

Analysis and Prevention of the Drawbacks of Artificial Intelligence Deep Faking Technology

Yuyan Liu *

LawSchool, Xinjiang University of Political Science and Law, Tumushuke, China

* Corresponding Author Email: qiaoluo717818@outlook.com

Abstract. The rapid development of artificial intelligence has led to the emergence of deepfake technology in a way that “foils reality”, making it visible to the public. Research has shown that the abuse of deepfake technology is undermining the social trust structure and the authenticity of information dissemination, and is highly likely to lead to risks such as fraud and infringement of reputation rights. Moreover, it also has problems such as ambiguous liability determination and difficulty in tracing the source. In response to this, it is necessary to establish a complete legal framework, such as defining the liability entities for the abuse of this technology and the issue of sentencing, enhancing the proactive defense capabilities of the technology, establishing comprehensive and real-time monitoring technologies through multi-party cooperation, and promoting social collaborative governance to enhance the awareness of prevention among the general public. While curbing the development of technology, people should also reserve space for its beneficial applications, promote the development of deepfake technology within the bounds of legality and compliance, and maintain social stability and protect public interests.

Keywords: Deepfake, legal regulation, risk management.

1. Introduction

The rapid development of artificial intelligence not only promotes the advancement of deepfake technology, making it gradually come into the public view, but also provides lawbreakers with an opportunity to commit crimes by using deepfake technology. As of now, deepfake technology has advanced to the point where it can “fool the eye”, appearing almost indistinguishable from the real thing. It is commonly seen in films and virtual environments, bringing entertainment and novelty to the general public while quietly posing risks. The rapid development of generative artificial intelligence has made image and video tampering much easier and more covert. There are two major problems in existing technology: the “black box” characteristic and the limited “generalization ability” [1]. At present, China still faces many challenges in establishing a legislative framework for deepfake technology. The existing laws and regulations have limitations and lagging aspects. For instance, in the 2024 AI face-swapping fraud case in Inner Mongolia, the victim suffered a loss of 400,000 yuan and could only be held accountable for the crime of ordinary fraud. Due to the fact that the concept of “sensitive personal information” has not been clearly defined in criminal law, the absence of protection for sensitive information will lead to errors in sentencing and fail to reflect the special protection of biological information. As technology continues to evolve and update, the risk coefficient of engaging in criminal activities will also keep increasing. The inherent threat of deepfake technology lies in its ability to somewhat undermine the public’s “perception” of videos, weakening their authenticity and reliability as a means of information dissemination and evidence. Its impact extends beyond the technical issue of “difficult to distinguish truth from falsehood” and also challenges the social trust structure [2]. In formulating response measures, there should be joint participation, mutual agreement and learning among all sectors of society. Legal, technical means and social governance should work together, and none of them can be lacking. Based on this, the research content of this article is the analysis and prevention of the drawbacks of deepfake technology.

2. The Current Situation and Challenges of AI Deep Forging Governance

2.1. Current Governance Situation

Since deepfake technology was applied to cybercrime activities, whether it is the forgery of voices or the creation of videos, photos, etc. for criminal purposes, it has disrupted the normal social order and aroused the anger and panic of the majority of the public. At present, many countries have successively introduced laws and regulations targeting the abuse of deepfake technology.

