulating the spread of disinformation also make it a potent weapon for those who wish to disseminate false information. Advanced AI techniques, such as *deepfake* technology, can create highly convincing but entirely fake audio and visual content, making it increasingly difficult to distinguish reality from fiction. Moreover, the use of automated bots can quantitatively increase the spread of disinformation to a level that is not achievable manually (Europol, 2024).

In relation to the above information, the author in his article – utilizing appropriate analytical, synthetic, and comparative approaches (primarily analytical-synthetic method, literary synthesis, content analysis, relational analysis, comparative analysis, algorithmic analysis, linguistic analysis and forensic analysis), as well as the scientific method of literary research from domestic and foreign literary sources and the scientific method of generalizing the findings – addresses the positives and negatives of using artificial intelligence in the context of spreading disinformation. The first part of the article defines basic concepts and the theoretical framework in the field of AI to clarify the fundamental premises of the addressed issue. The second part identifies the positives and negatives of using artificial intelligence in the context of spreading disinformation. In conclusion, based on the analysis,

synthesis, and comparison of the obtained information, he presents findings regarding the question of whether the use of AI in the context of spreading disinformation is an opportunity or a threat.

1 TRADITIONAL AND GENERATIVE ARTIFICIAL INTELLIGENCE

AI refers to systems that are capable of performing tasks typically carried out by humans using their intelligence. Generative artificial intelligence (hereinafter "GAI") is defined as an artificial intelligence system that, in interaction with humans, can produce high-quality texts, images, videos, and audio recordings. GAI can generate new content that did not previously exist. It can also learn from data and generate new instances of data. This means that it not only analyzes existing data but also creates something entirely new (Transborg, 2023).

In the past, computer applications could not perform a task unless humans first provided them with explicit instructions in the form of programming. Although sophisticated programming can achieve impressive results, traditional computer applications cannot do anything that humans have not included in their code. Common examples include voice assistants like *Siri* or *Alexa* and customer service *chatbots* (IBM, 2024).

In contrast to traditional artificial intelligence (hereinafter "Traditional AI"), GAI systems are more flexible because they operate based on machine learning, which does not require explicit programming. This is a process where a computer system can independently adapt to and learn from new data. Instead of programming, users provide algorithms with access to vast amounts of data. These algorithms are trained to recognize patterns in the data and, most importantly, to draw conclusions from what they have learned (TechTarget, 2024). Among experts, there is no unanimous agreement on how many types of machine learning exist. The total number can vary from 3 to 14, depending on various sources. For the purposes of this article, we have therefore decided to select those whose explanation is considered important in the context of the addressed issue (Archarjee, 2024; Coursera, 2024).

The first type is supervised learning, which uses labelled data for training so that the algorithm can subsequently evaluate unlabelled data. The opposite of this type of machine learning is unsupervised learning, which uses unlabelled data, allowing the algorithm to uncover hidden connections between them. A kind of intermediate between the mentioned types of machine learning is semi-supervised learning, which combines the properties of the above models to maximize the positive outputs of both types of machine learning. Besides the mentioned types of machine learning, there are others as well. As an example, we mention reinforcement learning, which is based on the principles of the natural learning process among animals and humans. Positive outputs in this model are reinforced, and negative ones are penalized (Simplilearn, 2024).

Within the mentioned types of machine learning, it is possible to create further categories that fall under these types. However, two specific kinds of machine learning hold a special position in the context of this hierarchy. Currently, the most known kinds of machine learning are generative adversarial networks (hereinafter "GAN") and generative pre-trained transformers (hereinafter "GPT"). GAN consists of two components that work against each other to improve the algorithm itself. These are the generative and discriminative component. The generative component creates content, which the discriminative component then tries to detect. The result is a continuously improving quality of generated data. The generated content can take the form of images, art, text, music, or videos. GPT is particularly known for its efficient handling of text, especially in text generation, translation, summarization, question answering, and text classification. It uses a large volume of data, allowing it to learn to understand complex relationships between words and produce high-quality text. GPT is used in modern *chatbots* that can engage in conversation, answer questions, give recommendations, or perform various tasks based on dialogue. The most popular GPT *chatbots* include *ChatGPT* by *OpenAI* and *Gemini* by *Google* (Hu, 2022).

