Veritas: Legal and Psychological-Pedagogical Research

2025. Vol. 1, No. 2 **ISSN 3083-6859 (online); ISSN 3083-6840 (print)**

UDC 004.8:340.134

Deepfakes: Definition of the Concept and Criteria for Distinguishing Between Harmful and Harmless Deepfakes

Andrii Hachkevych

Ph.D. in Law, Associate Professor, Lviv Polytechnic National University, 12, S. Bandera str., 79013, Lviv, Ukraine, andrii.o.hachkevych@lpnu.ua, ORCID: 0000-0002-8494-1937

http://doi.org/10.23939/veritas2025.02.012

Abstract. This article addresses the issue of combating deepfakes, which has recently gained significant relevance. With the emergence of publicly available artificial intelligence tools capable of generating highly convincing images, video clips and other types of content, as well as a favorable digital landscape for their dissemination, deepfake technology has become increasingly prevalent. Given the risks of deepfakes, reasonable expectations are placed on the law designed to protect our fundamental values, which are often jeopardized by deepfakes. The spread of harmful deepfakes poses risks to the individuals depicted, causes damage and destroys the reputation of affected organizations, and can be dangerous for society by serving as an effective means of disinformation and manipulation of public opinion. Therefore, the author examines deepfakes as a challenge, exploring the concept in depth and highlighting some of the contentious issues. This article outlines four criteria to differentiate harmful deepfakes from harmless ones: a) consent of the individual–whether a person featured in a deepfake has agreed to its creation and dissemination, b) absence of criminal acts–whether a deepfake involves any illegal activities, c) indication of artificial intelligence usage–whether a deepfake clearly demonstrates characteristics of being created by artificial intelligence, d) understanding of their intents considering serving a social good. The author also attempts to provide his own definition of the concept of a deepfake, and to outline its components: technological and intellectual. Alongside traditional images and videos, he considers that audio recordings should be regarded as a type of deepfakes. Furthermore, this article discusses the relationship between the dissemination of harmful deepfakes and relevant legal categories such as revenge porn, defamation, right to privacy, unfair competition, and disinformation known in contemporary legal systems.

Keywords: deepfake, definition of deepfake, harmful deepfakes, artificial intelligence, deep machine learning, revenge porn, defamation, right to privacy, disinformation, deepfake law.

Introduction

The rise of deepfakes–video clips, images and other types of content that have been convincingly altered with advanced artificial intelligence systems–has become widespread, posing threats to the rights and interests of individuals (such as revenge porn and violations of the right to publicity), organizations (through unfair competition), and society in general.

Suggested Citation:
Hachkevych, A. (2025). Deepfakes: Definition of the Concept and Criteria for Distinguishing Between Harmful and Harmless Deepfakes. *Veritas: Legal and Psychological-Pedagogical Research.* 1(2), 12–20. DOI: doi.org/10.23939/veritas2025.02.012

Journal homepage: https://science.lpnu.ua/veritas
Article history: Received: 25.05.2025. Revised: 23.06.2025. Accepted: 28.11.2025.

Deepfakes can be used to misinform and manipulate public opinion, for example, by creating fake news reports or altering political speeches. The availability of advanced artificial intelligence systems capable of generating highly believable images and videos, combined with a favorable digital environment for information exchange, has led to the growth of deepfakes.

Given the risks of deepfakes, reasonable expectations are placed on the law designed to protect our fundamental values, which are often jeopardized by their spread. In this article, the author offers a thorough examination of the deepfakes phenomenon as a legal challenge, highlighting the importance of addressing the issue of combating deepfakes, which is especially relevant in today's context.

Literature Review

Since 2017, when the term "deepfake" first emerged, there has been a growing number of publications addressing the problem of combating deepfakes, as well as an expansion in the scope of related research topics. Although the term originates from the pseudonym of a user who replaced one person's face with another in a video clip [19] its etymology has later been associated either with the use of deep machine learning techniques – as further explained in this article – or with the idea that a deepfake is a fake of exceptional realism (with deep meaning, among other things, profound or extraordinary).

