Proceedings of ICGPSH 2026 Symposium: EdTech & AI in Learning: Large Language Models in Business, Politics, and Humanities
DOI: 10.54254/2753-7048/2026.LD34995

Criminal Law Evaluation of Sex-Related Audio-Visual Works Produced by Artificial Intelligence Deepfake

Ziyi Yang, Lingjian Chen*

Henan University of Economics and Law, Zhengzhou, China
**Corresponding Author. Email: monkey403@126.com*

Abstract. The rampant spread of sex-related audio-visual works created by artificial intelligence deepfake seriously violates victims' personal dignity, and disputes remain over how to evaluate such acts. The application of the existing crime of infringing upon citizens' personal information is restricted by the conviction threshold centered solely on data quantity. Crimes involving obscene materials will lead to misalignment of legal interests and secondary objectification of victims. Proposing to add new crimes such as identity theft crime or sexual privacy infringement crime will trigger misaligned legal interests and dogmatic paradoxes regarding the authenticity of privacy, and may undermine technological neutrality and hinder the development of artificial intelligence. In response to the above problems, scholars should return to traditional criminal law dogmatics and establish a regulatory system centered on the crime of defamation.

Keywords: deepfake, sex-related audio-visual works, crime of defamation, technological neutrality

1. Introduction

Generative artificial intelligence has been widely adopted in various scenarios, especially the field of deepfake. The audio, video and images it generates are highly realistic and severely impair the judgment of ordinary people. Reviewing domestic and foreign criminal law research, most scholars discuss the criminalization of deepfake from a macroscopic perspective, while neglecting criminal evaluation targeting specific application areas. In recent years, most deepfake practices focus on generating pornographic videos featuring women, yet there is still a dilemma in evaluating such conduct. Scholars holding macroscopic viewpoints argue that deepfake content contains false information and triggers public trust crises, so deepfake acts should be uniformly regulated as a new type of crime. Some scholars propose establishing an independent crime of identity theft to handle such acts [1]. Other scholars base their arguments on sexual rights and advocate separate criminalization of acts that infringe sexual privacy [2]. Scholars who favor regulating specific conduct maintain that regardless of technical iterations, the core harm of deepfake lies in violations of traditional legal interests including victims' property and reputation. They hold that interpretable space within existing criminal law provisions should be fully utilized to regulate such conduct via established crimes. For instance, some scholars take personal information protection as the primary entry point and rationalize the regulatory path of the crime of infringing upon citizens' personal

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.LD34995

information through relevant interpretative theories [3]. To date, academia has not reached a unified and feasible framework for properly evaluating the production of deepfake sex-related audio-visual works.

Against this backdrop, this paper reflects on the above regulatory approaches to identify a reasonable standard for evaluating such conduct. By sorting out the current regulatory status of deepfake sex-related audio-visual works and analyzing the advantages and drawbacks of each regulatory idea, this paper proposes returning to traditional criminal law dogmatics. The crime of defamation shall serve as the core standard for judging the criminal liability of producers who create deepfake sex-related audio-visual works. Meanwhile, liability boundaries between content producers and technical developers must be precisely defined to strike a balance between cracking down on technical abuse and safeguarding technological neutrality.

2. Current situation and limitations of criminal regulation on sex-related deepfake audio-visual works

Deepfake is a portmanteau combining *Deep Learning* and Fake. Its core logic lies in adopting generative artificial intelligence (AIGC) to automatically synthesize and tamper with biological feature information including human body images, videos and audios through large-scale data training. Back in 2017, some users applied this technology to post face-swapped pornographic videos of Hollywood celebrities on *Reddit*, an American online forum, sparking public concerns over the proper application of deepfake technology. Such pornographic deepfake audio-visual materials deliver striking visual impact and are extremely hard to distinguish from authentic content. They target specific individuals precisely, not only degrading victims' social reputation but also inflicting devastating trauma on their personal dignity and mental health. Faced with such new-type crimes severely violating citizens' legitimate rights and interests, criminal law must provide corresponding regulatory responses.

