УДК 304.32.364-1

DOI 10.32840/cpu2219-8741/2024.4(60).14

I. Kyianytsia

PhD, associate professor
associate professor of department of journalism and advertising
e-mail: y.kyyanytsya@knute.edu.ua, ORCID: 0000-0002-9629-9865
State University of Trade and Economics,
19 Kyoto St., Kyiv, Ukraine, 02156

D. Fayvishenko

Doctor of Economics, Professor
Head of the Department of Journalism and Advertising
e-mail: fayvishenko.ds@gmail.com, ORCID: 0000-0001-7880-9801
State University of Trade and Economics
19 Kyoto St., Kyiv, Ukraine, 02156

ARTIFICIAL INTELLIGENCE ON GUARD AGAINST THE HARMFUL EFFECTS OF DEEPFAKES

The purpose of the study *is to outline the threats posed by modern technologies for creating deepfakes, to confirm the need for legal regulation of their spreading, and to make relatable proposals for recognizing deepfakes at the everyday level, in particular through the use of artificial intelligence.*

Research methodology. *The materials used to prepare this article were compiled using a combination of theoretical and empirical methods, including the analysis of sources that offer information about the role of deepfakes in the media environment and their impact on society as a whole. The analysis of foreign websites with legislative acts made it possible to systematize these sources on the relevant issues, as well as to strengthen the argument for the need to legally regulate the dissemination of such falsifications. The use of these methods, as well as the inductive generalization of the field under study, contributed to the structuring of the necessary material to obtain the relational basis for recognizing deep audiovisual counterfeits.*

Results*. A list of rules that can be used to recognize deepfakes is proposed, and a list of online resources for their detection is reviewed and systematized in order to increase the overall level of media literacy and awareness of threats that negatively affect the mental health of society.*

Novelty. *As a result of the analysis of the sources, as well as their systematization and generalization, recommendations for strengthening critical thinking among the population are proposed, and the need for visual training is emphasized in order not to be deceived by another example of a deep audiovisual fake.*

Practical significance. *The proposed rules can be used both for widespread use in society to develop critical thinking and for the development of a set of competencies and programmatic outcomes in media education disciplines.*

Key words: *deepfake, disinformation, manipulation, media addiction, media literacy, artificial intelligence*

I. Introduction

In today's environment of rapid development of information and communication technologies, there are both positive dynamics in the creation and dissemination of information and negative consequences accompanying these processes. Positive aspects include expanding opportunities for disseminating information of various genres and formats, including audiovisual, artistic, cultural and intercultural products that contribute to the formation of a global information society. At the same time, however, there is a rapid increase in the volume of disinformation, much of which is fake news and deepfakes. In particular, the development of deep neural network-based deepfakes is a significant factor in manipulating the media space, undermining trust in information sources.

The issues of building and recognizing deepfakes have a high rating among scientists, particularly in the fields of artificial intelligence, machine learning, and media literacy. The foundation for the creation of deepfakes was the work of American researcher I. Goodfellow, who proposed the GAN (Generative Adversarial Networks) approach as a basis for generating realistic images. It is the creation of deep fakes that directs scientists from around the world to analyze their impact on society and identify their unreliability. For example, T. Chu and A. Zadrovic study deepfakes using specialized image and video analysis algorithms, focusing on artifact recognition, while S. Jou and L. Li conduct research in the field

of recognizing anomalies in video and audio fakes, analyzing linguistic and behavioral features and developing methods to identify «flaws» in voice and facial expressions that are difficult to imitate by artificial intelligence, which allows for a more accurate distinction between the real and the generated. It should be noted that when studying the nature of deepfakes, we should also focus on the ethical and social consequences of their creation and dissemination, emphasizing the need for regulation and ethical responsibility when developing algorithms for deepfakes. These are the issues that are being raised in Ukraine in the context of information security, cyber hygiene, and media literacy, contributing to an understanding of the threats posed by deepfakes, as well as the development of technological and educational measures to recognize them. Currently, in Ukraine, A. Boyarchuk, T. Gordienko, O. Ihnatenko, V. Kudriavtsev, Y. Nikitina have studied the impact of deepfakes on public consciousness and media literacy, analyzed the structural characteristics of images and videos, and developed methods to counter deepfakes and fake news, emphasizing the importance of both technical and educational aspects in the fight against it. Thus, Ukraine is actively involved in the global process of combating digital fraud through the development of technologies, increasing information literacy and media hygiene among the population. Studies show that fake news is used to create realistic falsifications that threaten the perception of reality and are not only part of disinformation campaigns, but also a potential challenge to media literacy, national security, and social stability. It should be noted that modern disinformation campaigns are created not only by media professionals and amateurs, but also by artificial intelligence, which is a significant factor in the growth of not only information threats but also reputational risks for businesses, organizations in various fields of activity, and society as a whole. Such discrediting increases distrust in the veracity of any audiovisual content and creates an environment where it becomes difficult to distinguish truth from manipulation, which is the key goal of disinformation, the structure of which is a multi-level system that uses various means of influencing society, technologies and strategies to achieve manipulative goals and, as a result, undermines trust in the media and promotes information chaos.