Deepfakes, often in conjunction with artificial intelligence, are examined in various contexts in modern research. In our opinion, there are three important interrelated areas of such research.

First, deepfakes as a phenomenon. This area focuses on the essence of the concept of a deepfake, the emergence of deepfakes and their evolution, deepfakes' classification, the distinction between deepfakes and other similar concepts, etc. [18; 26; 28; 34; 46]. We would like to mention Westerlund's article, which provides a comprehensive analysis of the phenomenon of deepfakes, referring to hyperrealistic videos that apply artificial intelligence to depict someone saying and doing things that have never happened [46, p. 39]. Vasist and Krishnan summarize the state of the art studying the deepfakes phenomenon and define further prospects for research, understanding deepfakes as hyperrealistic synthetic media [41].

Second, deepfakes can be viewed as a technique or a counter-technique. This area is explored in various related to IT works discussing the creation of deepfakes and their detection [23; 30; 42; 43]. The study of Nguyen et al. should be mentioned first of all due to its contribution to the systematizing of methods for deepfakes' creating and detecting, enumerating AI-based tools and their characteristics, and explaining how they work in general. This article describes two approaches to the concept of a deepfake denoting content synthesized with the help of artificial intelligence: broad and narrow, depending on whether face-replacement technology or also other technologies are used to create a deepfake [35].

Third, deepfakes as a threat. While having a significant impact on society [15], deepfakes pose risks of various kinds, from creating distrust in the news to emerging a new means of disinformation [10; 40]. Botha and Pieterce discuss deepfakes in the same context of 21st-century information security threats as fake news and call deepfake a technology that combines and superimposes existing images and videos onto source content using Generative Adversarial Networks (GANs) [7]. The study by Kirchengast highlights the dangerous side of image manipulation and its harmful implications, such as political deception, voter manipulation, commercial fraud, and revenge porn. In the same study, deepfakes are defined as a type of human image synthesis, where an existing image is superimposed into a video to change the identity of those depicted in the video [20]. Meskys et al. identify four categories of deepfakes: ranging from rather harmful deep fake porn and deepfakes in political campaigns to less harmful deepfakes for commercial uses and creative deepfakes. They also provide the definition of the term "deep fakes" (separate writing) as face-swapping technologies that enable a quick creation of incredibly realistic images and videos [27].

Purpose

The purpose and objectives of this study are to address the rapid development of technologies for creating deepfakes. As the creation and dissemination of deepfakes become increasingly common, this article focuses on defining what a deepfake is and how to differentiate harmful deepfakes from harmless ones.

This article aims to provide a comprehensive review of the concept of deepfakes. Such purpose is caused by the lack of a generally accepted definition that can be used in the law.

The specific objectives of the article are as follows:

1. To explain the term "deepfake" etymology and its possible usages.
2. To identify the key features of deepfakes as existing forms of content.
3. To establish criteria for distinguishing between harmful and harmless deepfakes.

Methodology

This study is based on the concept of deepfake, which is often confused with the term "fake". First, we explore the origins of the term "deepfake" and its relationship to the more well-known concept of fake. After analyzing existing definitions and uses of the concept of deepfakes, we employ the method of generalization to clarify it further.

Using a systematic approach, we examine the key characteristics of deepfakes, including their various forms, types of depiction, and means used to achieve their notable feature–being hyper-realistic. A crucial part of our study involves developing criteria to differentiate between harmful and harmless deepfakes.

Considering the possible negative consequences of deepfakes related to such phenomena as revenge porn, defamation, right to publicity violations, and disinformation, we have established criteria that could serve as a foundation for ethical guidelines and legal regulations.

Results and Discussion

The term "deepfake", derived from English, combines elements related to deep learning–a form of artificial intelligence that mimics the processes of the human brain using artificial neural networks. The second part of the term, "fake", has been widely adopted in modern languages to convey various meanings, from false news ("fake news") to fake profiles of public figures on social networks. One of the definitions of fake news is that it is a form of forgery, a falsity disseminated specifically to misinform the audience [32].