2.1. Back-end regulation: deficiencies in applying crimes involving obscene articles

Since such videos are objectively filled with explicit depictions of sexual behaviors and carry the attribute of triggering sensory stimulation, the definition of obscene articles can cover counterfeit pornographic videos. Thus, evaluating deepfake conduct through crimes involving obscene articles bears practical applicability.

Nevertheless, deepfake pornographic videos are regulated not merely for containing obscene content. From the perspective of the Criminal Law's structural arrangement, crimes of spreading obscene articles fall under the category of Crimes of Disturbing the Order of Social Administration, whose core protected legal interest is abstract public morality. However, the severe social harm caused by sex-related deepfake audio-visual works does not essentially stem from infringement upon public morality, a social legal interest. Instead, it consists in the devastating and targeted blow to the personal dignity and reputation of specific individual victims via highly insulting fake sex-related videos. If regulated under crimes involving obscene articles, judicial authorities shall first examine whether the involved videos possess obscenity. Victims' facial features are then degraded to a physical component constituting "obscene articles", or treated as objective objects used to judge whether videos can arouse sexual desire. This strips victims of their subjective status as human beings, and completely alienates their portraits that carry personal dignity and self-identity into sexual objects and digital tools for public consumption and judicial evaluation.

DOI: 10.54254/2753-7048/2026.LD34995

2.2. Front-end regulation: deficiencies in applying the crime of infringing upon citizens' personal information

The crime of infringing upon citizens' personal information safeguards the confidentiality, integrity and disposability of citizens' personal information, with its core legal interest being citizens' right to informational self-determination [4]. In the criminal chain of deepfake production, the first step is to obtain victims' facial and other biometric identification information. Regulating such acts through this crime is reasonable, as it regards deepfake production as deprivation of the right to informational self-determination. It can remedy the defect that crimes involving obscene articles fail to protect individual rights in a targeted manner, shifting the criminal evaluation perspective from disrupted public order to impaired personal rights and interests [5]. Even so, the application of this crime still has prominent limitations.

2.2.1. Legal vacancy concerning biometric identification information

Producing sex-related deepfake audio-visual works presupposes access to biometric identification information. However, according to the *Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in the Handling of Criminal Cases Involving Infringement upon Citizens' Personal Information*, biometric identification information is not included within its scope, which only specifies health physiological information. Some scholars argue that health physiological information and biometric identification information cannot be fully equated. Biometric identification information covers physiological feature information and behavioral feature information; while physiological feature information can be interpreted by health physiological information, behavioral feature information is generally excluded [1]. Moreover, Clause 5.2.2 of the *Practice Guide to Cybersecurity Standards – Guidelines for Classification and Grading of Network Data*, formulated by the National Information Security Standardization Technical Committee in 2021, clarifies that personal biometric identification information and personal health physiological information are two distinct categories of data[1]. This results in biometric identification information being classified as other types of personal information. That is to say, criminal regulation can only be triggered when illegally acquired personal information reaches a volume of 5,000 pieces. Yet sex-related deepfake materials feature targeted attacks against single individuals; producers only need a small number of facial photos of a specific victim to generate realistic videos. Conduct targeting a single victim can hardly meet the quantitative threshold for criminal liability, rendering front-end regulatory measures largely ineffective.

2.2.2. Governance bias of valuing front-end conduct while neglecting back-end conduct

China's Criminal Law adopts a phased differentiated governance philosophy for personal information protection, focusing heavily on punishing front-end acts of illegal acquisition, sale or

[1] Practical Guide for Cybersecurity Standards – Guidelines for Classification and Grading of Network Data, Section 5.2.2: c) Personal biometric identification information: Original information and comparison information for personal biometric identification, including facial features, fingerprints, gait, voiceprints, genes, irises, and other biometric identification information. e) Personal health and physiological information: Information on personal medical treatment and health status, including personal medical information such as illnesses and hospitalization records, and personal health status information such as height and body temperature.