Having realized the multifaceted nature of disinformation in general, as well as the growing role of audiovisual communication, we can state that it is the deepfakes that pose the greatest threat to public consciousness, especially due to the strengthening of artificial intelligence technologies, which not only reduces the level of consumer responsibility, but also reduces the ability to recognize technically created fakes and deepfakes

II. Problem statement and research methods

Taking into account the growing problem of spreading disinformation through fakes and deepfakes, as well as its interdisciplinary nature by covering not only technical but also social, psychological, legal and ethical aspects, the purpose of this paper is to try to propose relational solutions that will help support the information ecosystem, aimed at strengthening trust between senders and receivers of information, as well as replenishing the media literacy portfolio with tools for detecting deepfakes based on the recognition of small errors in video and audio, observations of facial expressions and movements, as well as image quality analysis, atypical angles of the face, etc.

So far, by researching media materials containing deepfakes, we have tried to identify their typical characteristics and distribution channels, and analyzing video and audio has made it possible to identify mechanisms for verifying the authenticity of information, which will help increase the level of media literacy of the population and develop a comprehensive content verification system. By crystallizing the foundations for content verification, we also tried to find artificial intelligence-based relational methods that would allow us to quickly verify the authenticity of media files before they are distributed to the media. In our opinion, such research will contribute to the strengthening of ethical standards to enhance the sustainability of communication structures in the digital age.

III. Results

Nowadays, deepfake is a high-tech image synthesis technique based on the use of artificial intelligence and machine learning. The rapid spread of this technology among Internet users raises significant concerns based on the real risks of its misuse. Deepfake technology is developing much faster than experts predicted, and its potential for harm seems to be quite significant. The trend towards increasing quality and realism of fake content indicates that in the near future it will be difficult to distinguish between videos depicting real people and those fabricated by computer algorithms [6].

Analyzing the evolution of this phenomenon, it is important to note that the first deepfakes appeared in 2017, when a Reddit user with the pseudonym DeepFace posted pornographic videos depicting famous Hollywood stars who did not actually participate in these videos [1, p. 16]. Since then, the significance of such a tool for manipulating information has been growing every day. Using algorithms similar to those developed by NVIDIA to change seasonal frames in videos, the user trained neural networks on a large number of images, which significantly improved the quality of fake videos. However, this process required significant computational resources and time to create millions of images used to train generative models, and the first versions of the deepfakes were relatively easy to detect due to their unrealistic nature, but technological advances have significantly improved the accuracy and realism of such materials. In particular, modern generative models can not only fake videos and images, but

also reproduce voices with high accuracy. This is where it is important to realize all aspects (both positive and negative) of neural networks, which today can synthesize the speech of different people, creating significant risks to the security and trust in information in the media space, but at the same time, can expose existing fakes in media content.

The growing spread of deepfakes and their ability to misinform society requires both technological and regulatory mechanisms to counteract them. On the one hand, technological solutions should focus on the development of algorithms for detecting fakes by analyzing images and audio at the pixel level, as well as the introduction of digital signature systems to verify the authenticity of content. On the other hand, it is necessary to improve the legal framework for regulating the use of AI technologies to prevent their use for harmful or criminal purposes. Recognizing the threat of deepfakes to democratic processes, many countries pay attention to the regulation of media content that may contain deepfakes, in particular, the government of the state of California has adopted legislation that prohibits the use of deepfakes 60 days before elections in order to protect political processes from their negative impact [7], also in Texas, the state passed a law prohibiting the use of deepfakes for fraud and false accusations aimed at undermining confidence in the electoral process, and in general, the United States is constantly working on the «Deepfake Accountability Act», which requires labeling of synthetic media and provides for punishment for its malicious use [10]. In contrast, in the European Union, the distribution of deepfakes is regulated under general laws, such as the Digital Services Act, which requires online platforms to monitor content and remove malicious deepfakes by applying the General Data Protection Regulation (GDPR), which can be applied to deepfakes if they violate a person's privacy [9]. Similar measures have been taken in China, which also adopted a law regulating the use of deepfake technologies, confirming the global awareness of the risks of this technology to national security and social stability [9]. Also, the Indian government is developing legislative initiatives to regulate malicious content created with the help of deepfakes, in particular to protect against information falsification and fraud [2]. These examples demonstrate the global response to the risks of data breaches in the context of democratic political processes, cybercrime, and privacy.

