Criminal Regulation of False Information Generated by Artificial Intelligence

Chaonan Li
School of Law, Xinjiang University of Finance and Economics, Urumqi, Xinjiang, China

Abstract: The widespread application of generative artificial intelligence technology has boosted social progress, while also giving rise to a new type of criminal risk characterized by the large-scale and low-cost dissemination of false information. From the perspective of criminal law, this paper systematically analyzes the technical causes of false information generated by generative artificial intelligence and categorizes such risks into system-endogenous risks and malicious abuse risks. The study finds that such risks pose systemic challenges to the current criminal law system, which are centrally manifested in the predicaments including ambiguous attribution of liabilities among multiple subjects, difficulties in proving subjective fault, and the lagging application of existing criminal charges. To address the aforementioned issues, this paper argues that on the premise of abiding by the principle of legality of crime and punishment and the principle of modesty of criminal law, regulatory loopholes should be filled through the purposive expansive interpretation of existing legal norms. Meanwhile, it is necessary to construct the performance standards of the duty of care for liable subjects and establish rules for identifying subjective intent based on whether the duty of care has been fulfilled.

Keywords: Medical Data; Generative Artificial Intelligence; False Information; Criminal Regulation

1. Introduction of the Research Issue

With In the era of advancing maturity of artificial intelligence technology, generative artificial intelligence technologies represented by ChatGPT, Sora and DeepSeek have been gradually applied to all aspects of human life and learning by virtue of their robust language synthesis and deep learning capabilities. The growing sophistication of this technology has greatly enhanced social production efficiency and unlocked innovative potential, while also giving rise to new governance challenges. An urgent issue currently confronting society is the proliferation of false information generated through the use of artificial intelligence. The core distinction between generative artificial intelligence and traditional artificial intelligence lies in its technical underpinning of large language models, which integrates web crawling technology to collect big data on a large scale and conducts continuous learning to enrich its language databases. The operational process of a generative artificial intelligence system consists primarily of three stages: data input, model operation and synthesis, and data output. Specifically, the system rearranges and reorganizes existing information in accordance with an actor's instructions, and ultimately generates and outputs content that meets the requirements specified in the instructions. Owing to inherent limitations in the model's capacity to identify false information, it cannot independently and accurately detect all false content. If false information is present in the original dataset, the content ultimately generated by the system will also contain such false information. In recent years, criminal cases in which offenders exploit the robust deep learning and information synthesis capabilities of generative artificial intelligence to generate false information have become increasingly prevalent. A notable example is the 2024 case of Qin, who, for the purpose of profit, used generative artificial intelligence technology to create deepfake videos, produced and trafficked obscene video content, and whose conduct constituted a serious criminal offense. He was consequently convicted of the crime of producing and trafficking obscene materials for profit in accordance with criminal law. In 2023, the appellants Zhang, Fu and Li, with the intent of illegal possession, colluded with others to generate fake videos by means of information network technologies and defraud public and

private property of an extremely large amount, and were consequently convicted of fraud. In response to crimes involving the use of artificial intelligence, the Criminal Law of the People's Republic of China (hereinafter referred to as the Criminal Law) has also formulated appropriate regulatory provisions for offenses related to generative artificial intelligence. For example, Article 286-1 stipulates criminal penalties for network service providers who fail to perform their information network security management obligations as prescribed by laws and administrative regulations. Existing legal norms and governance measures targeting false information are primarily directed at illegal or false information produced and disseminated with human participation, whereas the regulation of the spontaneous generation of false information by generative artificial intelligence remains inadequate [1].