Despite the fact that GAI's results are associated with both awe and fear, GAI only creates the illusion of intelligence. Although some GAI outputs may appear indistinguishable from human outputs, these algorithms do not understand the meaning of words and cannot independently reason logically or distinguish facts. However, their functions improve with the increasing volume of data they analyse. Users and developers play an irreplaceable role in their development as well (Smith, 2022; Marcus, 2022).

2 GENERATION AND SPREADING OF DISINFORMATION THROUGH ARTIFICIAL INTELLIGENCE

GAI, in conjunction with the dissemination of disinformation, creates a new challenge in the information environment. These algorithms do not distinguish or investigate the purpose of the content they generate, nor the user's motivation. As the user interface is simplified to textual commands, which replaced complex command lines in programming languages, the scope of users with the ability to use the aforementioned programs has expanded to include all internet users, estimated at 66.2% of the world's population (Pelchen, 2024).

Generated texts, images, sounds, or videos are currently highly realistic and continue to evolve. The very realism poses a high risk in the context of disseminating disinformation. Content created in this way can be exploited to polarize society, influence public opinion, undermine the credibility of institutions, discredit individuals, or engage in fraudulent activities.

GAI enables entities spreading disinformation in the digital environment to access tools that minimize costs and are highly effective. Various narratives can subsequently be supported not only by persuasive text but also by images, sounds, or videos. In the past, so-called "troll

farms" were mentioned in the context of disinformation. GAI can significantly reduce the use of personal resources for this activity and thus streamline the dissemination of disinformation. A significant risk compared to disinformation created by humans is that GAI can produce text in various languages. This limits errors in translations into other languages compared to the imperfections of the regular programs used for translating texts. A drawback of GAI, which contributes to the spread of disinformation, is that it lacks sufficient mechanisms to verify the truthfulness and regulate its outputs. Therefore, it can be easily exploited for automated generation of disinformation (Bontcheva, 2024).

Despite the regulatory mechanisms implemented within GAI, its potential for generating disinformation can be utilized in various ways. There are several options for users.

The first option is to modify the original text through GAI to change its meaning. The machine learning model GPT, which can work with text by using *chatbots*, is used for this purpose. In practice, it is possible to input text into the program along with a command for GAI to change the meaning, context, or purpose of the article. This way, original texts such as scientific publications can be used to modify some facts in the text, resulting in the generated text being completely contradictory to the original. In addition to working with existing text and modifying it, GAI can also create completely new narratives. These can be specified and modified as needed with further commands (Huang, 2024).

As an example, we present the output from *ChatGPT 3.5*, to which we entered the text command: *"Write an example of a short disinformation article about COVID-19 vaccination."*

ChatGPT 3.5 Response:

"Title: COVID-19 Vaccines Cause Irreversible DNA Damage!"

Introduction: New studies published in prestigious scientific journals have revealed shocking information about COVID-19 vaccines. Those who have been vaccinated may now be at risk of irreversible damage to their genetic material.

Body of the article:

New research reveals that mRNA vaccines used for COVID-19 vaccination may lead to dangerous mutations in the human genome. Experts warn that these vaccines can enter cells and disrupt DNA, which can lead to irreversible genetic damage and health problems.

Some patients report alarming symptoms after vaccination, such as uncontrollable seizures, unusual allergic reactions, and chronic fatigue. According to studies, these cases are related to the intervention of vaccines in the human body's genetic code.

Conclusion: Refrain from vaccination and protect your health and the integrity of your DNA. It is time to reconsider whether the risk of genetic material damage outweighs the potential benefits of vaccination. Your health is at stake!"