In Ukraine, the date of signing the Declaration on the Future of the Internet (28.04.22) can be considered the beginning of the formation of a comprehensive regulatory mechanism to counteract dangerous phenomena that pose risks to the information security of society [3]. Understanding the significant risks that cause disruption of the Internet, such as partial or complete disconnection, network fragmentation, a surge in cyberattacks, the spread of online censorship and disinformation, it is digital services that must comply with international human rights standards and ensure freedom of expression, encouraging pluralism. However, the question arises as to the specific mechanisms for filtering illegal content and the extent of responsibility of platforms and users in this process, as the fine line between censorship and diversity of opinion makes it difficult to apply this principle. Thus, despite the existence of the Declaration aimed at combating computer crimes, in particular manipulation of electoral processes through deepfakes and other disinformation technologies, our country remains at the initial stages of implementing such legislation.

Taking into account the above arguments, we can conclude that the problem of creating and distributing deepfakes calls into question the ethical and legal foundations of modern media institutions, raising serious concerns, as the possibility of using deepfakes to manipulate information can threaten public security and political stability. Accordingly, it is very important not only to develop ethical standards that would ensure the responsible use of artificial intelligence-based technologies aimed at protecting citizens from disinformation, but also to develop and implement legal norms and technological tools to identify deepfakes and reduce their harmful impact on society.

At the moment, we believe that household recommendations for improving media literacy and developing critical thinking are also very important. Given that deepfakes are characterized by a number of features, including the absence or unnatural blink rate, distorted facial expressions, and head movements that do not correspond to the typical movements of the person depicted, when viewing audiovisual content, it is worth considering anatomical features such as the edges of the face, nose, and lip area, which may be areas where anomalies that give off deepfakes can be noticed. Given the rapid development of deepfake technologies, it is likely that in the near future there will be videos in which even experienced users will have difficulty distinguishing between real people and computer-modeled characters. In view of this, it is necessary to outline key characteristics that will help identify fake content, in particular: the quality of audiovisual content, which is higher in original videos than in fabricated ones:

– quality of audiovisual content – original videos are usually characterized by high image and sound quality, which captures details, natural movements and the absence of visual artifacts, while fake videos are often of lower quality, as computer algorithms face significant computational difficulties in reproducing realistic small details, such as skin texture, reflective elements and small movements of facial muscles;

– partial closure of the face or showing it at an acute angle – in fake videos, the face is mostly shown at a certain angle or partially closed, as neural networks that generate deepfakes have difficulty

reproducing realistic details from different angles and facial angles are often used to avoid complex movements that are difficult to model;

– disproportionate scaling of the face – camera movement contributes to the depiction of abnormal changes in the proportions of computer-modeled faces, which is associated with the difficulty of processing the scaling of objects at different angles and can manifest itself as changes in the size of individual parts of the face, which contradict the laws of perspective and are a sign of deep faking;

– unnatural boundaries between real and artificial parts of the face – when overlaying a face or other elements in a deepfake, clear or blurred boundaries between real and modeled parts may be observed, often demonstrated by signs of «transition lines» or differences in shades and texture at the boundaries of the joint;

– abnormal skin tone during facial movements – arises from difficulties in matching light and color characteristics during movement, as neural networks are not always able to properly adapt skin tones to changing lighting conditions and angles, resulting in unnatural shadows or abnormal shades in artificial areas;

– inconsistency between sound and lip movements – asynchrony between facial expressions and speech, when the overlay of audio on video may be inaccurate, resulting in a lack of synchronization between sound and lip movements and, as a result, the effect of «delayed» or «accelerated» facial expressions that do not match the sound, indicating the use of algorithms to replace the original face or sound.

