Proceedings of the 4th International Conference on Social Psychology and Humanity Studies
DOI: 10.54254/2753-7048/2026.ZJU31857

The New Criminal Hazards and Countermeasures Caused by Deepfake Technology in the Digital Age

Sijia Wang

Faculty of Social Sciences and Humanities, The University of Nottingham-Ningbo, Ningbo, China
18839378226@163.com

Abstract. In the digital age, with the rapid development of network technologies such as big data and artificial intelligence, while the lives, work, and studies of the general public have been greatly facilitated, many new types of illegal crimes have also emerged. The new type of crime brought about by deepfake technology has become a typical representative. For example, new types of crimes such as telecommunications fraud, pornography extortion, and online rumors. The paper, through the methods of literature review, case analysis, and comparative study, explores the harmful manifestations and difficulties in cracking down on deepfake crime, identifying the current governance shortcomings, proposing feasible technical investigation, legal regulation, and platform management optimization plans, and providing decision-making references for public security organs and regulatory departments. The paper finds that the new type of crime caused by deepfake technology is seriously harmful and difficult to crack down on, and effective control needs to be achieved through multi-dimensional comprehensive governance.

Keywords: Deepfake technology, New types of cybercrime, Artificial Intelligence Governance methods

1. Introduction

The term 'Deepfake' originates from the use of AI algorithms in the field of artificial intelligence, including deep learning techniques and forgery techniques [1]. Mainly including technology types such as AI face swapping, speech simulation, face synthesis, video generation, etc. This type of crime has the characteristics of high authenticity, low cost, and cross-platform dissemination [2]. While this technology demonstrates positive value in fields such as film and education, the new types of crimes caused by its malicious use are also experiencing explosive growth [3].

The significance of this study is to make up for the lack of systematic research on new types of crimes related to deep forgery technology, and to enrich the theoretical system of network security governance and new criminology. To provide technical and legal references for public security organs to combat new types of crimes, assist regulatory departments in improving legal regulations, enhance the public's ability to identify deepfake content, and maintain national security and social stability. This study focuses on the core question: What new forms of crime have been spawned by deepfake technology? What are the characteristics and hazards of crime? How to find strategies to deal with such crimes through a scientific research design?

DOI: 10.54254/2753-7048/2026.ZJU31857

2. Literature review

The term Deepfake was first used on social media platforms in 2017, on Reddit [4]. "Deepfake technology" refers to the technique of combining "deep learning" with "fake" to perform surreal digital forgery on images, audio, and video [2]. Its implementation relies on key algorithms such as encoding and decoding networks, convolutional neural networks (CNN), and generative adversarial networks (GAN) [5]. With the continuous advancement of big data and machine learning technologies, GAN, as the most widely used model, can continuously optimize the realism of forged data through adversarial iterations between generators and discriminators. The realism of the generated content has reached a level that is difficult to accurately distinguish between natural humans and artificial intelligence. This has also become a key prerequisite for it to be used by criminals to commit crimes. This kind of crime presents new behavioral characteristics and harmful forms through the Internet [6]. There are many positive examples of the use of deepfake technology in education, art, health, entertainment, and many other fields [7]. For example, in movies, different facial forms are used to shape the same character between youth and old age. On the other hand, due to the open-source algorithm code of the technology, it has been widely abused by users [1]. Users can produce fake audio and video using open source resources without professional skills, and with the massive data support provided by social media, the threshold for technology abuse continues to decrease.