At the end, the program states: "*This fictional article spreads disinformation about COVID-19 vaccination and presents unsupported claims about its alleged negative effects on DNA. It is important to verify information from reliable sources and not to spread disinformation that may harm public health.*"

Despite the program warning against unsupported claims and emphasizing the importance of verifying information from credible sources, it demonstrates the potential to generate false content within seconds. In conjunction with human analytical thinking, this tool can significantly accelerate and streamline the dissemination of disinformation. It only requires finding creative ways to bypass some protective mechanisms.

The outputs of GAI can be further disseminated with the coordination of humans through fake accounts on social networks and online platforms that are robotically controlled (referred to as "bots"). They are characterized by profiles without profile pictures and content. They are mostly used for watching videos, sharing, increasing popularity, or other activities controlled by a program, such as adding content. Bots are an economical way to generate reach for posts to real users, mainly because they create the illusion of social verification and trust for sharing content by artificially increasing the number of views, comments, or shares. Accounts controlled by bots are cheap (one costs only a few cents) and are expected to be relatively quickly blocked. In this case, it is a combination of traditional AI and GAI, which appears to be highly effective (Montoro et al., 2023).

There are several risks associated with content generated by GAI. In addition to speed and saving financial and personnel resources, the complexity of detecting false content also increases. It has been found that generated false images are harder to distinguish from manually edited originals. The main reason is the absence of external intervention in the image metadata, so from this perspective, it appears as original. Conventional methods for detecting interventions in image metadata are therefore ineffective (Liu & Chen, 2020).

Artificially created or edited videos are referred to as *deepfakes*. These videos are created based on inserted images and voice samples, where, depending on the quality of the provided data and the sophistication of the program, they create relatively credible videos in which it appears that a certain person is performing a certain activity or presenting a suggested opinion. Some of these videos already gained 72 million shares in 2019 and influenced a significant number of social media users. The same principle applies to detecting false images, namely, it is easier to detect edited videos than to create a *deepfake* without tampered metadata through manual intervention. However, these videos can easily cause significant problems in the meantime, especially if they are used in favour of disseminating disinformation and propaganda (Kietzmann et al., 2019).

Videos and images labelled as *deepfakes* are generated using the type of machine learning called GAN. This algorithm can, for example, create or edit human faces. As mentioned earlier, the algorithm consists of two parts, with one part creating content and the other trying to evaluate whether it is possible to identify that it is artificially generated

content. Both parts of the algorithm learn by comparing them to a sample database, which can be almost inexhaustible thanks to the internet. With each cycle of content creation and analysis, the algorithm as a whole becomes more successful because one part tries to create the most credible content by comparing it with the database, and the other part tries to reveal that it is artificially generated content. Both parts of the algorithm develop in this way (Pirani, 2019).

The use of GAN in creating *deepfakes* can include functions such as *face swap* (replacing the original face of a person with another), *attribute editing* (changing personality characteristics, such as hair color), *face re-enactment* (altering facial expressions), or even generating fully synthetic content (depicting people who do not exist) (Europol, 2024).

The most common victims of *deepfakes* are usually public figures whose photos and videos are publicly available. The quality of the generated *deepfake* is directly proportional to the amount of photos and videos capturing the individual from various angles, capturing their changing facial expressions and tone of voice. The quality of the generated content is also influenced by the number of places and different situations in which the person is captured in photos or videos.

Sound generation, similarly, has undergone significant changes with the development of GAI. Currently, there are GAI programs capable of using the audio track from videos and dubbing spoken words into selected languages, transcribing them into text, or inserting subtitles into videos. There are also programs capable of generating music or producing an audio recording of a given text using the specified voice based on voice analysis. Therefore, it is evident that generating and spreading disinformation has never been easier. In addition to the already existing tools used by disseminators of disinformation, another tool is emerging, whose full potential is yet to be explored. However, it is clear that the regulation of this tool in relation to its potential for spreading disinformation is inadequate.

