THE ROLE OF AI IN CREATING FALSE REALITIES TO MANIPULATE THE MASSES

UDC: [004.8+004.932]:32.019.5(478)
DOI: https://doi.org/10.53486/tids2025.20

VICTORIA LOZAN
PhD, Associate Professor
Academy of Economics Studies of Moldova
lozan.victoria@ase.md
ORCID ID: 0000-0002-5869-8515

ION COSTIN
IT Engineer
Vocational School from Soroca, Republic of Moldova
ioncostyn1@gmail.com
ORCID ID: 0009-0008-4540-3256

Abstract. Recently, products based on Artificial Intelligence (AI) technology have been used to create photo and video content aimed at influencing public opinion, especially during election periods. Alongside healthy use for enhancing creativity in the film industry, advertising, or journalism, offering new possibilities for artistic expression, and creating personalized content tailored to the specific needs of different audiences, AI platforms are also being used to generate content designed to influence public opinion. This is an extremely relevant and complex issue that raises numerous ethical, social, and political questions. Content creation platforms are being used to spread conspiracy theories, damage people's reputations, or manipulate electoral outcomes. In some cases, the falsified content is difficult to detect. Deep fakes can create scenes or false statements that appear authentic, which can lead to the spread of misinformation. AI-generated content can be used to amplify certain messages or ideologies, reinforcing social and political divisions. It is essential that the public is informed about the existence of these technologies and educated on critically evaluating the content they consume. Scientists and tech companies are working on developing algorithms capable of detecting AI-generated fake content. Dedicated tools, relying on complex algorithms, are used to create deep fakes. As the technology for generating fake content becomes more advanced, it becomes increasingly difficult for the public to distinguish between reality and fiction. Another dilemma is accessibility; as AI technology becomes more accessible, the costs of generating fake content will decrease, and the ability to produce such materials will become available to an ever-growing number of individuals. The purpose of this paper is to analyze AI-generated content (wefaceswap.com) and determine the aspects that define the fictitious nature of the material. While AI offers enormous opportunities for innovation and creativity, it also comes with significant risks related to misinformation, manipulation, and the erosion of trust in the media.

Keywords: Artificial Intelligence, deepfake, content, fiction, reality, manipulation, polarization.

JEL Classification: C80, C88, C92, O33, O36.

INTRODUCTION

The rapid development, over the last decade, of technology based on artificial intelligence and its implementation in various fields has substantially facilitated human activity. The use of modern technologies reduces the time required to complete certain activities, improving the final result. Technologies based on artificial intelligence are used in transport management (Abirami, *et al.*, 2024),

enabling traffic flow forecasting, enabling congestion management and intelligent routing, increasing efficiency in various facets of modern transportation systems. Due to its potential to improve functionality, decision-making, and efficiency, AI is used in Internet of Things systems (Khadam, *et al.*, 2024). The technology also makes a considerable contribution to software-based industrial networks, smart manufacturing, logistics, supply chains, construction and 5G/6G networks (Rojas, *et al.*, 2024). AI-based methods are implemented for software development and testing (Amalfitino, et al., 2023) and in engineering (Martínez-Fernández, Bogner and Franch). AI-based methodologies, algorithms, data sources, results, diagnoses are also applied in the medical field (Martinez-Millana, et al., 2022). AI-based technologies are also being implemented in self-driving vehicles, multimedia recognition, cybersecurity, space exploration, genetics, climate change, agriculture (Goel, *et al.*, 2023). AI is also being implemented in wireless drone networks for trajectory optimization, radio resource management, routing and topology control, edge and cache calculation, and improved security and privacy. (Zhou, *et al.*, 2024).

A valuable use of artificial intelligence technology is seen in the creative industry, where it enables content generation, information analysis, optimization of materials and post-production workflows, as well as efficient data extraction, enhancement and compression (Anantrasirichai & Bull, 2022).

This paper analyzes existing platforms for creating deepfakes, the way such content is created, and the impact of the spread of these products.