DOI: 10.54254/2753-7048/2026.LD34995

provision of personal information. In the context of sex-related deepfake production, however, source materials for deepfake are often photos lawfully and publicly posted by victims on social networks. Current criminal law lacks direct regulatory provisions governing illegal deepfake

processing after lawful information acquisition. More importantly, it is difficult to separate the evaluation of harmfulness between front-end information acquisition and back-end information processing. The substantial damage inflicted by deepfake on victims occurs at the stage of integrating sex-related content and disseminating finished audio-visual works. It is unrealistic to fully regulate sex-related deepfake audio-visual works merely through broad crackdowns on front-end personal information violations under criminal law.

3. Constructing a criminal law evaluation system centered on the crime of defamation

Both learning from foreign jurisdictions to add new crimes such as the crime of identity theft or the crime of sexual privacy infringement, and applying the existing crime of infringing upon citizens' personal information or crimes involving obscene articles have inherent drawbacks in application. Faced with this new type of digital violence of artificial intelligence deepfake sex-related audio-visual works, scholars shall first adopt an internal-oriented solution, review existing charges stipulated in the Criminal Law, and bring such acts into the regulatory scope of the crime of defamation. This approach can realize doctrinal logical consistency, and achieve targeted crackdown in judicial practice.

3.1. High adaptability of imputing liability via the crime of defamation

3.1.1. Compliance with protected legal interests of the crime of defamation

The essence of a crime lies in legal interest infringement. To determine the most applicable regulatory charge, it is essential to clarify the core legal interests infringed by deepfake sex-related audio-visual works. As mentioned above, such acts neither infringe real sexual privacy, nor cause core harm to abstract public morality. Instead, they deliver devastating blows to victims' personal dignity through realistic fake sex-related scenes, and such infringement essence is highly consistent with the legal interests protected by the crime of defamation.

The legal interest protected by the crime of defamation is objective social evaluation rather than subjective sense of reputation. In criminal law theories, the right to reputation is divided into internal reputation (subjective sense of reputation and dignity) and external reputation (objective social evaluation). Some scholars further propose that the crime of defamation protects functional legal interest of reputation, namely the conditions and capabilities for individuals to participate in social communication and gain social trust as social members [6]. Producers of deepfake sex-related audio-visual works take advantage of the public's common cognitive psychology of "seeing is believing", and forcibly splice victims' facial features with obscene human bodies. When such videos spread virally on the internet, no matter how firm victims' self-evaluation is, their functional reputation in the external stranger society will be damaged. Even if producers mark videos as fake content, the dissemination of pornographic videos will still reduce victims' social evaluation. Therefore, evaluating such acts by the crime of defamation shifts the protection perspective to personal dignity and focuses on victims' social evaluation, which embodies more humanistic care than crimes involving obscene articles that only focus on public morality.

DOI: 10.54254/2753-7048/2026.LD34995

3.1.2. High consistency with the objective behavioral pattern of the crime of defamation

The objective act of the crime of defamation is fabricating facts to defame others, and deepfake technology is a high-dimensional and visualized form of fabrication. Article 1 of the *Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in the Handling of Criminal Cases of Defamation and Other Crimes Committed via Information Network*[2] (hereinafter referred to as the *Interpretation on Network Defamation*) issued in 2013 defines fabrication in detail. A comparison between this provision and deepfake pornographic videos reveals high similarity between the two. Victims' facial photos or normal videos publicly released on social media can be deemed original information of others. Producers adopt deep learning algorithms such as Generative Adversarial Network (GAN) to integrate victims' facial features with characters in obscene videos, and generate videos recording non-occurring behaviors. This process constitutes fabrication, and the final generated pornographic videos of a specific person undoubtedly belong to facts damaging others' reputation. Traditional defamation relies on written and verbal language to fabricate groundless content. Although deepfake sex-related audio-visual works exist in the form of videos, from the perspective of legal essence, the core regulation object of the crime of defamation is false information capable of impairing others' social evaluation. Provisions of the Criminal Law do not impose exclusive restrictions on the presentation form of facts. The essential attribute of facts lies in verifiability of authenticity. Images and videos, just like written and verbal language, have the ideographic function of conveying information and expressing meanings. A deepfake sex-related video delivers clear information to audiences that a specific victim has committed obscene sexual acts at a certain time and place. This information points to the objective physical world and features verifiable authenticity afterwards. Therefore, bringing audio-visual materials such as images, audios and videos into the scope of facts defined in the crime of defamation does not exceed the literal meaning scope of provisions in the Criminal Law.