Is it reasonable to derive the meaning of "deepfake" from the neologism "fake"? On one hand, both terms are associated with misleading through misinformation. On the other hand, the word "fake" emphasizes lies in content, generally presented as news, while "deepfakes" indicate false representations in other formats, including video clips and images. Understanding the etymology of "deepfake" clarifies that not all falsified materials in media, especially on social media, are classified as deepfakes. We believe that the term "deepfake" should be considered beyond just a category of fakes to gain a more comprehensive understanding of combating deepfakes.

When discussing the essence of deepfakes, the question arises which element should be emphasized: the ability of a deepfake to appear believable, fostering mimicking authenticity, or the fact that deepfakes are created using specialized software based on deep machine learning, a branch of artificial intelligence.

The term "hyperrealistic" is often used to describe the believability as a feature of deepfakes, which derived from the art world to indicate something that looks so real that it could be mistaken for reality. Consequently, one often perceives deepfakes as genuine due to their high level of believability and convincing authenticity. Unfortunately, identifying a deepfake and recognizing its false nature, especially when it is generated using advanced software, typically requires sophisticated detection tools. The focus on deep machine learning, which is based on imitating the functioning of human brain neural networks, reflects the technological component of a deepfake. At the same time, the effect of high believability as another component can also be achieved through other means, such as editing images with graphic software or creating videos involving a celebrity's double.

From a legislative perspective, we believe that videos or images generated by artificial intelligence using deep learning should be treated the same way as those produced through other methods, as the impact of their disseminating can be equally damaging. Theoretical discussions, meanwhile, often highlight that a deepfake has been created with the help of artificial intelligence, with machine learning serving as one of its foundational elements [3; 8; 9].

The effect of high believability pertains to the presentation of information rather than its accuracy. A video may appear realistic until we analyze the likelihood of the depicted event actually occurring. In some cases, even after analysis, it may be still

challenging to determine the truth. Therefore, the definition of a deepfake should emphasize its hyper-realistic form and the likeness to the individual it includes, rather than its factual accuracy. Moreover, not all deepfakes pose a danger. For instance, a humorous video that is readily perceived as fake does not carry the same risks as a video that appears realistic.

Additionally, the involvement of artificial intelligence in the creation of deepfakes, namely the method of deep learning, presents a double-edged sword in formulating a definition. On one hand, artificial intelligence ensures that videos or images, or perhaps other types of content, become convincing due to the ability of generating a brief, unique analog of reality. On the other, the outcome of this process is not reality but rather a fictional representation embodied in a content, according to the parameters set by the user of the software. We can see that artificial intelligence not only imitates human abilities, but may also exceed them while producing a video clip or image. Artificial intelligence can copy a person–not in real life, but within the virtual space–showcasing their actions, speech, singing, or other behavior. This observation expands our understanding of artificial intelligence by including its significant peculiarity that is often overlooked in traditional definitions.

Despite the apparent connection with artificial intelligence (technological component) and high believability of the content (intellectual component) disseminated for some purpose, the concept of a deepfake lacks a generally accepted understanding.The phenomenon of deepfakes clearly refers not only to video clips in which one person's face is replaced with that of another, often a well-known individual. This narrow interpretation stems from the origin of the term “deepfake”, which was initially used for videos with altered faces shared on the social network Reddit).

Any synthetic content closely resembling reality can be categorized as a deepfake. We can also state that a deepfake may not feature a specific actor but instead depict events where the identity of individuals is irrelevant (for example, a group of protesters whose faces are hidden while chanting slogans or posted on some Ukrainian telegram channels video showing a statue of Taras Shevchenko in Lviv moving as it were alive). Although these deepfakes may not pose a challenge since they do not infringe upon anyone’s right to publicity or constitute defamation, they can harm societal interests through hidden manipulation and misinformation; deepfakes, where the identity of a depicted person is not important, can also be utilized for fraudulent purposes.