DEEPFAKE PRODUCTS AND MASS MANIPULATION

1. Deepfake generation applications

Deepfake technologies allow users to create convincing images and videos, generated by artificial intelligence, that mimic the appearance, sound, and mannerisms of real people. Currently, there is a wide range of platforms that allow the creation of deepfakes and any user is able to use them to create different projects, even to generate fake video or audio content. No advanced knowledge is required, just select the required platform, subscription type and device hardware configuration to ensure efficient running of the web application. Of course, desktop applications are also available. A description of the most popular platforms for creating deepfake can be found on (Taylor, 2025). The paper will analyze the possibilities of creating deepfakes using the (Wefaceswap, 2025) platform which offers a *Creator package* for $19.99/month (80 credits: 80 swapped images or 120 seconds of swapped video can be made) and the Microsoft Clipchamp app.

Deepfake is a type of synthetic media in which AI, specifically deep learning algorithms, are used to manipulate or generate visual and audio content that closely resembles real people, making it appear authentic. Some apps work by swapping the subject's face with existing media, others can create new images and video avatars from scratch.

Experts estimate that over 500,000 deepfake videos and voice recordings were shared on social media by the end of 2023 (Kaur, *et al.*, 2024).

Deepfake apps offer valuable tools for innovation and effective content creation. For example, applications are used in entertainment to enhance film production, creating special effects or “reducing the aging” of actors; in marketing, deepfakes personalize advertisements and attract customers with realistic avatars; in education and training, by simulating real-life scenarios for improved learning; for language translation, by synchronizing lip movements with translated audio for more localized content; in content creation, thanks to technology for rapidly producing personalized media content.

2. The process of creating deepfakes and ethical considerations

The deepfake creation process goes through several stages. The first stage involves collecting high-quality images and videos of people whose faces are to be changed. This data must cover a wide range of facial expressions and positions to ensure a successful deepfake. The second stage consists of using computer vision techniques to detect and isolate faces from the collected images and videos. Accurate face detection ensures precise results for the next steps. The third stage involves aligning the detected faces so that they have a similar position and orientation. Correct alignment is important for achieving a realistic deepfake, as it ensures that machine learning algorithms accurately analyze and reproduce facial features and expressions. In the fourth stage, facial features and expressions are encoded into a mathematical representation, important for training machine learning algorithms. The fifth stage consists of training machine learning algorithms. Depending on the complexity of the product and the capacity of the dataset, this process can take several hours or days to complete. The creation of the deepfake is completed by swapping faces in the video, using the encoded facial data to manipulate the video and produce a realistic deepfake.

The creation of a deepfake must be accompanied by a clear statement that the image or video was generated using a deepfake application or an Artificial Intelligence system. Transparency is important in the process of creating and distributing deepfakes. Maintaining an ethical framework in the process of using technology allows avoiding misleading the public. As long as deepfakes are not used to manipulate, defame, deceive, or violate the rights of others, they can be considered legal.

In the case of using photos, videos or voice recordings belonging to other people in the creation process, it is necessary to obtain their explicit consent. Lack of consent may lead to a violation of copyright, image rights or privacy, thus attracting legal consequences. Social platforms, such as Facebook/Meta, have begun to impose clear rules on labeling AI-generated content to combat misinformation and abuse.

With the implementation of AI technologies, companies face various technological, organizational and cultural challenges (Ångström, *et al.*, 2023).

3. Deepfakes detection

Depending on the platform and resources used, the quality of deepfakes varies. To identify a deepfake, an analysis of unnatural facial movements, awkward or infrequent blinking, and inconsistent expressions is required. Inaccurate lighting or shadows and blurry areas, especially around the face, can also indicate manipulation. Lip sync issues, where the mouth doesn't match the speech, are also an indication of manipulation. Another aspect would be unnatural reflections in the eyes or excessively smooth skin textures. Additionally, audio mismatches, such as voices not aligning with lip movements, can be a clear indication. These visual and audio inconsistencies are key indicators of a fake video created using deepfake technology.