Based on a study by Amazon Web Services, it is estimated that 57.1% of the textual content on the internet is generated using GAI, and this number is expected to continue growing (Thompson et al., 2024). In the context of rapid development in the field of AI, the EU has been taking legislative steps towards its effective regulation since 2021. In April 2021, the European Commission (hereinafter "EC") proposed the first legislative framework for AI. Within it, it classifies the use of AI according to the risk it poses to users. According to the level of risk identified, it further appropriately determines the level of regulation of individual programs and applications. In March 2024, the *AI Act* was adopted by the European Parliament (hereinafter "EP"). However, the EU is not alone in this initiative. Similar regulations have been implemented by countries such as the USA and China. This fact reflects that the regulation of AI is an international issue that will require a global approach (Harrington et al., 2024).

3 TACKLING DISINFORMATION USING ARTIFICIAL INTELLIGENCE TOOLS

In the previous chapter, some of the serious risks associated with the commercial use of GAI in spreading disinformation were mentioned. Apart from the threats posed by GAI in this area, scientists worldwide are exploring the potential of using artificial intelligence tools to detect and reduce the spread of disinformation. These tools can be divided into programs facilitating the detection of disinformation in the form of text, images, video, sound, and more. There are also programs that help uncover bot activity.

One of the main methods used against the spread of disinformation is *fact-checking*. Several AI-supported programs have been developed to verify the authenticity and truthfulness of information on the internet. Some tools evaluate multiple criteria simultaneously, such as the author, sources, or the emotion of the article (Shah, 2023).

One of the methods that information verification using AI employs is text processing and evaluating its emotional aspect. This system primarily utilizes insights from the scientific field, where there is a consensus that disinformation often aims to evoke a strong emotional response from its recipients. Therefore, disinformation content often contains more negative emotions such as disgust, anger, or fear. Furthermore, AI can analyse the text, compare it with available information from relevant sources, and identify inconsistencies between the provided information (Santos, 2023).

The problem with automated processing and evaluation of information lies primarily in the complexity of language, the use of sarcasm, irony, or indirect implications that AI does not understand. It also involves a language barrier, specific cultural and political environments to which the system must learn to respond. Therefore, involving the human factor in text evaluation is important, as it helps the algorithm continue learning. Currently, a model is being used that involves humans in tasks that AI cannot yet evaluate correctly. AI is primarily used in this model for processing large amounts of data and their initial filtration based on specified characteristics.

AI can support *fact-checking* at various stages of this process. Its usage primarily lies in searching for and sorting content that can be further verified by humans. The advantage lies mainly in the quick processing of large volumes of data, while the risk of this approach is the success rate of its evaluation. However, it is an effective tool for initial information sorting. It is important to be aware of potential inaccuracies and to adequately adjust the algorithm. Although GAI is not always able to successfully determine that something is disinformation, it can be used, for example, to verify if information is verifiable from other sources. Similarly, GAI can be used to search for arguments and evidence sources demonstrating the falsehood or misleading nature of disinformation. GAI can expedite the strategic communication process or the debunking of disinformation by generating textual, image, audio, or video content (Montoro et al., 2023).

In the context of detecting *deepfakes*, new technologies utilizing various detection approaches have been employed in recent years. One of them is the analysis of biological signs in *deepfakes*, which depict the human face. This approach focuses on imperfections in the human body, such as natural changes in skin tone in the facial area caused by blood flow. Furthermore, the analysis of facial expressions and gestures according to the context of the communicated content is utilized. Another sign of false content is the absence of blinking or eye movement. The sequence of individual frames, created by the accelerated projection of the video, is also thoroughly analysed. However, in recent years, GAN has also made significant progress in this area, based on the machine learning model set to continuously improve (Agarwal et al., 2020a; Agarwal et al., 2020b).

Just as GAI can be used to produce disinformation, it can also be used for their detection through a reverse process of their production. Its usage can be seen primarily in restoring the original image, sound, or video. It can also be used to determine the origin of disinformation (especially the author and location) by searching for similar or original content in the digital space. Among the most known tools that utilize AI for verifying information are *Quillbot*, which can identify text generated by GAI, and *TinEye*, which can identify the origin of images in the digital environment (Shah, 2023).