At present, some scholars in the academic community have keenly recognized the implications of this new type of criminal risk. Among existing research, Liu Xianquan points out that the application of generative artificial intelligence has led to a surge in risks associated with data utilization acts, shaking the current regulatory logic centered on "data control" [2]. Pi Yong has systematically analyzed the obstacles to the application of criminal law in relation to false information in various human-computer interaction scenarios, and revealed the inadequacy of the traditional criminal charge system in responding to "autonomous" technological outputs [3]. Deng Hongguang et al. have analyzed the compound risks of false information at the data layer, model layer and output layer, as well as the necessity of collaborative governance by multiple subjects, from the perspective of the governance system [4]. All existing academic research in the field points to a core proposition: against the backdrop of the current digital age, the criminal law normative system constructed with natural persons at its core is confronted with multiple systemic predicaments—including gaps in behavioral regulation, ambiguity in subject attribution, difficulties in subjective determination, and the failure of normative interpretation—when responding to the new criminal risks posed by generative artificial intelligence, which features a high degree of autonomy and social penetration. Against this backdrop, this paper aims to sort out and integrate the existing academic debates, and more importantly, attempts to explore a regulatory path that not only conforms to the principle of legality of crime and punishment and the principle of modesty of criminal law, but also effectively addresses technological risks and guides technological innovation amid the tension between technological iteration and the guarantee of the rule of law.

2. A Typological Analysis of the Risks of False Information Generated by Generative Artificial Intelligence

The causes of false information generated by generative artificial intelligence through deep synthesis are mainly manifested in two forms: first, the generation of false information stemming from inherent flaws of the system itself, such as its training data and the reinforcement learning from human feedback (RLHF) model; second, the creation of false information as a result of criminals exploiting the deep synthesis capabilities of generative artificial intelligence to input instructions related to false information.

2.1 System-Endogenous Risks

System-endogenous risks arise from the inherent uncertainty and unreliability of generative artificial intelligence technology. Under such risks, the generation of false information is not induced by any subject with unlawful intent; instead, it is a "defective output" resulting from the system’s intrinsic technical flaws during its autonomous operation, which is specifically manifested in the following aspects. First, the deep synthesis of generative artificial intelligence is based on statistical patterns rather than factual logic, which causes the system to confidently fabricate non-existent facts and cite fictitious information. Second, the training of generative artificial intelligence relies heavily on data, and the quality of data directly determines the output of the model. If the training data contains data related to false information, it will lead to the formation of an erroneous model, which may generate false information in specific contexts [5]. Third, under the influence of the reinforcement learning from human feedback (RLHF) model, the model generates information through human-like thinking but is incapable of in-depth causal reasoning. It may erroneously describe merely correlated events as having a causal relationship and produce misleading

analyses with fundamental logical flaws. In summary, system-endogenous risks can lead generative artificial intelligence systems to produce completely fictitious yet formally rigorous content—such as fake academic papers, legal cases, and news reports—in a highly confident tone. This may result in more severe harmful consequences if the users are minors with limited cognitive and discriminatory abilities.

2.2 Malicious Abuse Risks

Malicious abuse risks represent the most typical and direct form of criminal risk among the incidental criminal risks of generative artificial intelligence, referring to the act of criminals exploiting generative artificial intelligence technology to commit crimes and achieve criminal purposes. Under such risks, while advancing social progress, generative artificial intelligence technology has become a "facilitator" of criminal acts and lowered the threshold for committing crimes. Generative artificial intelligence technology maintains a neutral stance, merely providing technical support, and the false information generated is endowed by the actor's criminal intent; the root cause of such risks lies in the actor's subjective viciousness. For example, anyone can produce highly realistic forged videos and fake images, or even tens of thousands of human-like fraud scripts and imitate anyone's voice, simply by inputting instructions driven by criminal intent. Thanks to the facilitation of generative artificial intelligence, the severity of criminal consequences far exceeds that of traditional crimes of the same type. Under such risks, the generation of false information by generative artificial intelligence also poses new and pressing problems for the traditional constitutive element system of criminal law, such as the determination of subjective fault and the distinction between joint crimes and neutral aiding acts.