In parallel with the evolution of tools for creating deepfakes, methods and techniques for their detection are being researched and identified. In the case of complex and qualitative deepfakes, the analysis of the above aspects in detecting manipulation will not be successful. An important direction is to evaluate the performance of learning-based deepfake detectors in more realistic contexts, quantitatively measuring their robustness against different processing operations (Lu & Ebrahimi, 2024). An in-depth analysis of state-of-the-art techniques and tools for identifying deepfakes, encompassing image-based, video, and audio content can be found in (Sunil, *et al.*, 2025). Fundamental technologies, such as deep learning models, are explored and their effectiveness in

differentiating real from manipulated environments is evaluated. Additionally, new detection methods are being explored that use sophisticated machine learning, computer vision, and audio analysis techniques.

The integration of Vision Transformers with a DenseNet-based neural feature extractor is presented in (Siddiqui, *et al.*, 2025). The authors claim that "the approach produces results equivalent to the latest Vision Transformer techniques, without relying on complex tactics such as distillation or ensemble methods. Furthermore, the inference technique is simple yet effective, which uses a voting mechanism to identify numerous faces in a single video frame."

Deepfake detection can be achieved through a multi-scale interactive dual-stream network (MSIDSnet). The network is divided into spatial and frequency domain streams and uses a multi-scale fusion module to capture both the facial features of images that have been manipulated in the spatial domain under different circumstances and the fine-grained information about the high-frequency noise of the falsified images. The network fully integrates the features of the streams in the spatial and frequency domains through a dual-stream interactive module and uses the vision transformer (ViT) to further learn the global information of the faked facial features for classification (Cheng, *et al.*, 2024).

Another direction in detection is the examination of spatial and temporal properties. Fake videos disrupt statistical regularity in original videos. Therefore, the generalization of deepfake detection is stimulated by distinguishing regularity disruption that does not appear in real videos. Perturbing a real video with a Pseudo-Fake Generator creates a wide range of pseudo-fake videos for training. Such a practice allows for deepfake detection without using fake videos and improves generalization ability in a simple and efficient way. To capture spatial and temporal perturbations together, a spatiotemporal enhancement block is proposed to learn regularity perturbation in space and time on self-generated videos (Guan, *et al.*, 2023).

There are other detection techniques that are researched in specialized articles, such as the approach to datasets, algorithmic techniques and challenges, useful for an overall analysis; methods based on large models and biometrics; classification of machine learning and deep learning methods, and hybrid ones; adversarial robustness, real-time processing and evaluation metrics.

4. Manipulation of public perception

The evolution of deepfake creation technology comes with both positive and negative aspects, the resulting products can directly influence collective perception and behavior. Advances in generative technologies allow the production of hyper realistic audio-visual content, which can mislead the audience by creating artificial realities that are difficult to distinguish from authentic ones. This content has the ability to simulate the voice, facial expressions and behavior of real people, thus becoming a potentially dangerous tool for information manipulation.

Recent research (Ranka, *et al.*, 2024; Jungherr, *et al.*, 2024) indicates that the use of AI in the generation of false information can have a significant impact on democratic processes, including elections, where the manipulation of public perception can be decisive. In addition, the AI integration in areas such as digital journalism, social media, online education and political campaigns facilitates the amplified distribution of falsified narratives, often difficult to identify even by specialists. Experiments with real subjects show that audio deepfakes are harder to detect than video ones and that human discernment depends on the environment in which they are presented (Groh, *et al.*, 2024).

In this context, rigorous analysis of the technical and socio-cognitive mechanisms through which AI can contribute to the creation and dissemination of false realities is necessary (Tugarev, 2023). It is also important to assess the impact of these technologies on social cohesion, public trust and the democratic functioning of society. This study aims to investigate how artificial intelligence is used to generate and distribute misleading content, to identify how disinformation through deepfakes can polarize public opinion in electoral campaigns and diminish final scores.