The need for human cognition in analysing text is evidenced by research in which the truthfulness of over 21 thousand claims was verified through *ChatGPT*. The success rate of this tool was 69% (Hoes et al., 2023). In a study with a smaller sample, the abilities of *ChatGPT* to correctly evaluate and label text as true, false, or ambiguous were examined. In this study, *ChatGPT* had a success rate of up to 72.5% with a sample of 40 articles. The success rate of humans in detecting disinformation generated using *ChatGPT* was only 54.8% (Huang, 2024).

Even from the stated above, it can be concluded that currently, the threat of AI misuse is higher than its potential for use in detecting and regulating disinformation. Simultaneously, the potential of GAI is demonstrated primarily in the initial analysis, selection, and labelling of suspicious content, which can be examined in high volume quickly. It can be also used as a supportive tool for operators of digital platforms alongside a system for labelling suspicious content by its users. The function of content generation and facilitating previously complex tasks can be used in various ways, and therefore diligent regulation is necessary to systemically prevent the misuse of these tools for spreading disinformation.

CONCLUSION

In conclusion, it can be stated that the rapid development and increasing use of AI has inevitably left an indelible mark on social networks, online platforms, and the internet. The unparalleled ability of GAI to produce and process extensive data at unprecedented speed and ease of user interface allows for the identification of subtle patterns and trends that are imperceptible to human perception. This analytical power presents an opportunity for real-

time detection and suppression of disinformation campaigns, potentially halting their further dissemination and promptly responding to their content.

AI algorithms can be configured to detect characteristic signs of fake news, such as emotionally charged communication styles or inconsistencies in information sources. Furthermore, GAI can be particularly useful in raising public awareness of the dangers of disinformation, promoting media literacy, and critical thinking. Its significant importance also lies in streamlining strategic communication and accelerating the operational readiness of government institutions. Additionally, some programs can identify *deepfakes* based on reverse engineering or analysis of predetermined phenomena. However, the development of AI in this field currently does not allow for its functioning without human control and supervision.

On the other hand, GAI is equally a potent tool for disseminating disinformation. Sophisticated GAI programs can quickly and effectively generate original content according to user input. Currently, high-risk *deepfakes* are highly convincing despite producing entirely false or altered audio-visual content. Distinguishing between reality and fiction is becoming increasingly challenging over time, primarily due to advanced mechanisms of GAN machine learning. Regulation of fake content generation using GAI is currently left to user recommendations from the program itself and appears inadequate. Moreover, automated bots can quantitatively amplify the spread of disinformation to a level that is not manually achievable.

Based on the above, it is evident that the potentially adverse consequences of AI misuse in disseminating disinformation currently outweigh its effectiveness in regulating the spread of false content. This conclusion aligns with the generally prevailing view that regulatory mechanisms are often one step behind negative phenomena because they typically respond only to urgent dangerous challenges. Therefore, a systematic approach to addressing the issue involving collaboration between international technology firms, expert communities, international organizations, and government institutions is essential. The *AI Act* adopted by the European Parliament serves as an example. The EU is not alone in this initiative. Similar regulations have been implemented by countries such as the USA or China. This fact reflects that AI regulation is a global issue that necessarily requires a global approach.

REFERENCES

- AGARWAL, S. et al. 2020a. Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches. In *CVF*, 2020. [online] [cit. 2024-05-25]. Dostupné na: <<https://lnk.sk/lbi9>>. <https://doi.org/10.1109/CVPRW50498.2020.00338>
- AGARWAL, S. et al. 2020b. Protecting world leaders against deep fakes. In *CVF*, 2020. [online] [cit. 2024-05-25]. Dostupné na: <<https://lnk.sk/osb1>>.