2.2. The Main Challenges Encountered during the Governance Process

At the technical level, the existing detection technologies have entered a stage of “multi-modal + active defense + large” model confrontation in the technological competition. It has achieved a certain improvement in the ability to trace the source, but it still lags behind the development of forgery technology and is unable to accurately identify new types of false content. The abuse of deepfake technology leads to the situation where criminals can easily produce fake audio, video, images, etc., but it also makes it difficult to trace the source, locate the responsible party, and hold them accountable. Its management also has shortcomings. The AI algorithm is extremely large in size and is used by many people, resulting in a mixed group of users who are difficult to manage and monitor. It is not feasible to conduct monitoring at all times, and the subsequent handling cannot reduce the damage caused by the rapid spread. From a legal perspective, several countries have successively formulated boundaries for the responsibilities related to AI deepfake technology. For instance, China has issued the "Interim Measures for the Management of Generative Artificial Intelligence Services", which explicitly requires that AI-generated content must be labeled. The EU's "Artificial Intelligence Act" mandates the identification of deepfake content, among other things. However, the current laws and regulations are still not fully adequate. There is a lack of protection for citizens' personal information, it is difficult to determine the responsible parties and malicious intentions, and the infringements caused by deepfake technology involve multiple legal interests (such as personal privacy and social order), and the traditional criminal law has a lag in defining them [3]. The current criminal law provisions do not explicitly include deepfake technology within the scope of regulation. There are also no specific targeted crimes for the abuse of this technology, which fails to keep up with the continuous advancement of the technology and makes it difficult to carry out precise crackdowns in judicial practice. Moreover, during the evidence review stage, due to the high fidelity of deepfake technology, there are indeed some cases where it is used for this purpose. The procuratorial authorities need to invest more effort in distinguishing the authenticity, for example, some people use this technology to forge audio-visual materials, data and other evidence, which increases the workload of relevant staff and makes the case determination more difficult [4]. Deepfake technology involves multiple parties such as developers, users, and platforms. The traditional criminal law framework, which holds natural persons or entities as the responsibility subjects, is insufficient to cover this situation, and the determination of responsibility and division is rather difficult. To commit a crime, it is necessary to determine whether it was committed with subjective intent or with malicious intent. However, in the case of deepfake crimes, it is more difficult to determine as it involves algorithmic issues. The judicial authorities lack corresponding measures to verify whether it was committed with subjective intent. At the social level, people have now shifted from the initial "passive response" to an active response. People are conducting active defense through a tripartite approach of "legal regulation, technological confrontation, and platform regulation", and flexibly applying existing laws and regulations (such as the "Cybersecurity Law") to impose substantive penalties on such behaviors. However, the abuse of deepfake technology is a serious issue. The cost of creating fake content is extremely low and the process is extremely fast. The construction of the social defense system often lags behind the dissemination of fake videos, audio, images, etc., resulting in a certain "time lag". The dissemination of audio-visual content, images, etc. synthesized by AI will confuse the public's perception. The spread of false information, due to the increasingly mature forgery technology, becomes difficult to distinguish between true and false, which is likely to cause

social chaos and reduce the public's trust in news media [5]. Deepfake technology has been widely misused to create false and indecent content that spreads everywhere, which to some extent can cause irreversible psychological or mental damage to the victims. The dissemination of false information can also incite group confrontation. For instance, in 2023, a certain online account used deepfake technology to create a false video showing "conflict between different groups", which led to offline gatherings and eventually the account operator was held legally accountable; or by fabricating false conflict incidents, they "created" real conflicts, causing public anger and panic.

3. The Risks and Challenges Brought about by AI Deepfake Technology

The technical characteristics and diverse risks of AI deepfake technology have prompted the establishment of governance frameworks both domestically and internationally. However, the current governance still faces multiple challenges.

3.1. Impact on Information Dissemination and Social Trust

In the current era of Internet information, the rapid development and advancement of AI deepfake technology have led to the content produced by it gradually becoming "so realistic as to be indistinguishable from the real thing" and thus entering the public's view. It has certain advantages in fields such as film and television, virtual reality, etc. However, the false content generated through deepfake technology spreads rapidly through the internet, leading to a further proliferation of fake news and rumors. For instance, some individuals have used AI deepfake technology to create false news such as "A tunnel collapse occurred in Xicheng District, Shanxi Province, resulting in the death of five people" and "A 6.0 magnitude earthquake hit Xinzhou, Shanxi Province", and the videos they released were accompanied by a strong shaking effect and extremely realistic scenes, which interfered with the public's judgment of true information. Zhao used an AI-generated video of "traffic accident scene with casualties" in which there were even bloodied and disfigured "victims" in the picture. This video was extremely impactful and caused great unease among the local people. The above case did not merely involve visual "parody", but rather fabricated disaster scenes and sources of information, precisely targeting the points that could cause public panic, thereby undermining public trust and compassion. The emergence of such illegal acts has further exacerbated the crisis of social trust.