3. Dilemmas in the Cognizance of Criminal Liability for False Information Generated by Generative Artificial Intelligence

Various risks arising from false information generated by artificial intelligence have inflicted a multi-dimensional and in-depth impact on the current criminal law system. The cause of such impact lies not in legal lag, but in the systemic inadaptation stemming from the conflict between technical logic and legal logic at the fundamental paradigm level.

3.1 Dilemmas of Subject Identification under Multi-Party Participation

Regulations the actors involved in the generation of false information by artificial intelligence include users, system designers, and service providers, all of whom participate directly or indirectly in the generation process. Under the traditional criminal liability system, an actor bears criminal liability for harmful consequences based on criminal causality. There must exist a socially appropriate, positively causal or inhibitory relationship between the actor's conduct and the harmful result. Even in the case of indirect perpetration, the actor's dominant control over the result must be established[6].The algorithm of artificial intelligence is formed by statistically weighting training data with a large number of parameters, and it is correlated to varying degrees with both system designers and users [3]. Due to the autonomy of artificial intelligence systems, which can directly cause harmful consequences, in the process of the generation and dissemination of false information by generative artificial intelligence, the acts of subjects in the artificial intelligence value chain, such as service providers, model developers, and users, may all have a causal relationship with the generation and dissemination of false information, resulting in difficulties in attributing liability for false information generated by generative artificial intelligence [7].

3.2 Dilemmas in the Identification of Subjective Fault

Informed the subjective element of a crime is central to criminal liability. However, the interactive nature of generative artificial intelligence makes it impossible to prove the actor's subjective intent. In many current criminal cases involving false information, criminal liability shall only be imposed on the actor if he/she knows that the information generated by the generative artificial intelligence is false and still disseminates it [8]. Yet users may not be aware of the specific false details in deep synthesis content, and for service providers, current technology cannot in most cases establish that they knew users were using their services to commit specific criminal acts. At the same time, how to pursue liability for

negligence in cases involving false information generated by artificial intelligence has become an acute practical issue. This mainly involves two aspects. On the one hand, there is the establishment of the standard of duty of care. It is unrealistic to require system developers to foresee all harmful false information that their models might generate in all scenarios. How should this duty of care be defined: to ensure that models possess basic fact-checking mechanisms, or to require accuracy approaching human-level performance? On the other hand, there is the boundary of permissible risk. The development of new technology is inevitably accompanied by new risks, and to what extent should criminal law tolerate technological uncertainty? If false information constitutes an inevitable side effect of technological innovation, the technical designer may not be held liable. Conversely, if such a risk is avoidable, the designer shall be under a duty of care. Failure to fulfill such duty shall result in legal liability [9].

3.3 Dilemmas in Legal Application

China's current criminal law charge system demonstrates structural inapplicability when addressing false information generated by generative artificial intelligence. A prominent problem is the overly narrow scope of offenses. For example, Article 286-1 of the Criminal Law of the People's Republic of China defines false information as "false information about dangers, epidemics, disasters, or police situations," which cannot cover equally or even more harmful false information generated by generative artificial intelligence, such as false political statements, financial information, and academic achievements. Second, there is a misalignment in charge regulation. The crime of illegally obtaining computer information system data under Article 285 of the Criminal Law targets "obtaining data stored, processed, or transmitted in a computer information system by other technical means". However, generative artificial intelligence systems conduct deep synthesis to create new information based on existing data, which does not conform to the defined act of "obtaining". In addition, issues exist regarding subject applicability. The term "network service providers" as stipulated in the 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 such as Illegal Use of Information Networks and Assistance in Network Criminal Activities (hereinafter referred to as the "Interpretation") does not properly cover generative AI platforms. Although generative artificial intelligence overlaps with search engines in some functions, it does not essentially fall within the scope of network service provision, making it impossible to apply the crime of refusing to perform information network security management obligations to such subjects [10].

