ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

Defining and Regulating Criminal Legal Risks of AI-Generated Content

Xihui Liu

Law School, Shandong University of Technology, Zibo, Shandong, China
19806143396@163.com

Abstract: *With the explosive development of generative artificial intelligence (AI), its outputs, while empowering society, have also triggered significant criminal legal risks. This paper focuses on three typical behaviors: using AI to generate false information, creating deepfakes, and infringing copyright, analyzing in depth the challenges in their qualification under the current criminal law system. The study finds that these difficulties are mainly reflected in the ambiguity of act characterization, the complexity of identifying criminal subjects, the difficulty in proving subjective culpability, and gaps in the scope of criminal law regulation. To address these dilemmas, this paper proposes adhering to a proactive yet prudent view of criminal law, effectively regulating the criminal risks of AI-generated content through reasonable interpretation of existing charges, exploring the addition of specific offenses, and constructing a tiered legal responsibility system, thereby seeking a balance between encouraging technological innovation and maintaining social order.*

Keywords: *AI-Generated Content, Criminal Legal Risks, Deepfakes, False Information, Copyright, Criminal Law Regulation*

1. Introduction

Artificial intelligence, particularly generative AI represented by large language models and generative adversarial networks, is reshaping the methods of information production and dissemination at an unprecedented pace. From writing papers to generating code, from creating paintings to simulating audio and video, AI-generated content has deeply permeated various fields of social life. However, technology is a double-edged sword. When this powerful tool is maliciously exploited, it gives rise to new, more covert, and destructive illegal and criminal activities[1]. Using AI to generate and disseminate false information on a large scale can severely disrupt public order[2]; creating non-consensual intimate imagery or committing fraud through "deepfake" technology seriously infringes upon citizens' personal rights and property rights[3][4]; unauthorized use of copyrighted works to train AI and generate similar content poses a severe challenge to traditional copyright legal systems[5][6].

Confronted with these novel risks, China's current criminal law system appears inadequate. Traditional criminal law theory is based on human-centric conduct. When the actor, the act, and the harmful consequence are deeply intertwined with AI, a series of qualification difficulties arise. Should these acts be treated as traditional crimes merely employing a "technological tool," thus applying existing law, or do they require new legislative intervention due to their "qualitative change" effect? This has become an urgent contemporary question for both legal theory and practice. This paper aims to systematically sort out and analyze the criminal law qualification dilemmas of the aforementioned three types of behaviors within the framework of China's current criminal law, and attempt to propose constructive regulatory pathways, hoping to contribute to the theoretical deepening and legislative improvement in this field.

2. Challenges in Qualifying AI-Generated False Information under Criminal Law

The use of AI to generate and disseminate false information, due to its high efficiency and strong deceptive nature, poses a far greater threat to social order and public safety than traditional means. However, when applying Article 291-bis of the Criminal Law – the crime of "Fabricating and Knowingly Disseminating False Information" – core issues of matching constituent elements arise.

ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

2.1. Dilemma in Defining the Scope of "False Information"

According to the provision of this crime, "false information" is explicitly limited to "dangerous situations, epidemic situations, disaster situations, and police situations." This enumerative formulation reflects the legislator's cautious and restrained attitude, aiming to prevent the excessive intrusion of penal power into freedom of speech. However, the "creativity" of AI far exceeds the legislator's foresight. For instance, AI can generate "investigative reports" claiming a well-known food company uses carcinogenic substances, triggering social panic and industry turmoil; it can fabricate "inside information" about a financial institution's imminent collapse, causing bank runs; it can even forge "official documents" about major national policy adjustments, disrupting financial market order.

This information clearly does not fall within the traditional categories of "dangerous situations, epidemic situations, disaster situations, and police situations," yet its social harm may be no less serious. If the principle of legality is strictly followed, such behaviors would be difficult to subsume under this crime. Although considering the application of the crime of "Picking Quarrels and Provoking Troubles" as a "catch-all" offense is possible, the constitutive elements of this crime require "disrupting social order," its standard of proof is relatively high, and its application carries a degree of ambiguity, potentially leading to inconsistent application of criminal law and risk of over-expansion. This creates a gap in criminal law protection: behaviors of comparable harmfulness face entirely different legal evaluations merely due to the formal difference in information content [2][7].