Deepfakes are no more disinformation than regular fake news, but they can affect public perception and contribute to polarization, especially among the less digitally literate. Nationally, the website https://stopfals.md is a portal that publishes and exposes forgeries appearing in the digital environment. The independent press from the Republic of Moldova also informs citizens about the deepfakes circulating on the Internet. Next, we will analyze the impact of the deepfake that appeared on October 16-17, 2024, which has President Maia Sandu as its main actor, on the eve of the presidential elections and the Constitutional Referendum in the Republic of Moldova. The video is currently not available to the public, so reference will be made to Detector Media, which translated the Romanian audio track used in the fake video (Koldomasov & Pylypenko, 2024), where the message of the President of the Republic of Moldova is contained, who *"repents"* for her position, campaigns against herself and calls for a *"no"* vote in the referendum: *"The Republic of Moldova will not be accepted into the EU even in 30 years. I know for sure that once the war in Ukraine ends, we will be next. If I remain president, I will have to ban all independent media and open criminal cases against opposition politicians. Vote for any candidate, just don't vote for me or for this fake Euro-referendum. My defeat is the only chance to save the life of the Republic of Moldova and your children."* This deepfake appeared and was distributed simultaneously on several local pro-Russian Telegram channels ("Republica Gagauza", "Prydnestrovets" and "112.md — Moldovan News 24/7"), disguised as campaign material for Maia Sandu.

Watchdog, between August 20-23, 2024, conducted a survey on a sample of 1011 people aged 18 and over. The respondents were asked the question: *"But if you had to choose in a referendum between the accession of the Republic of Moldova to the European Union (EU) and the accession of the Republic of Moldova to the Eurasian Economic Union (EEU), what would you choose?"*. The results are shown in Figure 1.

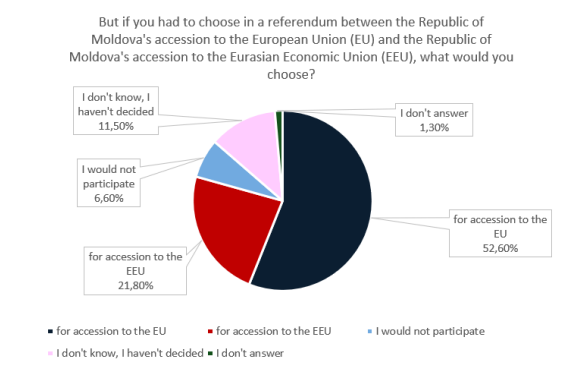


Figure 1. Watchdog survey results.

Source: *https://watchdog.md/news/208261/creste-impactul-campaniilor-de-dezinformare-si-influentare-a-opiniei-publice-in-republica-moldova/*

In early October, the Institute for Public Policy (IPP) conducted a survey on a sample of 1,100 people, in which respondents were asked to express their opinions on the foreign policy options of the Republic of Moldova. When asked whether they would support the accession of the Republic of Moldova to the European Union, 54.5% declared that they would vote in favor. The results of the survey are presented in Figure 2.

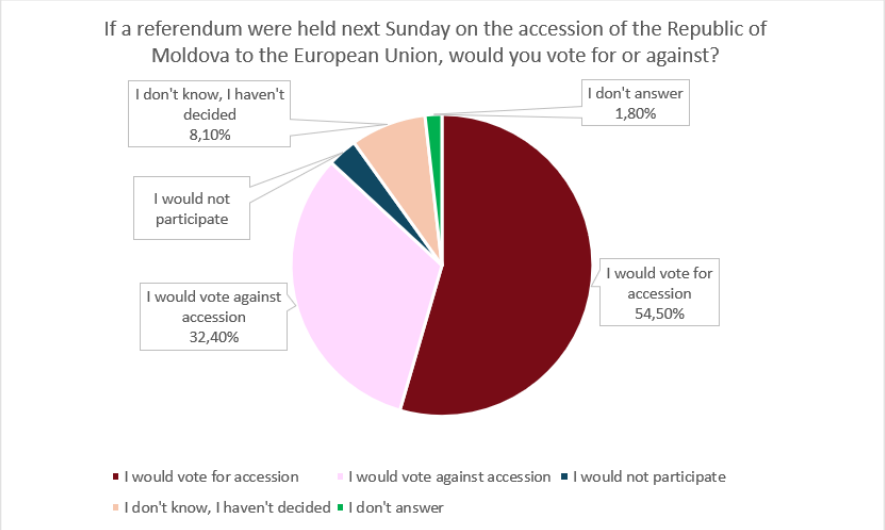


Figure 2. Results of the survey conducted by the Institute for Public Policy.
Source: *http://bop.ipp.md/ (Accessed: 5.9.2025)*

The results of the Republican Constitutional Referendum of October 20, 2024 on supporting the amendment of the Constitution in view of the accession of the Republic of Moldova to the European Union are available on (CEC, 2024), 50.35% of voters voted for and 49.65% - against. If the votes from the diaspora are excluded, then internally there are 45.38% - for and 54.62% - against. A very large gap compared to the results of the survey conducted in early October 2024 by IPP, about 9.12%.

