Proceedings of ICILLP 2025 Symposium: Property Law and Blockchain Applications in International Law and Legal Policy
DOI: 10.54254/2753-7048/2025.LD29243

The Dilemmas and Breakthroughs in Criminal Regulation of Deepfakes -Focusing on Technological Alienation and Legal Responses

Yixiao An

School of Political Science and Law, University of Jinan, Jinan, China
15898981553@163.com

Abstract. As a product of the artificial intelligence era, deepfake is of landmark significance in fields such as entertainment and art. However, its alienation has simultaneously brought about negative consequences such as personal privacy violations, threats to political security, disruption of economic order, crises of social trust, etc. Studies have shown that with characteristics of high deceptiveness and low accessibility, deepfake technology is susceptible to exploitation by criminals for fraud, defamation, and the dissemination of obscene materials. Currently, China is facing challenges such as limitations in post-incident protection models, inappropriate criminalisation standards, inadequate safeguards for biometric information, etc., in regulating deepfake crimes. Based on overseas legislative experiences and China's existing legal framework, this paper proposes optimization pathways for addressing challenges in the criminal regulation of deepfakes, which include legislative suggestions such as establishing specific criminal charges, strengthening platform liability, and enhancing criminal law protection for personal biometric information, as well as comprehensive consideration of conviction and sentencing standards at the judicial level. The study concludes that deepfake technology should be viewed rationally and comprehensively. Effective governance can be achieved by refining regulatory frameworks to ensure the technology's reasonable application and prevent its uncontrolled misuse or distortion.

Keywords: Deepfake, Criminal Law, Crime, Criminal Regulation

1. Introduction

In recent years, with the continuous advancement of computer technology, deepfake technology has been a matter of great concern for its capability to generate highly realistic videos and audio recordings, and has brought increasingly sophisticated artistic works. However, it is a double-edged sword. The negative aspects of this technology have also precipitated widespread societal risks and harms. For instance, in the civil public interest lawsuit concerning personal information protection brought by the People's Procuratorate of Xiaoshan District, Hangzhou City, Zhejiang Province against Yu, Yu acquired proficiency in using 'AI face-swapping' software through online tutorials. He then collected facial data from over a hundred individuals via internet channels. Without their

DOI: 10.54254/2753-7048/2025.LD29243

consent, Yu utilized these faces to generate fabricated obscene videos and images using this software. He subsequently disseminated these materials within specialized online groups for financial gain. Furthermore, Yu sold the "AI face-swapping" software to others, provided usage tutorials, supplied the personal facial model materials required for face-swapping, and offered customized face-swapped video services to others. He charged sales fees and service charges, and infringed upon the lawful rights and interests of many individuals.

The adverse effects of deepfake technology have sparked profound global concern, as beyond infringing upon victims' personal privacy, damaging reputations, and inflicting psychological trauma, it poses an actual and urgent threat to political security, economic order, and societal trust. In the realm of political security, deepfake technology may be employed to fabricate leaders' speeches, potentially inciting public sentiment, triggering international misjudgements, or even escalating conflicts. In economic order, fraudsters may employ deepfake to mimic the voices of corporate executives or clients to deceive victim companies, inflicting substantial losses. In terms of social trust, deepfake technology undermines the traditional notion that "Words are but wind, but seeing is believing". Its widespread application in fabricating video and audio content of influential public figures and disseminating misinformation erodes the foundational trust in information authenticity to a significant degree [1].

This paper aims to provide solutions to the challenges encountered in the criminal regulation of deepfake through a combination of approaches, including theoretical analysis, thereby contributing to the protection of legal interests and social stability.

2. Technical characteristics and alienation representation

2.1. Core principles and characteristics of deepfake

Deepfakes, as the name suggests, combine deep learning with faking, with Generative Adversarial Network (GANs) as the core technology [2]. While traditional deep learning typically operates within a single-chain architecture, GANs take a different approach by introducing an adversarial training mechanism. Its framework is built upon two neural network modules - the generator and the discriminator. The generator is specially used to produce highly realistic fake images, videos, and audio data, and is the core component of deepfake today. Besides the GAN framework, the technical system deepfake integrates three key technologies - facial feature analysis, multimedia content synthesis, and effect enhancement.