2.2. Complexity in Identifying the Actor and Subjective Intent

In traditional cases, the act of "fabricating" usually refers to the actor's independent, original fabrication. However, in the context of AI generation, the actor's role shifts to that of a "prompt engineer" and content selector. The actor does not directly fabricate specific falsehoods but rather sets instructions, guiding the AI model to synthesize and infer the output from its massive training data. How, then, is the "fabrication" act to be identified? Is it the user's prompt, or the computational process of the AI model, that constitutes "fabrication"?

Closely related is the proof of subjective intent. The actor might argue that they were merely "curiously testing the AI's capabilities" and "did not intend" for the AI to generate such realistic and harmful false information. The prosecution needs to prove that the actor not only had the intent to "fabricate" but also had knowledge that the fabricated information fell within the statutory category of "false information." When user prompts are vague and the AI "overperforms," proving the actor's direct intent becomes exceptionally difficult. This may require inferring the actor's indirect intent based on factors such as the actor's professional background, subsequent dissemination behavior, and whether verification was attempted, but this undoubtedly increases the difficulty of judicial determination.

3. Criminal Risks and Regulatory Boundaries of Deepfake Technology

Deepfake technology elevates the criminal risks of AI-generated content to a new level, primarily involving the crimes of insult, defamation, fraud, and the crime of infringing on citizens' personal information.

3.1. Crimes against Personality Rights: Insult and Defamation

Using deepfake technology to superimpose another person's face onto pornographic videos is currently the most common malicious application. This act may first constitute the crime of insult under Article 246 of the Criminal Law. Deepfake videos, using "violence or other methods" to publicly damage another person's reputation, if the circumstances are serious, fully meet the constitutive elements of the crime of insult. Their realistic visual effects cause devastating trauma to the victim's reputation, privacy, and mental health. Simultaneously, if the fabricated content is not merely humiliating but involves the fabrication of specific facts sufficient to damage another's reputation, such as forging a video of a public official accepting bribes, it may also constitute the crime of defamation. Here, deepfake technology, as a super-tool for fabricating "facts," endows fictional content with unprecedented "credibility."

However, pursuing accountability through criminal pathways for such behaviors still faces a dual dilemma[4][8]. On the one hand, such cases face the constraint of the public prosecution threshold. The crimes of insult and defamation are generally crimes handled only upon complaint, placing the burden of proof on the victim. Deepfake technology is highly covert, making it difficult for victims to

ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

independently track and identify the perpetrator. Although the law allows for conversion to public prosecution when "seriously harming social order or national interests," the standard for "serious harm" involves significant discretionary space in judicial practice. On the other hand, the issues of platform liability and accomplice status need clarification[9]. The dissemination of deepfake content relies on online platforms. If a platform fails to take necessary measures promptly after receiving notice from the rights holder, does it constitute an accomplice? How to define its state of "knowing or should have known"? This requires effective linkage between criminal regulation and the platform responsibility provisions in the Cybersecurity Law and the Personal Information Protection Law.

3.2. Intertwining of Property Crimes and the Crime of Infringing on Citizens' Personal Information

The criminal risks derived from deepfake technology are not isolated. Its specific application in the realm of property crime clearly demonstrates the intertwining and superposition of different legal interests. When the technology is used to simulate specific individuals (such as relatives, friends, or superiors) to carry out precise fraud, the behavior certainly meets the constitutive elements of the crime of fraud under criminal law evaluation, with the technology serving merely as a tool for committing a traditional crime. However, stopping at this evaluation underestimates its true risk dimension and regulatory complexity.