The conclusion may be subjective without concrete evidence, but the polarization of public opinion and the numbers of the Referendum results can be seen. Directly or indirectly, it is a consequence of the deepfake that appeared on the eve of the elections.

Analysis of surveys conducted by (IPP, 2024) attests to a significant increase in 2024 in the use of social networks as a source of information for the population of the Republic of Moldova. The most popular are Facebook, Instagram, Telegram, TikTok, YouTube, odnoklasniki.ru, Viber, Vkontakte. Each person, thanks to the implementation of AI-based algorithms, is provided with certain content depending on their interests. Algorithms analyze browsing history and place individuals in an information bubble specific to their preferences. This process is done subtly by algorithms without the subject's awareness.

5. Making deepfakes with Wefaceswap

To better understand the generation process and analyze the result obtained, deepfake images and videos will be created using the Wefaceswap platform and the Microsoft Clipchamp application. The images and videos available on the website https://ase.md/ and the institution's Facebook page will be used as data sources. Another source is the video promoting Vocational Technical Education in the Dual System available on the official Youtube channel. Figure 3 presents the sources used.

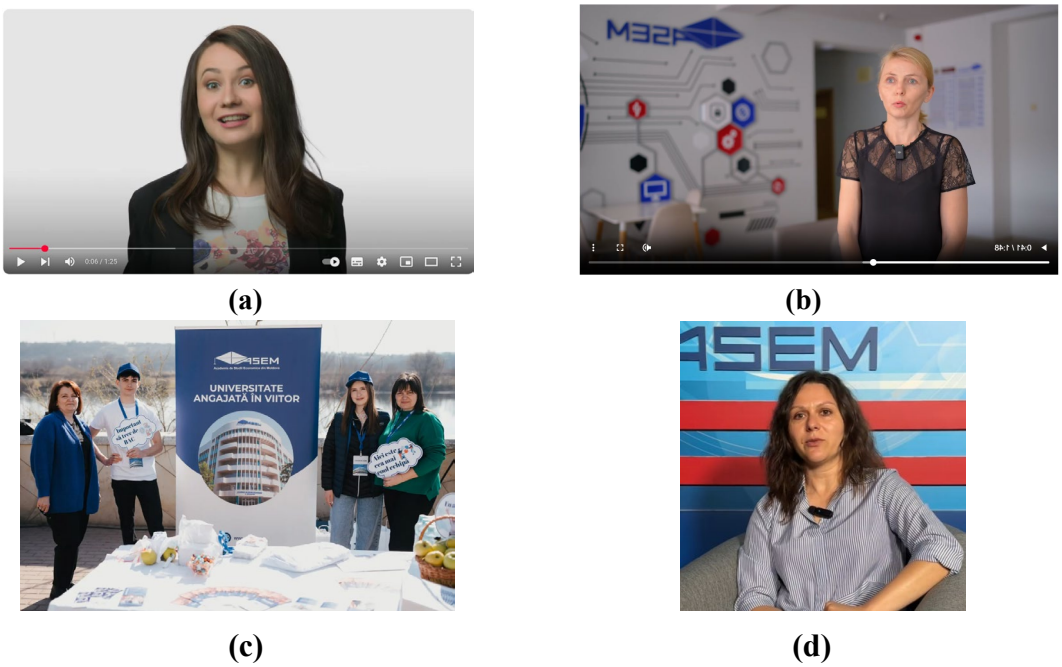


Figure 3. Data sources for generating deepfakes.