According to previous research, the new types of crimes brought about by deepfake technology are mainly manifested in four categories: fraud, cyberbullying, data manipulation, and false testimony [8]. Data from 2022 shows that 38% of large enterprises have experienced over 50 identity frauds, with the highest loss reaching $480000 [9]. Deepfakes exacerbate the harm of fraud by increasing credibility, with counterfeit objects concentrated on business leaders and relatives. They use identity trust to reduce victims' vigilance and become an important incentive for business and personal property losses. Cyberbullying refers to criminals who humiliate and extort victims by forging nude photos, videos of inappropriate behavior, and other content. The core harm is the destruction of personal reputation and mental health, and the victims include ordinary groups such as students and teachers. 96% of the videos produced by deepfake technology in data manipulation crimes are pornographic content, with female victims accounting for more than 90% [1]. At the same time, false content is also generated to mislead the public. The harm covers both individual and societal levels. At the individual level, both individuals and ordinary people may suffer from privacy violations, while at the societal level, public opinion can be manipulated and even trigger international crises. The criminal content of false testimony cases is to generate false audio and video as evidence in court. Causing judicial injustice and wrongly accuses innocent people.

Deepfake cybercrime has four core characteristics. Firstly, the criminal behavior is diverse, infringing multiple legal interests. The legal interests infringed in the current cases of cybercrime involving deepfake technology mainly focus on the areas of "infringement of citizens' personality rights," "infringement of citizens' property rights," "endangering national security," and "disrupting public order" [10], such as generating nude images of victims based on GAN technology and forging faces to bypass payment authentication. The second point is that the cost of committing crimes is low, the operation is simple, and criminals do not need professional hardware and technology. They can create fake audio and video through relevant software, the moral burden and psychological concerns of the perpetrator in committing a crime are reduced due to the fact that the software used during the crime is as common as the software used in daily computer operations [1]. The third feature is that criminal activities are significant across regions. Relying on the Internet and social media, they break through the physical space restrictions [5]. The fourth point is to challenge

DOI: 10.54254/2753-7048/2026.ZJU31857

the current legal system. Deepfake technology has shaken the foundation of trust in electronic evidence, and the abuse of technology has challenged the simple common sense of "seeing is believing" (false videos can impersonate evidence). The current criminal law focuses on regulating dissemination behavior rather than production behavior [11]. And there is a lack of targeted regulation on the abuse of biometric information.

Different countries have different laws and regulations to deal with new criminal cases caused by deepfake technology. Taking the laws and regulations of China, the United States, and Australia as examples and comparisons. According to Chinese law, AI-generated content must be watermarked and labeled with the source, and companies must verify user identity and not violate personal data processing regulations [12]. Focus on content traceability and user supervision. Due to the unique national conditions in which different states in the United States independently establish their own laws, some states are at the forefront of laws and regulations regarding the establishment of deepfake technology crimes. The regulation focuses on privacy protection, election security, and protection of minors. In 2021, New York amended the law to grant victims the right to private litigation, and hold perpetrators accountable for using illegal Deepfake images to harass extortionists as a serious crime. In 2020, New Jersey enacted legislation to penalize deepfakes that maliciously damage a candidate's reputation during the election period. In 2023, Texas enacted legislation to prohibit the production of deepfake pornographic videos targeting individuals under the age of 18 [13]. Australia has not yet introduced targeted regulations and relies on existing defamation and privacy-related laws [14].

3. Data collection and analysis

The system collects academic papers, industry reports, and legal regulations related to deepfake technology, crime governance, and legal regulation from both domestic and foreign sources. From Xinhua News Agency BBC, Forbes and other authoritative media have collected typical cases of deepfake crimes from 2020 to 2025, covering four types of fraud, cyberbullying, etc., recording information such as case background, criminal methods, loss amount, and handling results. Present core data through charts, including the trend of changes in the number of articles related to "deepfakes" each year and the ranking of the number of different types of content.

Obtain in-depth crime statistics on forgery published by Cybercrime Magazine and Interpol, including the number of cases, scale of losses, and characteristics of victims. Sort out the age, gender, and occupational distribution of the criminals in the case, the proportion of victim types, the industry and regional characteristics of the crime, and clarify the high-risk groups and key areas of crime. Compare the regulatory focus and implementation effects of different countries' regulations, summarize effective governance experience and existing shortcomings, and provide a reference for optimizing the domestic regulatory system. Through the above analysis, a comprehensive presentation of the new criminal situation brought about by deep forgery technology is presented, laying an empirical foundation for the proposal of subsequent response strategies.