Based on the aforementioned technologies, works produced by deepfake are often highly realistic and difficult to distinguish from reality with the naked eye, making deepfake inherently "highly deceptive." Simultaneously, due to the widespread availability of open-source tools, such as the once wildly popular yet quickly removed Chinese App "ZAO", and simplified workflows and easiness in learning, deepfake also exhibits a "low threshold" characteristic [3].

2.2. The alienation of deepfake process in the application

On July 8, 2017, the State Council of China issued the Development Plan for New-Generation Artificial Intelligence. As big data technology and algorithms are deeply integrated and applied in practice, the development of artificial intelligence has progressed steadily, exerting profound and far-reaching impacts on numerous aspects of economic and social life [4]. Deepfake technology, built upon big data and artificial intelligence deep learning frameworks, represents an innovative approach within the field of intelligent audio-visual processing. The technology demonstrates

DOI: 10.54254/2753-7048/2025.LD29243

immense application potential and considerable development prospects across multiple domains, including the arts, education, and personal autonomous learning [5].

However, with the continuous advancement of deep learning and biometric recognition technologies, coupled with the aforementioned “Low threshold”, deepfake technology is no longer out of reach in people's lives. Due to its “highly deceptive” nature, deepfake is easily exploited by perpetrators for illegal activities. When such actions meet the elements of a crime, such as defamation, perpetrators may face criminal prosecution by relevant authorities, contradicting the original intent of creating deepfake technology. In this way, deepfake technology has gradually deviated from its original purpose of entertainment and artistic creation during its application, and transformed into a criminal tool, becoming as a new variable threatening social security.

3. Typical criminal forms of deepfake abuse and judicial practice challenges

3.1. Analysis of specific criminal application scenarios

Deepfake technology, due to its highly realistic simulation capabilities, can be exploited by criminals for various illicit activities, which can be primarily categorized into three types based on the nature of the crime and its harmful consequences, including fraud-related offenses, defamation and slander offenses, and crimes involving the production, sale, and dissemination of obscene materials. Next, analysis will be conducted using typical case studies.

One type is fraud-related crimes, specifically manifested as perpetrators using deepfake technology to forge identities, instructions, or credentials to commit property fraud. In Hong Kong's first multi-person AI face-swapping fraud case that occurred in 2024, a finance officer at a multinational corporation in Hong Kong was deceived by deepfake technology [6]. During a video conference, the officer saw multiple “senior executives from the UK headquarters” and ultimately lost HK$200 million to the scam. In this case, the fraudsters obtained the executives' facial features and voices through public channels, then used deepfake technology to synthesize a convincing video conference scenario where the victim was the only real person present. The deepfake's “highly deceptive” nature enabled the perpetrators to achieve their criminal objectives.

The second type involves crimes of insult and defamation, specifically manifested as perpetrators using deepfake technology to create and disseminate false audio-visual content that damages others' reputations. In Case No. Liang Gong (Feng) Kuai Xing Fa Jue Zi [2024] No. 241 handled by Liangshan County Public Security Bureau in Jining City, Shandong Province, Ren used AI face-swapping photo editing software to alter Liu's personal photos into pornographic images and posted them online, constituting an act of insult. In Case No. Luan Gong (Chang) Xing Fa Jue Zi [2021] No.0041 handled by the Luanping County Public Security Bureau in Chengde City, Hebei Province, Liang provided Man's facial photo and an explicit video to others who used AI face-swapping technology to synthesize a video. Liang also took screenshots of the fabricated video and repeatedly posted them on the overseas Twitter platform using VPN software, accompanied by insulting text and Man's contact information, causing an extremely adverse social impact. His actions constituted defamation. In both cases, the perpetrators used AI face-swapping technology as the core tool to carry out harmful acts, resulting in damage to the victims.