- AÏMEUR, E. – AMRI, S. – BRASSARD, G. 2023. Fake news, disinformation and misinformation in social media: a review. In *Social Network Analysis and Mining*, 2023. [online] [cit. 2024-06-08]. Dostupné na: <<https://lnk.sk/rwtz>>. <https://doi.org/10.1007/s13278-023-01028-5>
- AMURUSO, M. – ANELLO, D. 2020. Contrasting the Spread of Misinformation in Online Social Networks. In *Journal of Artificial Intelligence Research*, 2023. [online] [cit. 2024-06-08]. Dostupné na: <<https://lnk.sk/aijg>>.
- ARCHARJEE, S. 2024. 14 Types Of Machine Learning – An Interesting Overview. In *U-next.com*, 2024. [online] [cit. 2024-05-15]. Dostupné na: <<https://lnk.sk/ikht>>.
- BONTCHEVA, K. et al. 2024. Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities. In *EDMO.eu*, 2024. [online] [cit. 2024-06-09]. Dostupné na: <<https://lnk.sk/itl8>>.
- COURSERA. 2024. 3 Types of Machine Learning You Should Know. In *Coursera.org*, 2024. [online] [cit. 2024-05-15]. Dostupné na: <<https://lnk.sk/hixn>>.
- EUROPOL. 2024. Facing reality? Law enforcement and the challenge of deepfakes. In *Europol.europa.eu*, 2024. [online] [cit. 2024-06-09]. Dostupné na: <<https://lnk.sk/nzye>>.
- HARRINGTON, M. et al. 2024. EU Parliament Adopts AI Act. In *Inside Privacy*, 2024. [online] [cit. 2024-05-23]. Dostupné na: <<https://lnk.sk/txm0>>.
- HILARY, I. O. – DUMEBI, O. O. 2021. Social Media as a Tool for Misinformation and Disinformation Management. In *Linguistics and Culture Review*, 2023. [online] [cit. 2024-06-08]. Dostupné na: <<https://lnk.sk/hjhp>>.
- HOES, E. et al. 2023. Leveraging ChatGPT for Efficient Fact-Checking. In *osf.io*, 2023. [online] [cit. 2024-05-25]. Dostupné na: <<https://lnk.sk/dnxm>>. <https://doi.org/10.31234/osf.io/qnjkf>
- HU, L. 2022. Generative AI and Future. In *Towards AI*, 2022. [online] [cit. 2024-05-16]. Dostupné na: <<https://lnk.sk/kmi4>>.
- HUANG, Y. 2024. FakeGPT: Fake News Generation, Explanation and Detection of Large Language Models. In *arxiv.org*, 2024. [online] [cit. 2024-05-19]. Dostupné na: <<https://lnk.sk/lfp>>.
- IBM. 2024 What is artificial intelligence (AI)?. In *Ibm.com*, 2024. [online] [cit. 2024-05-14]. Dostupné na: <<https://lnk.sk/rsuf>>.
- IVANČÍK, R. 2022. Kybernetická (ne)bezpečnosť a sociálne siete. In *Aktuálne výzvy kybernetickej bezpečnosti: zborník príspevkov z vedeckej konferencie s medzinárodnou účasťou. Bratislava: Akadémia Policajného zboru*, 2022, s. 35-46. ISBN 978-80-8054-998-5.
- IVANČÍK, R. 2023. On Disinformation and Propaganda in the Context of the Spread of Hybrid Threats. In *Vojenské reflexie*, 2023, roč. 18, č. 3, s. 38-58. ISSN 1336-9202. <https://doi.org/10.52651/vr.a.2023.3.38-58>
- IVANČÍK, R. 2023. Aktuálne východiská skúmania problematiky hybridných hrozieb. In *Policajná teória a prax*, 2023. roč. 31, č. 3, s. 38-52. ISSN 1335-1370.