3.2. Excluding dissemination with indirect intent from criminal conviction

In the internet era, the immeasurable harmful consequences triggered by deepfake sex-related videos largely stem from blind reposting behaviors by numerous netizens. Imposing criminal punishment on all disseminators will lead to excessive crackdown scope and adverse effects. Therefore, after doctrinal analysis on the crime of defamation, disseminators with indirect intent shall be excluded from criminal liability.

Although Article 246 of the *Criminal Law of the People's Republic of China* does not explicitly stipulate "for the purpose of damaging others' reputation", a purpose crime does not require explicit statutory provisions. Some criminal charges are defined as purpose crimes in judicial practice and

[2] *Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in Handling Criminal Cases Involving Defamation Committed via Information Networks* (hereinafter *Network Defamation Interpretation*), Article 1: Any of the following circumstances shall be deemed as "fabricating facts to defame others" as stipulated in Paragraph 1 of Article 246 of the Criminal Law: (1) Fabricating facts that damage others' reputation and disseminating such facts on information networks, or organizing or instigating others to disseminate such facts on information networks; (2) Altering original information content involving others on information networks into facts that damage others' reputation and disseminating such facts on information networks, or organizing or instigating others to disseminate such facts on information networks. Anyone who knowingly disseminates fabricated facts that damage others' reputation on information networks, if the circumstances are flagrant, shall be deemed as "fabricating facts to defame others."

DOI: 10.54254/2753-7048/2026.LD34995

criminal law theories, even if the perpetrator's specific purpose is not clearly stipulated by law [7]. According to mainstream criminal law theories and judicial adjudication practices, the crime of defamation generally requires the perpetrator to possess the purpose of damaging others' personality and reputation. It means that to constitute this crime, the perpetrator shall not only know that the disseminated content is fabricated false facts, but also actively pursue the consequence of deterioration of the victim's social evaluation in terms of will element, namely the perpetrator must have direct intent subjectively. In the digital era, though the multiplier effect of online dissemination amplifies the objective harm of such acts, the perpetrator's foresight of harmful consequences shall not be equated with the pursuit of such consequences to curb social harm, so as to avoid the misunderstanding of objective imputation. Most ordinary netizens repost such videos merely out of curiosity, herd mentality or sensory stimulation, without the subjective purpose of damaging others' personality. Criminal regulation on such netizens will result in excessive expansion of criminal punishment. In such cases, civil liabilities including compensation for damages and apology, or relevant administrative penalties can cut judicial costs and achieve better regulatory effects.

3.3. Precisely defining the boundary of criminal liability between online platforms and technical developers

Against the backdrop of rapid artificial intelligence development, when criminal law intervenes in the governance of deepfake sex-related audio-visual works, it shall severely punish digital crimes without hindering the development of domestic artificial intelligence industry. Hence, the concept of technological neutrality shall be adhered to, to reserve reasonable fault tolerance space for underlying technology research and development as well as platform operation. Nevertheless, technological neutrality is dual-natured. Instrumental rationality requires limited legal intervention, while technological development will also bring ethical evaluation on technology application scenarios from the perspective of social rationality [8].