Source: ***(a)*** *https://www.youtube.com/watch?v=Nm4QBGQ2cHc*
(b) *https://ase.md/la-asem-a-fost-deschis-un-laborator-pentru-studierea-sistemului-informational-geografic/*
(c) *https://ase.md/asem-la-targurile-educationale-regionale-din-causeni-si-soroca/*
(d) *https://www.facebook.com/reel/767532952118987?locale=ro_RO*

The first deepfake will be created, replacing the face of the student in figure 3.c with the face in figure 3.d. The result is shown in figure 4.



Figure 4. Deepfake of changing face in an image.

Source: *Made by the author on the Wefaceswap platform based on the sources in figure 3.c and figure 3.d.*

As can be seen in Figure 4, the deepfake image appears authentic, although if the original images are examined, a difference in age is observed.

The second deepfake will replace the face in the video in figure 3.a with the face in figure 3.d. As a result, the resulting product can be considered successful, if there were no obvious synchronization errors where the face in the video remains the original one. Figure 5 shows some screenshots from the deepfake made.



Figure 5. Screenshots from the Deepfake video Promoting Dual Education.

Source: *Made by the author on the Wefaceswap platform based on the sources in figure 3.a and figure 3.d.*

In this case, due to the rapid movements of the person in the original video, accurate face identification is not possible. To achieve a successful and high-quality deepfake, more powerful platforms and a larger dataset for the person in figure 3.d are needed.

The following deepfake consists of modifying the face in figure 3.b with the face in figure 3.d. Screenshots of the obtained deepfake are presented in figure 6.



Figure 6. Screenshots from the Deepfake video Launch of the GIS Lab.

Source: *Made by the author on the Wefaceswap platform based on the sources in figure 3.b and figure 3.d.*

As can be seen from Figure 6, the face in the deepfake does not correspond to the one in Figure 3.d. A new person was obtained that has features specific to the subject in the original video and is additionally assigned facial features corresponding to the subject in Figure 3.d. Due to the slow movements in the original video, no obvious manipulations can be detected, although in the close-up frames, vibrations are observed in the face – an indication of manipulation.

From the deepfakes made, it can be concluded that to obtain a qualitative result, a platform that performs complex manipulations is needed. As data sources, it is necessary to use people of the same age, stature and appearance. It would be advisable for the original videos, the inertia of the subjects who will be subjected to manipulation.

CONCLUSIONS

The development of AI-based technologies has seen a remarkable rise in recent times, especially deepfake platforms. The products made are becoming more and more authentic and difficult to detect. Therefore, there is a need to establish methods for detecting deepfake. In some cases, detecting manipulation is simple and can be done through detailed analysis of the deepfake video, in other cases dedicated applications are needed.

Deepfakes can influence public opinion based on misleading information, thus undermining trust in the political system, mass media, etc. In case the accessed content arouses suspicions, it is necessary to investigate its source and the channels on which it is distributed. It should not be ignored

that deepfake products also have positive aspects, streamlining content creation for marketing, the film industry, education, social media, etc.

Anyone can create a deepfake, without any deep IT knowledge. It is important to indicate this fact when creating a material using a deepfake application or artificial intelligence. If a person's photos are used to create a deepfake, their permission is required, otherwise legal problems may arise.