At this point, the concept of a deepfake can encompass a wide array of content, provided it contains technological and intellectual components. The understanding of deepfakes might be broadened by including video clips, images, audio recordings, and texts. The common practice of using modern software to replace the voice of the original performer in a well-known song with another voice highlights the necessity of viewing combating deepfakes as a wider issue than just fake videos.

Furthermore, we should note that the term “deepfake” is not fixed; it is used in various contexts. In addition to referring fake images, video clips, or other forms of content representation, “deepfake” can also denote the technology responsible for creating such content. This dynamic nature of the term is illustrated in many contemporary articles, where authors explore the technology behind deepfakes [1; 5; 33; 38].

We believe that the following four criteria should be used to distinguish harmful fakes from harmless ones.

Criterion 1: Consent. Deepfakes represent situations that individuals did not actually engage in, making them fundamentally different from photographs or video clips that capture real-life events. However, deepfakes utilize genuine characteristics of a person, such as their appearance, facial expressions, and body structure. In several countries, these characteristics are protected under the right to publicity, which safeguards various manifestations of a person–from name to signature–against unauthorized commercial use.

The right to publicity shares similarities with trademark protection, linking it to the realm of intellectual property [11]. While the protection of human identity through the right of publicity is well-established in U. S. law, particularly highlighted by three landmark cases involving Bette Midler, Vanna White, and Tom Waits in the latter half of the 20th century [24], various legal systems also recognize certain personality rights [31].

Consequently, any commercial use of a person's voice, image, and other attributes should require their consent to avoid potential legal issues. With respect to deepfakes, the right to publicity and an individual's ability to control the meaning of own identity [25] applies in cases where deepfakes are used for advertising or other commercial purposes. Furthermore, we believe that obtaining consent from the individual depicted is essential to differentiate between harmful and non-harmful, non-commercial uses of the technology.

Criterion 2: Prohibitions of criminal law. While prohibiting certain acts under the threat of penalties, criminal law encompasses various crimes involving the use of deepfakes, such as revenge porn, fraud, defamation, and manipulation across different countries.

Revenge porn is related to the publication of explicit photographs and videos without the individual's consent. According to the Merriam-Webster dictionary, it occurs when sexually explicit images of a person are posted on the Internet without his or her consent, especially as a form of revenge or harassment [29]. Despite their falsity, deepfakes can also be considered revenge porn, given the factor of their perception as accurate. In addition, revenge porn should be separated from child pornography, depending on the age of the individuals depicted in the photos or videos, focusing primarily on adults [13].

The article's title, "The Development of Blackmail through Deepfakes" [6], clearly reflects the interconnection of deepfakes and another crime, closely related to revenge porn, when the individual is coerced into performing specific actions, first and foremost paying a large sum of money under the threat of negative consequences. The creation of deepfakes and their dissemination may lead to defamation. However, not all countries have given rise to criminal liability in addition to civil liability. Defamation, meaning disseminating false information that harms individuals' or organizations' interests, has ignited debates over whether it should fall under criminal law [4].

In our brief overview of deepfake-related prohibitions, we should also mention the manipulations that become more widespread in recent political campaigns [22]. Some countries, including Singapore [44], have already enacted laws criminalizing the use of deepfakes.

Deepfakes may serve as a tool for fraud and other illegal activities, including identity theft, illegal migration, and espionage [2]. Sandal et al. have recently explored the phenomenon of deepfakes within the criminal justice system and covered actual prohibitions in more detail [37].

The challenges posed by deepfakes are expected to be considered for improving criminal legislation and developing case law. A comparative approach, where states borrow norms from one another to introduce new crimes into their legal systems, is preferable. Thus, a deepfake should not be a means of committing a crime based on the corpora delicti of the criminal law.

Criterion 3: Content labeling. Scholars are increasingly studying the idea of marking the origin of content created by artificial intelligence about creating and disseminating deepfakes [14; 21; 45]. This suggestion is also being supported at the level of government agreements. One of the key principles of the G7 is to "develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, including watermarking or other techniques to enable users to identify AI-generated content" [12]. The rules on labeling are becoming a part of the legislation, first of all, under the progressive EU AI Act (Article 50).