3.2. Threats to Individual Rights and Interests

Deepfake technology can be used to fabricate indecent videos of individuals, false statements, etc., infringing upon their portrait rights, reputation rights, and even being used as a tool for fraud. For instance, there is the case of impersonating a live-streaming for product promotion. Recently, actor Wen Zhengrong was involved in an AI face-swapping live-streaming promotion. His team reported an average of 50 fake accounts per day, but the situation still persisted despite repeated bans. Yu Menglong learned to use the "AI face-swapping" software and obtained others' facial information from various sources. Without the consent of the individuals, he used this software to swap faces and forge obscene videos and pictures. As of now, this technology is no longer just about changing faces in videos. It has already been able to penetrate into areas such as property security. Fraudsters no longer merely use fake voices for fraud, but instead forge real-time video calls with relatives and friends, and simulate situations such as "encountering an emergency and needing funds for emergency expenses". The abuse of AI deepfake technology to a certain extent makes it difficult for the public to distinguish between real and fake. During the application of this technology, it is prone to infringe upon citizens' portrait rights, reputation rights, etc., thereby violating citizens' legitimate rights and interests and endangering social public order, etc [6]. The reduction in the cost of spreading false information among others means that the perpetrators of such acts receive little punishment and thus cannot be effectively deterred from engaging in such behavior. Maliciously spreading false rumors to humiliate others is one of the most egregious forms of infringement. Criminals will use AI to replace someone's face and spread it on the internet or on malicious websites. Even if the victim

refutes the rumor, it is difficult to completely erase the fake video circulating on the internet. The victim will suffer extremely and irreversibly continuous harm both mentally and psychologically. When everything can be fabricated, genuine evidence will also seem utterly unconvincing. If one presents evidence against injustice and is accused by the opposing side of "it was fabricated by AI", and if the detection technology is not up to par, it will be impossible to prove that what the opposing side claims as fabrication is false, nor can it be proven that the presented evidence is not fabricated. This can easily lead to a deadlock. If one has not done something but the other party presents evidence to prove that they did, and it is impossible to determine whether it was generated by an AI or not, there is a very high probability that the person will be unable to defend themselves convincingly. The occurrence of the above situation and the resulting damage to rights will bring about a deeper crisis of trust for both society and individuals.

4. Countermeasures Against AI Deepfake Technology

In response to the aforementioned problems of technological lag, incomplete laws, and weak social defense, a systematic solution needs to be constructed from the perspectives of technology, law, and society.

4.1. Technical Solution

Regarding deepfake technology, relevant departments and technicians should increase the research and development of detection and identification methods for whether AI algorithm technology and precise traceability technology have been used, further expand the scope of active defense and expand the response measures. They can adopt technologies such as digital watermarking and blockchain certification to trace the origin of the generated content. Or establish an intelligent recognition system, through which the facial dynamic features and speaking rhythm of the people in the video can be analyzed to determine whether they have engaged in forgery; in terms of platform review, a real-time detection system can be established to conduct detailed manual review of the suspected content identified. At the same time, the developers of deepfake technology should also actively participate in this process, promoting the diversification of the development team and enhancing their sensitivity to whether the related technology is being used for criminal activities [7].

4.2. Legal Solution Approach

The relevant departments should further improve the laws and regulations, and add specific charges for the series of criminal issues arising from the abuse of deepfake technology. They should also clearly define the constitutive elements, classification of criminal subjects, and sentencing standards for the abuse of deepfake technology. It is also possible to refine the laws and regulations, clearly defining the legal responsibilities and boundaries of the counterfeiters, dissemination platforms, etc., or through judicial interpretations, clarifying the criteria for criminal charges to reduce the lag of the law. Legislation should be enacted to prohibit the creation or dissemination of deepfake content without the individual's consent, especially when such content is intended with malicious intent [8]. Legal regulations need to be flexible. Legislation can also precisely focus on the national domain. Through a three-dimensional framework of "defining prohibited behaviors + clarifying responsible entities + establishing a rapid relief mechanism", it can not only curb malicious use but also provide space for the benign application of technology [9]. For cases where the platform failed to fulfill its review obligations and allowed false information to spread on the platform, the platform should be severely punished to force it to conduct more rigorous reviews or use verification tools to assist in the review process. Add the "Abuse of Deepfake Technology Crime", and clearly define the elements of this crime (such as forging others' biometric information, spreading false content causing serious consequences). Jiang Ying also pointed out in his research the shortcomings of the current criminal law and the issue of insufficient punishment intensity. He clearly stated that "the rational response of the criminal law to deepfake should be based on the judicial level", and that it should be