The new forms of crime generated by deepfake technology can be roughly divided into four categories.

Fraud: Using voice or image deepfake technology to impersonate a CEO, family member, or business leader in order to obtain funds

Cyberbullying: Creating false nude images or inappropriate videos of others (such as students, cheerleaders, etc.) for harassment or extortion.

Data manipulation: Generating pornographic content or fabricated political videos to mislead the public, disrupt elections, or trigger international crises.

DOI: 10.54254/2753-7048/2026.ZJU31857

False testimony: forging audio/video as false evidence in court to frame innocent people.

The characteristics of new forms of crime generated by deepfake technology can also be roughly divided into four categories.

Diverse Criminal Acts, Multiple Legal Interests Infringed: Compared to traditional crimes, new types of crimes that rely on deepfake technology are gradually shifting towards non-contact methods, with various types of criminal techniques emerging one after another. It involves the infringement of different individual rights (personality rights and property rights), social order, and even national security, with a wide range of legal interests affected.

Low Cost and Simple Operation: Relevant algorithms and codes are open-source on technical websites. Users can create false content through simple operations on apps or webpages without professional hardware or skills, significantly reducing the threshold for crimes and moral burden.

Strong Cross Regional Nature and Concept: With the development of digital society, new types of crimes represented by deepfake technology have gradually blurred concepts such as real space, geographical distance, and national boundaries, breaking the limitations of geographical distance and space, and providing cross-regional possibilities for various new types of crimes.

Bringing multiple challenges to the current legal system: False audiovisual works generated through deepfake technology may be used as "electronic evidence" to replace objective evidence and scientific viewpoints, shaking the trust foundation of the current evidence system. The current legal regulations are difficult to substantially regulate deepfakes and the protection of personal biometric information is not yet perfect.

4. Discussion

Public security organs should increase their crackdown efforts, strengthen organizational support and departmental linkage, establish a collaborative mechanism among multiple departments such as criminal investigation, cybersecurity, and grassroots police stations, and achieve precise strikes through intelligence analysis. For difficult cases, experts can be hired to participate in tackling them.

The key to upgrading the application of investigation technology in the technical department is to use voiceprint identification to capture sound wave deviations, audio identification to verify the consistency between mouth shape and audio, and image identification to extract subtle facial features, combined with project guidance and case synthesis mode, to break down information barriers and improve investigation efficiency.

The government should improve the legal regulatory system and clarify the civil responsibilities of manufacturers, disseminators, and platforms; Convict and sentence those who endanger national security and infringe upon citizens' rights and interests on criminal law-related charges, and build a strong legal defense line.

Strengthen the management of platform operators, require them to fulfill the obligation of identifying deepfake content, establish a rapid blocking mechanism, and cooperate with investigation and evidence collection. Public security, cyberspace and other departments regularly conduct industry supervision and investigation, simplify verification procedures, and shorten the time difference for the spread of false information.

5. Conclusion

This study aimed to investigate the harms, challenges in crackdowns, and comprehensive governance countermeasures of the new type of "deepfake" crime, and the findings indicate that this crime poses severe threats to ethical morality, legal boundaries, and normal regulatory order; The

DOI: 10.54254/2753-7048/2026.ZJU31857

analysis revealed that deepfake crimes are characterized by diverse types, low cost and simple operation, cross-regional and multiple challenges to existing laws, which supports the initial hypothesis that a multi-dimensional and collaborative governance system is urgently needed to address this issue.

This research contributes to the existing body of knowledge by systematically constructing a "law enforcement crackdown-technical investigation-legal regulation-platform management" integrated governance framework for deepfake crimes. The findings extend previous theories by providing evidence that the synergy of multiple subjects can effectively make up for the deficiencies of single-dimensional governance, thus filling the research gap of fragmented countermeasure discussions in existing literature.