These simple recommendations require care and critical thinking, and it is also important to check the source, time and circumstances of the recording, as deepfakes are often distributed without context or with minimal information, which makes it difficult to verify their authenticity, so in cases where it is impossible to visually identify a fake, modern technological solutions such as deepfakes recognition software become an effective tool. One of them is Deepware, a software that analyzes video footage using machine learning algorithms to detect artificially created or modified images and videos, where at the initial stage, a neural network learns basic image characteristics such as textures and pixel patterns, which are often disturbed by digital manipulation. Next, it analyzes parameters such as the location of eyes, nose, smile angles, head movements, etc. in more depth, as these details can vary significantly in the modeled content; and finally, it calculates the probability that the content is fake, providing the user with information in the form of a percentage. Deepware's neural network, trained on large amounts of authentic data, not only helps users to notice the slightest discrepancies between real and synthetic images, but also allows them to identify possible editing and assess the level of trust in the analyzed content, which is especially valuable for media, legal and research organizations that need to verify the authenticity of video materials. The key aspect of this technology is that it can both detect the presence of fakes and determine the percentage probability of manipulation, which allows users to more accurately assess the degree of authenticity of content [13]. But this is not the only software that works on the basis of various data analysis algorithms and can be useful for detecting disinformation content. Currently, there are tools for detecting information anomalies, including:

– AI or Not – uses artificial intelligence to analyze both images and videos to determine whether an image has been generated or altered using deep learning algorithms, evaluating various factors such as synthetic artifacts and fake details to give a result on the authenticity of the image;

– FotoForensics – analyzes images to detect changes and manipulations using the Error Level Analysis (ELA) technique, which allows you to see differences in the compression levels of different parts of the image;

– AI Voice Detector – analyzes audio recordings to detect voice fakes created by artificial intelligence, using techniques to compare voice models and synthetic samples to detect anomalies in speech, tone, or pronunciation;

– Resemble Detect – uses machine learning to identify synthetic voices by analyzing the characteristics that may be inherent in deepfake technologies, such as unnatural intonation, timbre, and rhythm of the voice;

– Error Level Analysis (ELA) – works by analyzing the levels of compression errors in different parts of the image, which allows you to identify any changes or superimposed elements that do not have the original image structure and any differences in the error level may indicate the presence of editing or alteration [12].

Based on models that mimic the structure and functioning of the human brain, using artificial neural networks to solve complex problems, this software is extremely effective when working with unstructured data, such as images or audio. That is why it is necessary to constantly replenish the portfolio of tools for working with deep learning, in particular through libraries and frameworks that allow you to apply or create your own approaches to building neural networks. Among the most well-known are PyTorch, TensorFlow, and Keras, which are actively used to design and train models for recognizing media manipulation. The availability of such libraries simplifies the creation of adaptive and accurate algorithms that can detect anomalies in audio and video files that may indicate the presence of deepfakes.

Given the above software and available data libraries for detecting manipulation in digital media, it is also worth paying attention to additional sources for verifying information, as the accuracy of detecting deepfakes depends not only on technical means but also on the context in which the content is distributed. It is important to understand that manipulative media can be associated with political or social motives that depend on the region, and the context of its origin should be taken into account. Also, the reputation of a website or platform can indicate the level of risk of manipulation or spreading false information. Using simplified tools such as Google Reverse Image Search or TinEye to find the original sources of content can also help identify possible fake or manipulated content and help determine whether a video or image has been previously published in a different context or has come from different sources. Thus, in addition to using technological tools to verify content, it is worth paying attention to a multifactorial assessment, which includes the reputation of sources and the country of origin of content, to minimize the risks of disinformation and manipulation.

IV. Conclusions

Deep fakes or deepfakes pose a serious threat to society, as they can effectively manipulate public opinion by falsifying images, voices of famous people, and introducing regionally colored attributes. Due to these aspects, their ability to spread falsifications that threaten not only democratic processes but also the rule of law in general is growing dramatically. In today's information environment, when artificial intelligence algorithms are constantly improving, detecting fakes is becoming increasingly difficult, as they become extremely realistic and impossible to separate from authentic content.