The deeper risk lies in the formation of a closely intertwined structure of legal interest infringement between such fraud crimes and the crime of infringing on citizens' personal information. Producing deepfake videos capable of deceiving visual judgment and emotional trust technologically presupposes the illegal acquisition of large amounts of the victim's sensitive personal information, such as facial biometrics and voiceprints, which are unique and unchangeable[10][11]. Therefore, the front-end acts of illegally obtaining, providing, or even trading this personal information may themselves independently constitute the crime of infringing on citizens' personal information. This interlocking crime chain means that deepfake fraud is often the downstream monetization link, while its upstream exists as an active illegal black market for personal information[1][10].

Based on this, the criminal law regulation of deepfake technology must transcend the traditional "ex-post punishment" model and instead establish a "source governance" systemic mindset. The focus of judicial efforts should be shifted forward, strengthening the severe crackdown on upstream crimes such as the illegal acquisition, production, and provision of citizens' personal information used for deepfakes. Only by cutting off the supply chain of its "data fuel" can the space for such crimes be compressed at the source, ultimately building a comprehensive crackdown posture against the deepfake crime ecosystem.

3.3. Dilemmas in Qualifying Copyright Infringement by AI-Generated Content

Whether AI-generated content constitutes a "work" in the sense of copyright law, and the nature of AI's use of copyrighted materials during its "learning" process, are the most controversial frontier issues in the field of intellectual property criminal law today.

3.3.1. The "Work" Attribute of AI-Generated Content and Copyright Ownership

This is the premise for discussing criminal liability. If the AI-generated content itself is not recognized as a "work," then acts of copying and disseminating it cannot be discussed in terms of the crime of copyright infringement. According to China's Copyright Law, a "work" is defined as an "intellectual achievement that is original and can be reproduced in a tangible form." Where does the "originality" of AI-generated content originate? Is it from the designer of the AI model, the user providing the training data, or the end user performing instruction tuning?

The mainstream view currently holds that AI itself cannot be a copyright subject[5]. Whether its output constitutes a work depends crucially on whether the human contribution in the generation process reaches the height of "originality"[5][6][12]. If the user merely inputs simple instructions (e.g., "paint a starry night in Van Gogh's style") and the AI produces a complex and somewhat random result, the basis for recognizing the user's copyright and thus criminal law protection is weak. Only when the user's instructions are specific and unique, exerting substantial control over the expressive form of the generated content that reflects their personalized choices and judgment, is the output more likely to be recognized as the user's "work." This uncertainty directly affects the reach of criminal law protection.

3.3.2. The Transformation of the "Act of Reproduction": From Direct Copying to "Pattern Learning"

The core act of the traditional crime of copyright infringement is "reproduction and distribution," i.e.,

ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

directly and mechanically copying another's work without permission. However, the infringement mode of AI has undergone a fundamental change. It does not directly copy segments of a specific work but rather, by "reading" a massive number of works (including many copyrighted works), learns an abstract "style" or "pattern," and subsequently generates "new" content that is not substantially similar to any existing work in its entirety but highly approximate in style and essence.

Does this behavior constitute "reproduction" in the criminal law sense[5][13]? If the training data includes copyrighted works used without permission, does the training process itself constitute an act of "reproduction"[6][13]? From a technical perspective, the training process indeed involves temporary reproduction of data. From the perspective of legal interpretation, this aligns more closely with 'learning' and 'comprehension,' rather than a 'public-facing expression.' Directly recognizing this as infringement of the reproduction right, and considering the entire commercial AI generation activity as "distribution," undoubtedly poses a significant challenge to traditional copyright theory. Criminal law, as the last resort, should be more conservative on this issue, avoiding premature and hasty intervention into a civil infringement area still under intense debate[5][12].

3.3.3. Determination of the Subjective "Purpose of Profit"

The crime of copyright infringement requires the actor to have the "purpose of profit." In the context of using AI to generate potentially infringing content, the determination of this subjective element is also becoming more complex. For example, a company uses pirated book data to train its AI model, then offers free text generation services to attract user traffic, and finally profits through advertising or premium services. The causal chain between its "purpose of profit" and the specific infringing act is long and indirect, posing new challenges for judicial determination.

4. Exploration of Regulatory Pathways for Criminal Risks of AI-Generated Content

Faced with the aforementioned qualification dilemmas, we must neither remain stagnant, allowing the law to lag behind technological development, nor overreact, readily resorting to punishment that stifles innovation. We should adhere to a proactive yet prudent view of criminal law and adopt a multi-level, systematic regulatory approach.