The issue of content labeling raises several organizational and legal dilemmas.

1. Given the ongoing debate around generative artificial intelligence, what content should be marked?

2. What kinds of signs are the most suitable for different types of content? Should these markings be embedded in the file's technical information or demonstrated in the content of a deepfake?

3. What are the responsibilities associated with refusing from labeling? This brings up another question about whether the responsibility for labeling lies with the developer of the AI system that generates the content, with the supplier, or potentially with the user.

Criterion 4: Public Good. The final criterion, which we consider the most controversial, cannot be ignored, as more and more deepfakes are created and disseminated to misinform and manipulate public opinion. This criterion relates to subjective assessments regarding how much a deepfake contributes to the public good.

On one hand, there may be no direct legal prohibitions against deepfakes in the country where they are created or disseminated. The principle of freedom of information might be undoubtedly recognized as the grounds for justifying harmful intentions (but unpunished). On the other hand, even in the absence of criminal law provisions, the release of deepfakes can violate existing ethical standards, erode public trust in information technology, and negatively impact fundamental democratic values. This type of thing is typical to political campaigns and media coverage of significant global events.

Considering these factors, we offer some examples of deepfakes that benefit society: (1) the area of healthcare – using face replacement technology to protect patient privacy [48] and making patient prototypes to develop the professional level of doctors [16]; (2) the area of education–re-creating historical events and utilizing multimedia technologies more effectively [36] (3) entertainment–reviving an image of a deceased prominent person, like Salvador Dali, as a guide in a museum [17], or diversifying game scenarios [39].

Conclusions

According to the study results, we define deepfakes as highly believable video clips, images, audio recordings, and other content created using artificial intelligence for specific purposes. The nature of these purposes largely determines whether deepfakes can harm the legitimate rights and interests of individuals, organizations, and society. Two key features should be considered when defining deepfakes: the intellectual and technological components. We argue that it is unreasonable to limit the definition of deepfakes to harmful video clips, often of a pornographic character, created through deep machine learning. Instead, one should recognize different forms of content that can be used to create deepfakes; and deepfakes can also depict non-recognizable human beings. This topic warrants further exploration, particularly concerning whether special laws should be enacted to combat deepfakes or whether the existing legal framework is sufficient.

Even in the absence of special laws that could contain the features showing what constitutes a deepfake as opposed to other forms of content, the criteria described in this article will be helpful in a deeper study of the phenomenon of deepfakes. Based on legal grounds, these criteria provide a practical tool for assessing deepfakes from the standpoint of legality, aiding in their classification into harmful and harmless.

Acknowledgements. None.

Funding. The author declares no financial support for the research, authorship, or publication of this article.

Author contributions. The author confirms sole responsibility for this work. The author approves of this work and takes responsibility for its integrity.

Conflict of interest. The author declares no conflict of interest.

Institutional review board statement. Not applicable.

REFERENCES

1. Alanazi, S., Asif, S. (2024). Exploring deepfake technology: creation, consequences and countermeasures. H*uman-Intelligent Systems Integration, 6,* 49–60. DOI: https://doi.org/10.1007/s42454-024-00054-8

2. Alanazi, S., Asif, S., Moulitsas, I. (2024). Examining the societal impact and legislative requirements of deepfake technology: a comprehensive study. *International Journal of Social Science and Humanity, 14(*2), 58–64.

3. Almars, A. (2021). Deepfakes Detection Techniques Using Deep Learning: A Survey. *Journal of Computer and Communications, 9,* 20–35. DOI: https://doi.org/10.4236/jcc.2021.95003

4. Anstey, B. J. (2017). Criminal defamation and reputation as 'honour': a cross-jurisdictional perspective. *Journal of Media Law, 9*(1), 132–153. DOI: https://doi.org/10.1080/17577632.2017.1311467

5. Arslan, F. (2023). Deepfake Technology: A Criminological Literature Review. *Sakarya Üniversitesi Hukuk Fakültesi Dergisi, 11*(1), 701–720. DOI: https://doi.org/10.56701/shd.1293642