The third type involves crimes of production, sale, and dissemination of obscene materials, specifically manifested as perpetrators using technical means to synthesize “face-swapped” pornographic videos, or distributing and selling such videos created by others to gain illegal profits. Case No. (2023) Yu 0203 Xing Chu 11 from the Shunhe Hui District People's Court in Kaifeng City, Henan Province, serves as a prime example. In this case, Zhang uploaded many AI-generated face-

DOI: 10.54254/2753-7048/2025.LD29243

swapped obscene videos to overseas online platforms with the intent to profit. After establishing contact with clients and reaching transaction agreements, Zhang used AI face-swapping tools to produce customized obscene videos based on sample videos or specific requests provided by the clients. Then he delivered the finished products to clients and collected corresponding fees, constituting the crime of producing, selling, and disseminating obscene materials for profit.

3.2. The dilemmas facing criminal justice regulation

In response to the emerging “deepfake”, China has actively addressed its negative aspects. However, relevant authorities have encountered numerous challenges when attempting to regulate deepfake crimes through existing criminal law frameworks in judicial practice.

First, China's afterward protection model has limitations. Traditional offences such as insult and defamation require actual harmful consequences as a prerequisite for criminal liability, and thus cannot prevent the abuse of technology itself [7]. When deepfake abuse substantially infringes upon core legal interests like national security or public order, the infinite dissemination and amplified harm inherent in digital media can cause consequences to escalate unpredictably and uncontrollably. Therefore, relying solely on post-crime criminal punishment as a single measure is difficult to achieve the desired governance objectives of effectively restoring damaged legal interests and mitigating societal harm. Furthermore, if unlawful acts do not directly involve typical criminal elements such as property fraud or illegal dissemination of private images, perpetrators cannot be criminally convicted and punished under existing legal frameworks, even when severe psychological harm to victims occurs. In such cases, legal remedies have to shift toward civil compensation to protect victims' legitimate rights and interests as much as possible.

Secondly, in China's judicial practice, the volume of illegal information disseminated online is nearly equivalent to a necessary criterion for criminal liability. Applying this standard to deepfake-related crimes would raise numerous issues [8]. With the assistance and influence of technologies like algorithmic recommendations, using metrics such as click-through rates or page views as direct benchmarks for criminal liability could easily allow such cases to meet the threshold for criminal charges through algorithmic layering and deep data mining techniques, so this approach to determination is clearly unreasonable. Moreover, the dissemination of certain deepfake content may inherently serve purely entertainment purposes. The volume of such content's circulation merely reflects its reach and audience distribution objectively, without necessarily indicating whether it has caused actual harm to the legal interests protected by criminal law.

Finally, China's protection of biometric information remains inadequate. The current Criminal Law classifies biometric information as “health physiological information,” yet the scope of these two categories does not fully overlap. For instance, behavioral characteristics such as gait patterns are not covered under “health physiological information” [9]. Furthermore, in China's current legal framework, there is a lack of adequate evaluation and regulation for behaviors involving “illegal use of biometric information obtained through legitimate means.” Taking the crime of infringing upon citizens' personal information as an example, the relevant authorities have prioritized combating offences such as the unlawful sale of information, essentially targeting criminal activities within the illegal circulation of information, that is, focusing on the “unlawful acquisition” aspect. However, the primary harm of deepfake lies in “illegal use.” In numerous practical scenarios, the vast array of legally accessible materials available online already fully satisfies the requirements of deep learning. Perpetrators often do not need to resort to “illegal acquisition” to obtain information, meaning many deepfake activities operate outside the scope of criminal regulation.

DOI: 10.54254/2753-7048/2025.LD29243

4. Current status and lessons learned from criminal regulation of deepfakes domestically and abroad

4.1. Examination of extraterritorial legislative experience

In response to the issue of deepfake abuse, many countries and regions have chosen to address it through legislation, drawing on international experiences.