4.1. Interpretive Path: Fully Tapping the Potential of Existing Criminal Law Norms

Within the existing legal framework, partially harmful AI-generated content-related behaviors can be brought under regulation through reasonable interpretation.

On one hand, a moderate expansive interpretation of "false information" can be adopted [2][7]. Prior to legislative amendment, through judicial interpretation or guiding cases, false information comparable in harmfulness to "dangerous situations, epidemic situations, disaster situations, and police situations" – sufficient to cause social panic or economic turmoil – could be considered under the "other methods" of the crime of Picking Quarrels and Provoking Troubles, but its scope of application must be strictly limited, adhering to the principle of necessity.

On the other hand, the criminal punishability of deepfake behaviors should be clarified [4][8]. By publishing typical cases, it should be clarified that using deepfake technology to produce or disseminate obscene materials, or for insult, defamation, or fraud, will be severely punished according to law. Simultaneously, the crackdown on upstream acts of illegally obtaining and providing citizens' personal information should be strengthened[10], cutting off the "raw material" supply chain for deepfakes.

4.2. Legislative Path: Exploring the Addition of Specific Offenses and Amending Constituent Elements

When the interpretive path cannot effectively cover new risks, legislative intervention becomes necessary.

Consider adding the crime of "Producing or Disseminating Deepfake Items"[8][14] . Drawing on foreign legislative experience (e.g., some U.S. states' Deepfake Laws), consider establishing a specific offense. This crime could be structured as a conduct crime, directly criminalizing the act of producing or disseminating deepfake images of others without consent, and setting different sentencing levels based on whether the content is pornographic or has other serious harms. This would provide more preemptive and comprehensive protection for citizens' personality rights.

ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

Amend the crime of "Fabricating and Knowingly Disseminating False Information[2][7]." To enhance regulatory flexibility, it is advisable to adopt a "list plus catch-all provision" model for defining "false information," rather than relying on a closed list. This could be achieved by incorporating a clause covering "other information that gravely disrupts social order," thereby capturing unforeseen yet harmful scenarios. This would provide necessary flexibility for judicial practice, while simultaneously clarifying the standard for "seriously disrupting social order" through judicial interpretation to prevent abuse.

4.3. Systematic Path: Constructing a Multi-dimensional Legal Responsibility System

Criminal law regulation is only a last resort and must work in synergy with other legal means, specifically including the following three approaches.

Strengthen administrative supervision and civil compensation[1][12]. Cyberspace, industry and information technology, and other relevant departments should strengthen supervision over AI service providers, requiring them to establish content review mechanisms, prominently label deepfake content, and fulfill "notice-and-takedown" obligations. In the copyright field, civil infringement litigation should be prioritized to resolve disputes, clarifying the copyright boundaries of AI-generated content and the legal standards for using training data.

Promote technological governance and industry self-regulation[1][15]. Governments and regulatory bodies should encourage research institutions and technology companies to advance the research, development, and deployment of AI-powered content identification and tracing technologies, leveraging technological solutions to address technological challenges. They should also actively promote the establishment of ethical guidelines and technical standards by industry organizations, mandating that enterprises comply with social responsibilities and mitigate the risks of technology misuse at the source.

Clarify the responsibilities of all parties. Regulatory authorities should clearly define the division of responsibility and liability among all relevant parties—including AI model developers, service providers, and end-users—in relation to criminal risks arising from AI systems. For platforms, the application of the "safe harbor" rules and the "red flag" standard should be refined[9], encouraging them to adopt proactive preventive measures.

5. Conclusion

The wave of generative artificial intelligence has arrived, and the criminal legal risks it brings are real and pressing. The analysis in this paper demonstrates that in areas such as using AI to generate false information, creating deepfakes, and potential copyright infringement, the current criminal law faces severe challenges in act characterization, subject identification, and subjective attribution. These challenges are rooted in the deep tension between the autonomy and "black box" nature of AI technology and the human-conduct-centric tradition of criminal law[16][17][18].