Blindly adding exclusive crimes for deepfake such as identity theft will easily bring neutral technical acts including algorithm-based facial feature modification into criminal regulation, violating the principle of technological neutrality. Artificial intelligence technology for fabricating facts is innocent in essence, and image synthesis based on such underlying codes is an outcome of scientific and technological progress. Deepfake pornographic videos deserve criminal regulation only because producers adopt artificial intelligence technology to fabricate and disseminate reputation-damaging facts that reduce others' social evaluation, which constitutes legal interest infringement and triggers criminal law intervention. Regulating deepfake sex-related audio-visual works via the crime of defamation sets the regulatory node after dissemination conduct. It means scientific and technological enterprises merely developing underlying algorithms such as Generative Adversarial Network (GAN), and platforms providing neutral online data storage and transmission services will not be held criminally liable. Adhering to technological neutrality and the safe harbor principle never means exempting platforms and developers from legal supervision. When the safe harbor mechanism is abused to foster cyber illegal industries, criminal law shall impose regulatory sanctions. For online platforms and technical providers, dedicated cyber crimes can be applied based on the severity of their breach of statutory obligations. For example, if platforms fail to perform the notice-and-takedown obligation under the safe harbor principle after receiving victims' complaints, or technical developers refuse to fix obvious security vulnerabilities of underlying technology (such as large-scale generation of prohibited content) and fail to perform supervision duties, and refuse to rectify after ordered by regulatory authorities, causing serious consequences, they shall be severely

DOI: 10.54254/2753-7048/2026.LD34995

punished for the crime of refusing to perform information network security management obligations stipulated in Article 286-1 of the *Criminal Law*.

4. Conclusion

The rapid development of artificial intelligence technology has spawned various digital violence crimes. Based on in-depth analysis of existing regulatory paths and reflection on legislative expansion, this paper establishes a regulatory path centered on the crime of defamation to solve the criminal law evaluation dilemma of deepfake sex-related videos. Substantively, the protection of objective social evaluation by the crime of defamation meets the protection demands of deepfake victims, and the act of producing deepfake pornographic videos conforms to the constitutive elements of fabricating facts in defamation. Meanwhile, subjective intent shall be strictly limited to direct intent in judicial application. Most importantly, this path clarifies liability division between technical developers and online platforms, safeguards the principle of technological neutrality and protects legitimate rights and interests of developers and platforms. It does not mean indulging platforms and developers from legal supervision, and dedicated cyber crimes such as the crime of refusing to perform information network security management obligations can be applied for regulation when necessary.

References

[1] Li, H. S. (2020). Criminal sanctions against the abuse of personal biometric information: A case study of artificial intelligence-based deepfakes [in Chinese]. Tribune of Political Science and Law, 38(4), 144–154.
[2] Chen, R. (2024). Criminal regulation of sexually explicit information generated through deepfake technology [in Chinese]. Legal Science, 3, 76–90.
[3] Huang, C. C. (2020). The normative transformation of regulating the crime of infringing citizens' personal information in the era of big data: Taking AI face-swapping pornographic videos as an example [in Chinese]. Journal of Huazhong University of Science and Technology (Social Sciences Edition), 2, 105–113.
[4] Liu, Y. H. (2019). The legal interest protected by the crime of infringing citizens' personal information in the context of civil code codification: Informational self-determination from the perspective of the integration of civil and criminal law and Article 111 of the General Provisions of the Civil Law [in Chinese]. Journal of Zhejiang Gongshang University, 6, 20–32.
[5] Zheng, G. J. (2025). Criminal law responses to the misuse of deepfake technology [in Chinese]. Science of Law (Journal of Northwest University of Political Science and Law), 43(3), 56–68.
[6] Che, H. (2021). The legal interest structure and procedural mechanism of the crime of defamation [in Chinese]. Chinese Criminal Law Journal, 1, 63–77.
[7] Ouyang, B. Q. (2007). Purpose crimes under typological thinking: Advocating the concept of genuine non-statutory purpose crimes [in Chinese]. Journal of East China University of Political Science and Law, 4, 66–73.
[8] Zheng, Y. S. (2018). Resolving the dilemma of technological neutrality: A jurisprudential reconsideration of the relationship between law and technology [in Chinese]. Journal of East China University of Political Science and Law, 21(1), 85–97.