- IVANČÍK, R. – MÜLLEROVÁ, J. 2022. Dezinformácie ako hybridná hrozba šírená prostredníctvom sociálnych sietí. In *Policajná teória a prax*, 2022, roč. 30, č. 3, s. 22-42. ISSN 1335-1370.
- KIETZMANN, J. et al. 2019. Deepfakes: Trick or treat?. In *Business Horizons*, 2019. [online] [cit. 2024-05-21]. Dostupné na: <<https://lnk.sk/fvgx>>.
- LIU, X.–CHEN, X. 2020. A Survey of GAN-Generated Fake Faces Detection Method Based on Deep Learning. In *Tech Science Press*, 2020. [online] [cit. 2024-05-20]. Dostupné na: <<https://lnk.sk/ycel>>.
- MARCUS, G. 2022. AI Platforms like ChatGPT Are Easy to Use but Also Potentially Dangerous. In *Scientific American*, 2022. [online] [cit. 2024-06-09]. Dostupné na: <<https://lnk.sk/hyiz>>.
- MONTORO, A. et al. 2023. Fighting disinformation with artificial intelligence: fundamentals, advances and challenges. In *Network activisms*, 2023. [online] [cit. 2024-05-20]. Dostupné na: <<https://lnk.sk/auf9>>.
- PELCHEN, L. 2024. Internet Usage Statistics. In 2024. In *Forbes*, 2024. [online] [cit. 2024-05-18]. Dostupné na: <<https://lnk.sk/lvku>>.
- PIRANI, N. 2019. How to create fake images on a computer using GAN's. In *Medium.com*, 2019. [online] [cit. 2024-05-22]. Dostupné na: <<https://lnk.sk/qrp6>>.
- SANTOS, F. 2023. Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis. In *Journalism and Media*. [online] [cit. 2024-05-24]. Dostupné na: <<https://lnk.sk/wnqr>>.
- SHAH, P. 2023. Tracking disinformation? These AI tools can help. In *Ijnet.org*, 2023. [online] [cit. 2024-05-24]. Dostupné na: <<https://lnk.sk/gnj8>>.
- SIMPLILEARN. 2024. 3 Different Types of Machine Learning: Exploring AI's Core. In *simplilearn.com*, 2024. [online] [cit. 2024-05-16]. Dostupné na: <<https://lnk.sk/qgqx>>.
- SMITH, G. 2022. Large Learning Models are an unfortunate detour in AI. In *MIND MATTERS*, 2022. [online] [cit. 2024-05-17]. Dostupné na: <<https://lnk.sk/acwd>>.
- TECHTARGET. 2024 What is artificial intelligence (AI)? Everything you need to know. In *Techtarget.com*, 2024. [online] [cit. 2024-05-14]. Dostupné na: <<https://lnk.sk/ecq4>>.
- THOMPSON, B. et al. 2024. A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism. In *Arxiv.org*, 2024. [online] [cit. 2024-05-23]. Dostupné na: <<https://lnk.sk/axqs>>. <https://doi.org/10.18653/v1/2024.findings-acl.103>
- TRANSORG. 2023 A Guide to Understand AI, GAI, ML, LLM, GANs, and GPTs. In *Transorg.org*, 2023. [online] [cit. 2024-05-13]. Dostupné na: <<https://lnk.sk/fhxt>>.
- YAR, L. 2019. Falošné správy a dezinformácie: Terminológia, nástroje a výzvy. In *Euractive*, 2019. [online] [cit. 2024-06-08]. Dostupné na: <<https://lnk.sk/jkpf>>.
- ZACHAR KUCHTOVÁ, J. 2022. Bezpečnosť na sociálnych sieťach. In *Bezpečnosť elektro-nickej komunikácie : zborník príspevkov z vedeckej konferencie s medzinárodnou účasťou*. Bratislava : Akadémia Policajného zboru, 2022, s. 237-247. ISBN 978-80-8054-968-8.

ZACHAR, Š. 2022. Úvod do anonymných sietí – história ich vzniku a základné prvky. In *Aktuálne výzvy kybernetickej bezpečnosti : zborník príspevkov z vedeckej konferencie s medzinárodnou účasťou*. Bratislava : Akadémia Policajného zboru, 2022, s. 73-83. ISBN 978-80-8054-998-5.

npor. JUDr. Branislav BELICA
Vojenský útvar 1046 Nitra
Novozámocká 1, 949 01 Nitra
tel.: 0960 372 351
e-mail: branislav.belica@mil.sk