The solution lies in seeking a dynamic balance between technological innovation and legal regulation. We should first address the most urgent challenges through reasonable interpretation of existing laws, maintaining the stability and adaptability of criminal law. For institutional loopholes that cannot be resolved through interpretation, the feasibility of legislative improvement should be prudently studied, considering the addition of specific offenses. Ultimately, an effective regulatory system must be a comprehensive governance system combining criminal, civil, and administrative means, and synergizing law, technology, and ethics[1][15][19]. As future legal professionals, we bear the mission of our time to understand technology, interpret the law, and shape rules. We must find the path of wisdom between guarding social fairness and justice and embracing the technological revolution.

References

[1] Zhang, Y. Criminal Risks of Generative Artificial Intelligence and Its Countermeasures. Chinese Criminal Science, 2023 (5),78-95.
[2] Jia, Y. On the Criminal Law Qualification of Using Artificial Intelligence to Generate False Information. Law Science Magazine, (2023)44(7), 122-136.
[3] Wang, S. Z. The Alienation of Crime in the Digital Age and Its Criminal Law Response: An Analysis Based on Deepfakes. Legal Forum, (2022)37(6), 112-123.
[4] Wang, H. W. Criminal Governance of Deepfake Technology: From False Information to Personality Rights Infringement. Chinese Criminal Science, (2021) (5), 105-121.

ISSN 2616-7433 Vol. 7, Issue 8: 28-33, DOI: 10.25236/FSST.2025.070805

[5] Sun, D. C. Copyright Identification and Criminal Protection Boundaries of AI-Generated Content. Intellectual Property, (2022) (11), 56-71.
[6] Yang, N. Research on Copyright Issues of AI-Generated Content in the Digital Economy. Electronics Intellectual Property, (2022) (8), 45-58.
[7] Liu, Y. H. The Change of Legal Interest in Data Crime and the Path of Imputation in the AI Era. Chinese Journal of Law, (2023)45(4), 99-117.
[8] Jiang, Y. The Dilemma and Outlet of Criminal Law Regulation on 'AI Face-Swapping' Behavior. Journal of National Prosecutors College, (2023)31(2), 155-170.
[9] Guo, Z. L. Research on the Rules for Identifying 'Knowing' in Platform Liability for Cybercrime. Law Review, (2022)40(4), 178-190.
[10] Li, C. Judicial Expansion and Restriction of the Crime of Infringing on Citizens' Personal Information in the Context of Deepfake Technology. ECUPL Journal, 2023(4), 133-145.
[11] Solove, D. J. The Myth of the Privacy Paradox. George Washington Law Review, 2021 (1), 1-51.
[12] Zhou, G. Q. The Criminal Law Response to New Types of Cybercrimes: From the Perspective of AIGC and Deepfakes. Peking University Law Journal, 2023(5), 1157-1175.
[13] Lemley, M. A., & Casey, B. Fair Learning. Texas Law Review, 2021(4), 743-786.
[14] Gao, Y. D. The Reconstruction of the Criminal Law Evaluation System for AI-Synthesized Videos. Chinese Criminal Science, 2023 (2), 89-103.
[15] Zhang, M., & Zhu, R. Regulating AI-Generated Content: A Chinese Perspective. Computer Law & Security Review, 2023, 48, 105-123.
[16] Chen, X. Difficulties and Solutions in the Criminal Imputation of Generative AI Models. Tribune of Political Science and Law, 2023 (1), 88-101.
[17] Liu, X. Q., & Lin, Y. J. The Challenge of Criminal Imputation for AI's 'Black Box' Decision-Making and Its Resolution. Contemporary Law Review, 2024(1), 87-99.
[18] Lao, D. Y. Risk Distribution and Criminal Imputation: Reflection on Causation Theory. Tribune of Political Science and Law, 2020 (6), 3-18.
[19] Ouyang, B. Q., & Wang, Q. The Construction of a Criminal Imputation System for Generative Artificial Intelligence. Journal of Comparative Law, 2023 (3), 134-149.