To combat such manipulative technologies, it is not enough to use only digital content authentication tools, such as programs that can detect modifications in video in real time. Thus, it is necessary to develop and implement systems that can provide an effective assessment of the authenticity of content, analyzing both visual and audio features to determine whether it has been manipulated or falsified. In our opinion, it is also very important to implement educational initiatives aimed at developing media literacy, user awareness, and improving the skills of critical analysis of online content. The primary safeguards against the influence of digital fakes are a developed ability for visual and auditory analysis, an understanding of digital manipulation techniques, and the skill to evaluate the credibility of information.

All of this highlights the need for active research and ongoing cooperation in this vector between scientists, technology companies, and governments to improve both tools for detecting manipulation and educational programs for a wide audience. Joint efforts can help strengthen law and order and protect democratic processes in the context of the modern digital era, where a conscious consumer of information, using their own knowledge and artificial intelligence, can detect deepfakes and prevent their negative consequences.

Список використаної літератури

1. Вальорска М. А. Діпфейк та дезінформація : практ. посіб. / пер. з нім. В. Олійника. Київ : Академія української преси ; Центр вільної преси, 2020. 36 с.
2. Діпфейки та ШІ у передвиборчій агітації в Індії: експерти побоюються, що технології можуть змінити вибори в усьому світі. URL: https://zn.ua/ukr/TECHNOLOGIES/dipfejki-ta-ii-u-peredviborchij-ahitatsiji-v-indiji-eksperti-pobojujutsja-shcho-tekhnolohiji-mozhut-zminiti-vibori-v-usomu-sviti.html (дата звернення: 15.10.2024).
3. Подобний О., Слатвінська В. Діпфейк в контексті декларації про майбутнє інтернету. *Юридичний науковий електронний журнал.* 2022. № 5. С. 594–596. URL: http://lsej.org.ua/5_2022/142.pdf (дата звернення: 25.10.2024).
4. Юртаєва К. В. Кримінологічний аналіз використання технології Deepfake: коли фейк стає злочином. *Вісник кримінологічної асоціації України*. 2021. № 1 (24). С. 31–42.
5. Bahar M., Sharmin A. Deep insights of deepfake technology: A review. URL: https://www.academia.edu/76656464 (date of request: 15.10.2024).
6. Chesney R., Citron D. Deepfakes and the new disinformation war. The coming age of post-truth geopolitics. URL: https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war (date of request: 21.10.2024).
7. Cole S. California's deepfake law aims to ban the technology for election misinformation. *Vice News*. URL: https://www.nytimes.com/2024/09/17/technology/california-deepfakes-law-social-media-newsom.html (date of request: 20.10.2024).
8. Declaration for the Future of Internet. URL: https://digital-strategy.ec.europa.eu/en/library/declaration-future-internet (date of request: 15.10.2024).
9. Hine E., Floridi L. New deepfake regulations in China are a tool for social stability, but at what cost. *Nature Machine Intelligence*.
10. H.R.3230 – deep fakes accountability Act 116th Congress (2019–2020). URL: https://www.congress.gov/bill/116th-congress/house-bill/3230/text (date of request: 15.10.2024).
11. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC. URL: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng (date of request: 15.10.2024).

12. Tolosana R., Vera-Rodriguez R., Fierrez J., Morales A., Ortega-Garcia J. DeepFakes and beyond: A survey of face manipulation and fake detection. URL: https://www.researchgate.net/publication/338355353_DeepFakes_and_Beyond_A_Survey_of_Face_Manipulation_and_Fake_Detection (date of request: 20.10.2024).
13. Westerlund M. The emergence of deepfake technology: A review. *Technology Innovation Management Review*. 2019. № 9. P. 40–53. URL: https://timreview.ca/sites/default/files/article_PDF/TIMReview_ (date of request: 18.10.2024).