6. Blancaflor, E., Garcia, J. I., Magno, F. D., Vilar, M. J. (2024). Deepfake Blackmailing on the Rise: The Burgeoning Posterity of Revenge Pornography in the Philippines. In *Proceedings of the 2024 9th International Conference on Intelligent Information Technology* (ICIIT '24). Association for Computing Machinery, New York, NY, USA, 295–301. DOI: https://doi.org/10.1145/3654522.3654548

7. Botha, J. G., Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. In: Payne, B. and Wu, H. (Eds.), *Proceedings of the 15th International Conference on Cyber Warfare and Security* (pp. 57–66).

8. Campbell, C., Plangger, K., Sands, S., Kietzmann, J., Bates, K. (2022). How Deepfakes and Artificial Intelligence Could Reshape the Advertising Industry: The Coming Reality of AI Fakes and Their Potential Impact on Consumer Behavior. *Journal of Advertising Research, 62*(3), 241–251. DOI: https://doi.org/10.2501/JAR-2022-017

9. Chadha, A., Kumar, V., Kashyap, S., Gupta, M. (2021). Deepfake: An Overview. In: Singh, P. K., Wierzchoń, S. T., Tanwar, S., Ganzha, M., Rodrigues, J. J. P. C. (Eds.) *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security.* Springer. DOI: https://doi.org/10.1007/978-981-16-0733-2_39

10. Chesney, R., Citron, D. (2019). Deepfakes and the New Disinformation War. *Foreign Affairs, 98*(1), 147–155.

11. Dogan, S., Lemley, M. (2006). What the Right of Publicity Can Learn from Trademark Law. *Stanford Law Review, 58,* 1161–1220.

12. European Commission. (2023, October 30). *Hiroshima Process International Guiding Principles for Advanced AI system.* Retrieved from: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system

13. Franks, M. A. (2017). 'Revenge Porn' Reform: A View from the Front Lines. *Florida Law Review, 69,* 1251–1287.

14. Gamage, D., Sewwandi, D., Zhang, M., Bandara, A. (2025). *Labeling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media.* Retrived from: https://arxiv.org/abs/2503.05711

15. Hancock, J., Bailenson, N. (2020). The Social Impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 149–152. DOI: https://doi.org/10.1089/cyber.2021.29208.jth

16. Kaur, A., Hoshyar, A. N., Wang, X., Xia, F. (2024). Beyond Deception: Exploiting Deepfake Technology for Ethical Innovation in Healthcare. In *Proceedings of the 1st International Workshop on Multimedia Computing for Health and Medicine* (pp. 70–78). Association for Computing Machinery. DOI: https://doi.org/10.1145/3688868.3689196

17. Kidd, J. Rees, A. (2022). Chapter 10: A Museum of Deepfakes? Potentials and Pitfalls for Deep Learning Technologies. In: T. Stylianou-Lambert, A. Bounia, A. Heraclidou (Ed.), *Emerging Technologies and Museums: Mediating Difficult Heritage* (pp. 218–232). Berghahn Books. DOI: https://doi.org/10.1515/9781800733756-012

18. Kietzmann, J., Lee, L. W., McCarthy, I. P., Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146. DOI: https://doi.org/10.1016/j.bushor.2019.11.006

19. Kikerpill, K., Siibak, A., Valli, S. (2021). Dealing with Deepfakes: Reddit, Online Content Moderation, and Situational Crime Prevention. In: J. B. Wiest (Ed.). *Theorizing Criminality and Policing in the Digital Media Age* (*Studies in Media and Communications, Vol. 20*) (pp. 25–45). Emerald Publishing Limited. DOI: https://doi.org/10.1108/S2050-206020210000020008

20. Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law, 29*(3), 308–323. DOI: https://doi.org/10.1080/13600834.2020.1794615

21. MacKenzie, W. I., Jr, Weber, R., Barr, H. M., Lanius, C., Tenhundfeld, N. L. (2025). Deepfake Label Recall: Combating Disinformation with Labels is Especially Effective for Those Who Dislike the Speaker. *International Journal of Human–Computer Interaction,* 1–15. DOI: https://doi.org/10.1080/10447318.2025.2466068