In June 2023, the European Parliament reviewed and approved the Authorization draft of the EU Artificial Intelligence Act. This legislation introduces tiered classification and regulation for high-risk AI systems, such as deepfake. Its "Technical Robustness and Safety" provisions mandate that AI systems must possess risk-minimization capabilities throughout their entire lifecycle, along with resilience to unexpected issues. This ensures effective defense against third-party illegal tampering or misuse, while guaranteeing functional integrity and operational controllability. At the data governance level, the bill innovatively establishes a "non-targeted collection ban," explicitly prohibiting the construction or expansion of facial recognition databases through indiscriminate extraction of publicly available internet data or closed-circuit television surveillance footage. The Ban extends throughout the entire AI system lifecycle, barring the development, market release, or actual deployment of systems based on illegally obtained data. Additionally, the bill imposes transparency requirements on generative AI systems. For content generated, AI sources must be mandatorily labeled and content legitimacy review mechanisms must be established. If training data involves copyrighted material, standardized summaries of key elements must be publicly disclosed.

The United States has enacted the Deepfake Accountability Act to address the misuse of deepfake technology, aiming to regulate its application through a legal framework and curb the spread of misinformation. The core provisions of the Act focus on the disclosure obligations of content creators, explicitly requiring that any video media files generated or modified using deepfake technology must be permanently marked with "non-erasable digital watermarks and textual descriptions" to clearly indicate the content's altered or synthetic nature. Failure to fulfill this obligation will be deemed illegal [10].

Overall, the EU places greater emphasis on the security of AI systems themselves and data compliance, while the US focuses on labeling generated content and ensuring accountability. They share the common goal of reducing the risk of deepfake abuse through legal constraints.

4.2. Analysis of China's existing legal framework

In the information age, the misuse of deepfakes can have a significant impact due to the rapid and widespread dissemination of data. The potential risks it poses cannot be overlooked, so China has also implemented certain legal responses. Article 11 of the Regulations on the Management of Online Audio-Video Information Services (hereinafter referred to as the "Regulations"), which came into effect on January 1, 2020, imposes explicit labeling obligations on both creators of deepfake content and online service providers. It requires users and providers of online audio-video information services to clearly label audio-visual materials produced and disseminated using deep learning technology that do not correspond to actual circumstances in a conspicuous and prominent manner, facilitating public identification. However, while the Regulations establish clear labeling requirements, they lack provisions addressing legal liability. They fail to set corresponding punitive clauses for violations of labeling obligations, which, to some extent, weakens the deterrent effect and enforceability of the regulations [11].

DOI: 10.54254/2753-7048/2025.LD29243

In addition to the Regulations, legal and regulatory documents such as the Decision on Strengthening the Protection of Online Information, the Cybersecurity Law, and the Civil Code have all contributed to maintaining cybersecurity. Administrative and civil regulatory approaches undoubtedly have their positive effects. However, when the harm caused by deepfake reaches a certain scale, criminal sanctions remain indispensable. Currently, China's criminal regulatory framework faces numerous challenges, including inadequate evaluation of crimes involving citizens' personal information, which demand urgent and unavoidable resolution.

5. Pathways for optimizing China's criminal regulatory framework for deepfake

5.1. Suggestions at the legislative level

Legislation, as the first step of criminal regulation, holds significant importance for the development of the entire optimization pathway. Therefore, three recommendations are proposed for the legislative text.

First, legislative bodies should consider establishing new specialized offenses such as “Deepfake Information Dissemination Crimes” or revising existing provisions like defamation and fraud to clarify their elements and sentencing guidelines, so as to precisely target core acts of malicious creation and dissemination of deepfake content as well as aiding acts.

Second, drawing on international practices, it’s necessary to strengthen the primary responsibility of platforms by stipulating their obligations to review, label, and remove deepfake content, along with legal liabilities for non-compliance. In addition, it’s necessary to reconstruct platforms' duty of knowledge at the criminal law level, fully integrating presumptions of subjective and objective fault, enabling legislative bodies or judicial interpretations to introduce the standard of “should have known [12].”

Third, the criminal law protection of personal biometric information should be strengthened as legislation has a forward-looking nature [13]. The “highly deceptive” nature of deepfake technology implies that irreplaceable personal biometric information is highly susceptible to future misuse. Relevant authorities may prevent such misuse by adopting the concept of “identity theft” as defined in the United States.

In summary, improving legislation should simultaneously address the crackdown on criminal activities, platform oversight, and personal information protection, aiming to curb the misuse of deepfake technology at its source and establish a comprehensive legal prevention and control system covering the entire process.