Reference

1. Valorska, M. A. (2020). *Dipfeik ta dezinformatsiia* [Deepfake and disinformation] (V. Oliinik, Trans.). Kyiv: Akademiia ukrainskoi presy ; Tsentr vilnoi presy [in Ukrainian].
2. Dipfeiky ta ShI u peredvyborchii ahitatsii v Indii: eksperty poboiuiutsia, shcho tekhnolohii mo-zhut zminyty vybory v usomu sviti. Retrieved from https://zn.ua/ukr/TECHNOLOGIES/dipfejki-ta-ii-u-peredviborchij-ahitatsiji-v-indiji-eksperti-pobojujutsja-shcho-tekhnolohiji-mozhut-zminiti-vibori-v-usomu-sviti.html [in Ukrainian].
3. Podobnyi, O., & Slatvinska, V. (2022). Dipfeik v konteksti deklaratsii pro maibutnie internetu [Deepfake in the context of a declaration about the future of the internet.]. *Yurydychnyi naukovyi elektronnyi zhurnal,* 5, 594–596. Retrieved from http://lsej.org.ua/5_2022/142.pdf [in Ukrainian].
4. Iurtaieva, K. V. (2021). Kryminolohichnyi analiz vykorystannia tekhnolohii Deepfake: koly feik staie zlochynom [Criminological analysis of the use of Deepfake technology: When a fake becomes a crime]. *Visnyk kryminolohichnoi asotsiatsii Ukrainy,* 1 (24), 31–42 [in Ukrainian].
5. Bahar, M., & Sharmin, A. (2021). Deep insights of deepfake technology: A review. Retrieved from https://www.academia.edu/76656464 [in English].
6. Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war. The coming age of post-truth geopolitics. Retrieved from https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war [in English].
7. Cole, S. (2019). California's deepfake law aims to ban the technology for election misinformation. *Vice News*. Retrieved from https://www.nytimes.com/2024/09/17/technology/california-deepfakes-law-social-media-newsom.html [in English].
8. Declaration for the Future of Internet. (2022). Retrieved from https://digital-strategy.ec.europa.eu/en/library/declaration-future-internet [in English].
9. Hine, E., & Floridi, L. (2022). New deepfake regulations in China are a tool for social stability, but at what cost. *Nature Machine Intelligence* [in English].
10. H.R.3230 – deep fakes accountability Act 116th Congress (2019–2020). Retrieved from https://www.congress.gov/bill/116th-congress/house-bill/3230/text [in English].
11. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC. Retrieved from https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng [in English].
12. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2019). DeepFakes and beyond: A survey of face manipulation and fake detection. Retrieved from https://www.researchgate.net/publication/338355353_DeepFakes_and_Beyond_A_Survey_of_Face_Manipulation_and_Fake_Detection [in English].
13. Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9, 40–53. Retrieved from https://timreview.ca/sites/default/files/article_PDF/TIMReview_ [in English].

Стаття надійшла до редакції 25.11.2024.

Received 25.11.2024.

Кияниця Є. О., Файвішенко Д. С. Штучний інтелект на варті проти шкідливого впливу дипфейків

Мета дослідження *полягає в окресленні загроз, які становлять сучасні технології створення діпфейків; підтвердженні необхідності правового регулювання їх розповсюдження, а також релятивних пропозицій щодо розпізнавання дипфейків на побутовому рівні, зокрема за допомогою штучного інтелекту.*

Методологія дослідження. *Під час опрацювання матеріалів, на основі яких було підготовлено цю статтю, використано комплекс теоретичних та емпіричних методів, зокрема аналіз джерел, в яких запропоновано інформацію про роль дипфейків у медійному просторі та їх вплив на суспільство в цілому. Аналіз закордонних сайтів, на яких представлені законодавчі акти, дав змогу систематизувати зазначені джерела з актуальної проблематики, а також посилити аргументацію щодо необхідності правового регулювання розповсюдження таких фальсифікацій. Використання зазначених методів, а також індуктивне узагальнення досліджуваного поля сприяло структуруванню необхідного матеріалу для отримання релятивних засад розпізнавання глибоких аудіовізуальних підробок.*

Результати. *Запропоновано перелік правил, використання яких може стати в нагоді при розпізнаванні дипфейків, а також розглянуто й систематизовано перелік онлайн-ресурсів для їх виявлення з метою підвищення загального рівня медіаграмотності та обізнаності про загрози, які негативно впливають на ментальне та психічне здоров'я суспільства.*

Новизна. *У результаті проведеного аналізу джерел, а також їх систематизації й узагальнення запропоновано рекомендації щодо розвитку критичного мислення населення, а також акцентовано на необхідності зорових тренувань для того, щоб не бути ошуканим черговим зразком глибокої аудіовізуальної підробки.*

Практичне значення. *Запропоновані правила корисні як для широкого використання серед суспільства для розвитку критичного мислення, так і для розробки комплексу компетентностей і програмних результатів, закладених у медіаосвітніх дисциплінах.*

Ключові слова: *дипфейк, дезінформація, маніпуляція, медіазалежність, медіаграмотність, штучний інтелект.*