22. Maddocks, S. (*2020). 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes. Porn Studies, 7*(4), 415–423. DOI: https://doi.org/10.1080/23268743.2020.1757499

23. Masood, M., Nawaz, M., Malik, K.M., Javed, A., Irtaza, A., Malik, H. (2023). Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence, 53,* 3974–4026. DOI: https://doi.org/10.1007/s10489-022-03766-z

24. McCarthy, T. (1995). The Human Persona as Commercial Property: The Right of Publicity. *Columbia University School of Law and the Arts*, 19(3–4), 129–148.

25. McKenna, M. (2005). The Right of Publicity and Autonomous Self-Definition. *University of Pittsburgh Law Review, 67*, 225–294.

26. Meikle, G. (2022). *Deepfakes.* John Wiley & Sons.

27. Meskys, E., Liaudanskas, A., Kalpokiene, J., Jurcys, P. (2019). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice, 15*(1), 24–31. DOI: https://doi.org/10.1093/jiplp/jpz167

28. Miller, D., Somoray, K., Stevens, H. (2025). A Shallow History of Deepfakes (January 22, 2025). Available at SSRN. Retrieved from: https://papers.ssrn.com/sol3/Delivery.cfm/5130379.pdf?abstractid=5130379&mirid=1

29. Merriam-Webster. (n. d.). Revenge Porn. In *Merriam-Webster.com dictionary*. March 10, 2025. Retrieved from: https://www.merriam-webster.com/dictionary/revenge%20porn

30. Mirsky, Y., Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys, 54*(1), 7. DOI: https://doi.org/10.1145/3425780

31. Moskalenko, K. (2015). The right of publicity in the USA, the EU, and Ukraine. *International Comparative Jurisprudence, 1*(2), 113–120. DOI: https://doi.org/10.1016/j.icj.2015.12.001

32. Mudra, I. (2016). The Concept of "Fake" and its Views in the Media. *TV and Radio Journalism, 15*, 184–188.

33. Mullen, M. (2022). A New Reality: Deepfake Technology and the World Around Us. *Mitchell Hamline Law Review, 48*(1), 210–234.

34. Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research, 154*, Article 113368. DOI: https://doi.org/10.1016/j.jbusres.2022.113368

35. Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q., Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding, 223*, 103525. DOI: https://doi.org/10.1016/j.cviu.2022.103525

36. Pandey, C. K., Mishra, V. K., Tiwari, N. K. Deepfakes: When to Use It. In *2021 10th International Conference on System Modeling & Advancement in Research Trends* (pp. 80–84). IEEE. DOI: https://doi.org/10.1109/SMART52563.2021.9676297

37. Sandoval, M. P., de Almeida Vau, M., Solaas, J., Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: a systematic review. *Crime Science, 13*, 41. DOI: https://doi.org/10.1186/s40163-024-00239-1

38. Sharma, M., Kaur, M. (2022). A review of deepfake technology: an emerging AI threat. In Soft Computing for Security Applications: Proceedings of ICSCS 2021 (pp. 605–619). Springer. DOI: https://doi.org/10.1007/978-981-16-5301-8_44

39. Tariq, S., Abuadbba, A., Moore, K. (2023). Deepfake in the Metaverse: Security Implications for Virtual Gaming, Meetings, and Offices. In *Proceedings of the 2nd Workshop on Security Implications of Deepfakes and Cheapfakes* (pp. 16–19). Association for Computing Machinery. DOI: https://doi.org/10.1145/3595353.3595880

40. Vaccari, C., Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society, 6*(1), 1–13. DOI: https://doi.org/10.1177/2056305120903408

41. Vasist, P., Krishnan, S. (2022). Deepfakes: An Integrative Review of the Literature and an Agenda for Future Research. *Communications of the Association for Information Systems, 51*, 556–562. DOI: https://doi.org/10.17705/1CAIS.05126

42. Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. *IEEE Journal of Selected Topics in Signal Processing, 14*(5), 910–932. DOI: https://doi.org/10.1109/JSTSP.2020.3002101

43. Wang, S.-Y., Wang, O., Zhang, R., Owens, A., Efros, A. A. (2020). CNN-Generated Images Are Surprisingly Easy to Spot… for Now. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8692–8701). Seattle. DOI: https://doi/10.1109/CVPR42600.2020.00872

44. Werner, J. (2024, October 24). *Singapore's parliament passes bill to combat manipulated online election content.* Retrieved from: https://babl.ai/singapores-parliament-passes-bill-to-combat-manipulated-online-election-content/

45. Wittenberg, C., Epstein, Z., Berinsky, A. J., Rand, D. G. (2024). Labeling AI-Generated Content: Promises, Perils, and Future Directions. *An MIT Exploration of Generative AI.* DOI: https://doi.org/10.21428/e4baedd9.0319e3a6

46. Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review, 9*(11), 40–53.

47. Whittaker, L., Letheren, K., Mulcahy, R. (2021). The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing. *Australasian Marketing Journal, 29(*3), 204–214. DOI: https://doi.org/10.1177/1839334921999479

48. Zhu, B., Fang, H., Sui, Y., Li, L. (2020). Deepfakes for Medical Video De-Identification: Privacy Protection and Diagnostic Information Preservation. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 414–420). Association for Computing Machinery. DOI: https://doi.org/10.1145/3375627.3375849

Andrii Hachkevych

Діпфейки: визначення поняття і критерії розмежування шкідливих та нешкідливих діпфейків

Андрій Гачкевич

Кандидат юридичних наук, доцент, Національний університет “Львівська політехніка”,
вул. С. Бандери, 12, 79013, Львів, Україна, andrii.o.hachkevych@lpnu.ua, ORCID: 0000-0002-8494-1937

Анотація. Стаття присвячена проблемі боротьби з діпфейками, яка набула особливої актуальності на сучасному етапі. Внаслідок появи загальнодоступних можливостей на базі штучного інтелекту створювати дуже правдоподібні зображення, відеокліпи та інші види контенту, а також формування сприятливого цифрового середовища для їхнього поширення, технологія діпфейків стала широко застосовуваною. Через небезпеку діпфейків виникла нагальна потреба у вдосконаленні права, що забезпечує охорону основоположним цінностям, для яких діпфейки часто становлять загрозу. Поширення шкідливого діпфейку загрожує правам людини, яка є дійовою особою діпфейку, спричиняє збитки для організації, на інтереси якої він впливає, руйнуючи позитивну репутацію, а також може бути небезпечним для суспільства загалом – як ефективний інструмент дезінформування та маніпулювання громадською думкою. Відповідно у цій статті автор розглядає явище діпфейків як виклик та розкриває суть поняття діпфейку, зауважуючи окремі спірні моменти. Названо чотири критерії, завдяки яким можна відрізняти шкідливі діпфейки від нешкідливих: згода особи (чи показана особа дала свою згоду), відсутність злочину (чи не має стосунку діпфейк до складу можливого злочину), позначення використання штучного інтелекту (чи сам контент містить вказівку на те, що він створений з використанням штучного інтелекту), а також сприяння суспільному благу (чи поширення діпфейку корисне для суспільства). Автор зробив спробу навести власне визначення поняття діпфейку, а також виокремити його складові: технологічну та інтелектуальну. Крім широко розповсюджених зображень та відеокліпів як форм діпфейків, зазначено доцільність долучати до діпфейків й аудіозаписи. У статті порушено питання взаємозв’язку між поширенням шкідливих діпфейків та відображенням у праві сучасних держав таких категорій, як порнопомста, дифамація, право на приватність, недобросовісна конкуренція, дезінформація та ін.

Ключові слова: діпфейк, визначення діпфейку, шкідливі діпфейки, штучний інтелект, глибоке машинне навчання, порнопомста, дифамація, право на приватність, дезінформація, закон про діпфейки.