5.2. Response strategies at the judicial level

At the judicial level, this paper focuses on suggestions for conviction and sentencing. When exercising sentencing discretion, judicial authorities should not only consider the dissemination scope of deepfake information, such as the number of views or shares, but also treat its degree of realism, extent of improper exploitation, and actual unlawful harm inflicted upon victims as substantive criteria for assessing harm severity. The degree of realism, in a sense, reflects the difficulty of identification and the cost of clarification, which can impair victims' judgment capabilities and thus directly influence the ease of committing the crime. The degree of improper use can sometimes serve as a key factor in assessing the perpetrator's subjective intent, thereby influencing the conviction and sentencing of their actions. Some deepfake acts may not inherently meet the elements of typical criminal offenses, yet they still inflict substantial psychological or other

DOI: 10.54254/2753-7048/2025.LD29243

harm on victims, which should not be dismissed lightly. Only through comprehensive consideration of conviction and sentencing can we maximize the punishment of crimes and protect the legitimate rights and interests of victims.

6. Conclusion

It is like VR depicted in the film The Lawnmower Man in 1992. People back then could scarcely have imagined that VR would become a reality two or three decades later. In today's era of rapid technological advancement, it's equally hard to predict whether more emerging technologies will emerge the moment we open our eyes tomorrow. However, just like deepfake, every new phenomenon or technology inevitably brings fresh opportunities and challenges, which is an inescapable pattern of progress. When confronting these novel creations, we should neither succumb to excessive enthusiasm nor pessimism, but rather adopt a rational, holistic perspective, and fully leverage our knowledge and skills to address emerging issues. Regulations should be established to ensure technology is implemented responsibly, preventing its uncontrolled development from leading to detrimental distortions. Throughout this process, we should uphold our strengths while remaining open to learning from others, continuously innovating ourselves. Only through such effective governance can technological innovations like deepfake truly serve the public welfare and benefit society.

References

[1] Zheng, G. J. (2025) Criminal Law Responses to the Abuse of Deepfake Technology. Science of Law(Journal of Northwest University of Political Science and Law), 3, 56-68.
[2] Robert, C., Danielle, K. C. (2019) 21st Century - Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security. Maryland Law Review.
[3] Wei, S. Y., Liu, Y. Z. (2020) Deepfake Technology Is Undermining Trust in Cyberspace. Chinese Journal of Computers.
[4] Jiang, Y. (2021) Dimensions and Limits of Criminal Regulation for Risks Posed by Artificial Intelligence "Deepfake" Technology. Social Sciences in Nanjing, 9, 101-109.
[5] Bobby, C., Danielle, C. (2019) Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. Cali-fornia Law Review.
[6] None. (2024) Deepfake High-Level Directive Transfer Swindles Two Billion from Multinational Companies. Wen Wei Po Daily in Hong Kong.
[7] Feng, M. Y., Jiang, T. (2023) Criminal Regulation of Deepfake Abuse. Hubei Social Sciences, 4, 127-135.
[8] Wang, W. J., Ma, F. (2021) The Dilemma and Solutions for Criminalizing the Algorithm-Driven Dissemination of Illegal "Deepfake" Information. Press Circles, 1, 64-74.
[9] Li, H. S. (2020) Criminal Sanctions for the Abuse of Personal Biometric Information: A Case Study of Artificial Intelligence "Deepfakes". Political and Legal Forum, 4, 144-154.
[10] Chen, R. (2024) Criminal Regulation of Deepfakes Involving Sexual Information. Law Science, 3, 76-90.
[11] Li, M. L. (2021) Criminal Law Governance Approaches for Deepfakes. Journal of Law and Technology (Chinese and English Edition), 6, 40-47+73.
[12] Liu, X. T. (2025) Normative construction of platform criminal liability in the governance of deepfake technology. Advances in Social Behavior Research, 4, 47-53.
[13] Li, T. (2020) Construction of Criminal Regulatory Framework for Deepfake Technology. Academic Journal of Zhongzhou, 10, 53-62.