REFERENCES

1. Abirami, S., Pethuraj, M., Uthayakumar, M. & Chitra, P., 2024. A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system. *Case Studies on Transport Policy,* Volume 17, article 101247.
2. Amalfitino, D. et al., 2023. Artificial Intelligence Applied to Software Testing: A Tertiary Study. *ACM Computing Surveys,* 56(3), pp. 1-38.
3. Anantrasirichai, N. & Bull, D., 2022. Artifcial intelligence in the creative industries: a review. *Artificial Intelligence Review,* Volume 55, pp. 589-656.
4. Ångström, R. et al., 2023. Getting AI Implementation Right: Insights from a Global Survey. *California Management Review,* 66(1), pp. 5-22.
5. CEC, 2024. *Comisia Electorală Centrală a Republicii Moldova.* [Online] Available at: https://pvt12024.cec.md/cec-template-referendum-results.html [Accessed 9 5 2025].
6. Cheng, Z., Wang, Y., Wan, Y. & Jiang, C., 2024. DeepFake detection method based on multi-scale interactive dual-stream network. *Journal of Visual Communication and Image Representation,* Volume 104, article 104263.
7. Goel, S., Guha, A., Kuppusamy, U. & Shanmugam, T., 2023. Survey on Artificial Intelligence and Its Applications. In: *Selvaraj, H., Chmaj, G., Zydek, D. (eds) Advances in Systems Engineering. ICSEng 2023. Lecture Notes in Networks and Systems.* vol 761: Springer, Cham, pp. 512-522.
8. Groh, M. et al., 2024. *Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video.* [Online] Available at: https://arxiv.org/pdf/2202.12883 [Accessed 10 5 2025].
9. Guan, J. et al., 2023. *Detecting Deepfake by Creating Spatio-Temporal Regularity Disruption.* [Online] Available at: https://arxiv.org/pdf/2207.10402 [Accessed 10 4 2025].
10. IPP, 2024. *Barametrul Opiniei Publice Republica Moldova.* [Online] Available at: http://bop.ipp.md/ [Accessed 9 5 2025].
11. Jungherr, A., Rauchfleisch, A. & Wuttke, A., 2024. *Deceptive uses of Artificial Intelligence in elections strengthen support for AI ban.* [Online] Available at: https://arxiv.org/pdf/2408.12613v1 [Accessed 5 2 2025].
12. Kaur, A., Noori Hoshyar, A., Saikrishna, V. & al., e., 2024. Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review,* Volume 57, article 159.
13. Khadam, U., Davidsson, P. & Spalazzese, R., 2024. Exploring the Role of Artificial Intelligence in Internet of Things Systems: A Systematic Mapping Study. *Sensors,* 24(20), article 6511.
14. Koldomasov, A. & Pylypenko, A., 2024. *"Europe stole Moldova": Russian propaganda about the elections in Moldova.* [Online] Available at: https://en.detector.media/post/europe-stole-moldova-russian-propaganda-about-the-elections-in-moldova?utm_source=chatgpt.com [Accessed 6 5 2025].
15. Lu, Y. & Ebrahimi, T., 2024. Assessment framework for deepfake detection in real-world situations. *EURASIP Journal on Image and Video Processing,* Volume 6.
16. Martínez-Fernández, S. et al., 2022. Software Engineering for AI-Based Systems: A Survey. *ACM Transactions on Software Engineering and Methodology,* 31(2), pp. 1-59.
17. Martinez-Millana, A. et al., 2022. Artificial intelligence and its impact on the domains of universal health coverage, health emergencies and health promotion: An overview of systematic reviews. *International Journal of Medical Informatics,* Volume 166, article 104855.

18. Ranka, H. et al., 2024. *Examining the Implications of Deepfakes for Election Integrity.* [Online] Available at: https://arxiv.org/pdf/2406.14290 [Accessed 13 5 2025].
19. Rojas, E. et al., 2024. A Survey on AI-Empowered Softwarized Industrial IoT Networks. *Electronics 2024,* 13(10), article 1979.
20. Siddiqui, F., Yang, J., Xiao, S. & Fahad, M., 2025. Enhanced deepfake detection with DenseNet and Cross-ViT. *Expert Systems with Applications,* Volume 267, article 126150.
21. Sunil, R. et al., 2025. Exploring autonomous methods for deepfake detection: A detailed survey on techniques and evaluation. *Heliyon,* 11(3), article e42273.
22. Taylor, M., 2025. *12 Best Deepfake Apps for Realistic Face Swaps.* [Online] Available at: https://akool.com/blog-posts/5-best-deepfake-apps-for-realistic-faceswaps [Accessed 8 05 2025].
23. Tugarev, L., 2023. Deepfake - manufacturing the reality with the aid of the Artificial Intelligence. *Moldoscopie,* 1(98), pp. 182-187.
24. Wefaceswap, 2025. *WEFACESWAP.* [Online] Available at: https://www.wefaceswap.com/ro [Accessed 12 4 2025].
25. Zhou, L. et al., 2024. A Comprehensive Survey of Artificial Intelligence Applications in UAV-Enabled Wireless Networks. *Digital Communications and Networks*.