properly handled through the application of the current criminal law, rather than "blind criminalization in legislation" [10]. In judicial practice, it is possible to establish positions related to the investigation of deepfake technologies or to establish AI evidence review standards. Promote AI-assisted identification tools, clarify the exclusion rules for deepfake evidence, assist in reviewing the anomalies of forged evidence, and improve work efficiency. Regarding the issue of whether it was done intentionally, the algorithm can also be used to analyze the fabricated data such as time and frequency, thereby assisting in determining whether it was done intentionally. Actively promote the application of evidence preservation technology and the AI system for reporting fake content. This not only reduces the cost for victims to protect their rights but also helps them preserve evidence and safeguard their own legal interests.

4.3. Societal-Level Solutions

To reassure the public, it is possible to conduct prevention and identification publicity through official accounts with high credibility, encouraging various media and platforms to release warning cases, thereby enhancing the public's awareness of prevention. Require platforms to label AI-generated content, and during the review process, both technical and manual reviews should be conducted. Establishing a rapid reporting channel to facilitate public reporting and removal of such content. Carry out appropriate cross-departmental collaboration and integrate the existing effective resources to jointly carry out the operation. Only through government-led legislation, enterprises implementing technical prevention measures, and the public enhancing their awareness of prevention, can people better curb the occurrence of bad phenomena and maintain a good social order.

5. Conclusion

Undoubtedly, the emergence and popularization of AI deepfake technology have brought convenience to many technological industries and possess certain innovativeness. However, the risks brought about by its development cannot be ignored either. It not only enhanced the work efficiency in certain industries, such as the film and television industry, making the generated content more realistic and efficient, but also gave rise to a series of risks, such as financial fraud, infringement of personal rights, and misleading public opinion, which could easily lead to social chaos and pose new challenges to the original legal system and governance model. Although the current predicament and corresponding measures have been clarified, there are still some shortcomings. The social trust crisis caused by deepfake technology needs to be addressed. The existing detection technologies are insufficient to cope with the constantly evolving fake technologies. The lack of regulatory efforts and the weak awareness of prevention among the general public in society also pose challenges. There is still room for further research. The governance work requires the joint efforts of all sectors. It is necessary to improve the legislative mechanism, break through technical barriers, and establish network risk prevention and control systems, etc., in an effort to maintain a relatively balanced state. In the future, it is necessary to reasonably predict the development trends of technologies, such as the improvement of forgery accuracy and the expansion of application scenarios. At the same time, relevant departments should establish a sound responsibility system, emphasize the review obligations of platforms, clarify the responsibilities of all parties, achieve effective control of the risks of deep forgery, and promote the application of technology within the framework of legality and compliance to realize its value.

References

- [1] Xu Z, Zhang X, Li R, Tang Z, Huang Q, Zhang J. Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. arXiv preprint arXiv:2410.02761, 2024.
- [2] Verma N. Deepfake technology and the future of public trust in video. Doctoral dissertation, 2023.

- [3] Zheng Gaojian. Criminal Responses to the Abuse of Deepfake Technology. *Legal Science (Journal of Northwest University of Political Science and Law)*, 2025, 43(3): 56-68.
- [4] Li Tianqi, Liu Xin. Evidence Risks and Regulatory Paths of Deepfake Technology. *Evidence Science*, 2022, 30(1): 70-82.
- [5] Liu Jianming. The Fatal Threat of Deepfake to Media and Humanity. *News Enthusiast*, 2021, (4): 8-13.
- [6] Wen Ming, Sun Yuanyuan. Research on Risks and Legal Regulation of Deepfake Technology Application. *Chinese Science and Technology Forum*, 2023, (4): 158-167.
- [7] Pawelec M. Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential. *AI and Ethics*, 2025, 5(3): 2641-2666.
- [8] Kamal N. Strengthening Privacy Laws to Combat Deepfakes: An Evaluation of Current Legal Protection and Future Directions. *International Journal For Multidisciplinary Research*, 2025.
- [9] Langa J. Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes. *BUL Rev.*, 2021, 101: 761.
- [10] Jiang Ying. The Dimensions and Limits of Criminal Regulation of "Deep Fakes" Technology in Artificial Intelligence. *Nanjing Social Sciences*, 2021, (9): 101-109.