This study has significant practical significance for the field of public security and social governance. The proposed countermeasures can provide direct decision-making references for public security organs to improve the efficiency of investigating deepfake crimes, help regulatory authorities optimize relevant legal systems, and guide platform operators to strengthen their main responsibilities, thereby effectively curbing the spread of deepfake-related illegal and criminal activities and safeguarding citizens' legitimate rights and interests and social public order.

This study is limited by its focus on macro-level governance countermeasures, with insufficient in-depth analysis of the specific operation details of technical identification methods such as voiceprint and image identification. One potential limitation is that the research does not involve quantitative analysis of the effectiveness of different governance measures in practical application, which may affect the accuracy of the proposed countermeasures in targeted implementation.

Future study could focus on conducting empirical research on the application effects of technical identification technologies in actual cases, and quantitatively evaluating the effectiveness of various governance measures through data collection and statistical analysis. In the future, the author will also expand the research scope to include cross-border deepfake crimes, exploring international cooperation mechanisms to address the transnational spread of such crimes.

Overall, this study provides new insights into the governance of technology-driven new-type crimes and highlights the importance of multi-subject collaborative governance in the digital age. By shedding light on the inherent characteristics and governance dilemmas of deepfake crimes, this research paves the way for the improvement of social governance systems in the era of artificial intelligence and contributes to the healthy development of the digital society.

References

[1] Qiu, Y. X. (2023). Research on deepfake cybercrime and its governance. Journal of Zhejiang Police College, 03, 87–97.
[2] Guo, E. Z., & Li, R. (2022). The harms and governance countermeasures of the new type of "deepfake" crime. Journal of Xinjiang Police College, 42(2), 51–58.
[3] Zhang, T. (2025). Criminal law responses to the technical risks of artificial intelligence "deepfake". Journal of Kunming University of Science and Technology (Social Sciences Edition), 25(5), 11–21.
[4] Kugler, M. B., & Pace, C. (2021). Deepfake privacy: Attitudes and regulation. Nw. UL Rev., 116, 611.
[5] Wu, X. D., Liu, B., & Wang, Z. J. (2023). An analysis and governance of new-type crimes related to deepfake technology. Journal of Political Science and Law, 40(6), 12–21.
[6] Ma, R. C. (2022). On the impact of new - type crimes on criminal law theory: Centering on cybercrimes. Academics, 4.
[7] Masca, M. E. (2021). Advantages and Disadvantages of Deepfake Technology. Geek Culture.
[8] Mankoo, S. S. (2023). DeepFakes-The Digital Threat in the Real World. Gyan Management Journal, 17(1), 71-77.
[9] Biometric Update. (2023, April 27). Surprise: Deepfake fraud on the rise. https: //www.biometricupdate.com/202304/surprise-deepfake-fraud-on-the-rise

DOI: 10.54254/2753-7048/2026.ZJU31857

[10] Li, S. L. (2024). Investigation difficulties and countermeasures of cybercrimes involving deepfake technology. Cyberspace Security, 15(4), 89–93.
[11] Zhang, Y. T. (2022). Legal regulation of the abusive use of "deepfake" in the era of artificial intelligence. Theoretical Monthly, (9), 118-130.
[12] Cyberspace Administration of China, Ministry of Industry and Information Technology of the People's Republic of China, & Ministry of Public Security of the People's Republic of China. (2023). Provisions on the administration of deep synthesis of internet information services. Gazette of the State Council of the People's Republic of China, (4), 22-25.
[13] Chawki, M. (2024). Navigating legal challenges of deepfakes in the American context: a call to action. Cogent Engineering, 11(1), 2320971.
[14] Celli, F. (2020). Deepfakes are coming: Does Australia come prepared?. Canberra L. Rev., 17, 193.