4. Criminal Law Regulatory Strategies for False Information Generated by Generative Artificial Intelligence

4.1 Teleologically Oriented Expansive Interpretation of Existing Charges

4.1.1 Teleological expansive interpretation of the crime of refusing to perform information network security management obligations

As a representative Service providers exercise significant control over content generation and other aspects of the interaction between systems and users. Therefore, it is necessary to impose a statutory review obligation on service providers. In a practical case, a company used artificial intelligence technology to generate false information exaggerating a disaster situation and refused to rectify the situation on the grounds that the false information was produced by a third-party technology company. Meanwhile, artificial intelligence service providers do not substantially fall within the definition of "network service providers" under the Interpretation. Accordingly, the author suggests that a teleological expansive interpretation should be applied to the existing statutory term "network service providers" in Article 286-1 of the current Criminal Law concerning the crime of refusing to perform information network security management obligations, through legislative interpretation, judicial interpretation or guiding cases. Artificial intelligence service providers provide users with model access interfaces and content generation services via the Internet, which essentially fall within the scope of "providing services by using information networks". Compared with traditional network access, storage, search and transmission services, generative artificial intelligence services are technically more sophisticated, yet their essence of "providing network services"

remains unchanged. Criminal law terminology must adapt to social development. As long as such interpretation does not exceed the possible meaning of the terms, it should be permitted to interpret them in line with the progress of the times so as to cover operators of generative artificial intelligence services and urge service providers to fulfill their information network security management obligations.

4.1.2 Teleological expansion of the definition scope of the crime of fabricating and intentionally spreading false information

Currently, the powerful synthetic technology of artificial intelligence has lowered the production and manufacturing costs of false information. However, the definition of the crime of fabricating and intentionally spreading false information under the existing Criminal Law is limited to "dangers, epidemics, disasters, and police situations". The specific types of false information generated by generative artificial intelligence have essentially reached the same level of harmfulness as the false information related to the "four types of situations" (dangers, epidemics, disasters, and police situations).For example, false financial information generated on a large scale by means of artificial intelligence may trigger abnormal fluctuations in the stock market and bank runs, and its harm to the economic order is equivalent to that of false "disaster" information; false information fabricated against major public policies may incite social antagonism and trigger social unrest, and its damage to social order is no less than that of false "police situation" information. These new types of false information are characterized by large-scale, precise, and highly simulated dissemination through artificial intelligence technology, and their destructive power has even exceeded that of false information produced by traditional means. Therefore, on the premise of adhering to the principle of legality (i.e., no crime without a law), the author suggests that some false information generated by generative artificial intelligence, whose level of harm is equivalent to that of the false information related to "dangers, epidemics, disasters, and police situations", should be incorporated into the evaluation scope of the crime of fabricating and intentionally spreading false information through teleological interpretation. Specifically, such false information shall be included in the category of "other acts that seriously harm social order and national interests" as stipulated in the Criminal Law, so as to accurately regulate new types of criminal acts and balance the protection of technological innovation and the maintenance of social public order.

4.2 The Refinement of the Specific Content and Performance Standards for the Duty of Care of Relevant Subjects

Under the technical scenario of artificial intelligence, different subjects shall bear different levels and types of duty of care due to differences in their status and control capabilities in the chain of technology development, deployment, and use. To reasonably define the boundary of criminal negligence liability for multiple subjects, it is necessary to further clarify the specific content and performance standards of the duty of care. In the process of criminal regulation, the intentional indifference of the subject shall be inferred based on whether the liable subject has fulfilled its duty of care.

4.2.1 Duty of care of system developers

As the source of technology, the duty of care of design and development entities shall focus on the "duty of risk prevention", that is, taking reasonable measures during the technology design stage to reduce the risk of the system being abused. Firstly, during the data collection and cleaning phase, technical and management measures shall be adopted to reasonably identify and eliminate known sources of large-scale false information, thereby preventing the system from internalizing the generation mode of false information due to data contamination [11]. Secondly, in the model training stage, safety alignment mechanisms such as reinforcement learning from human feedback shall be adopted to set reasonable safety barriers and reduce the possibility of the system being induced to generate false information. Thirdly, sampling monitoring of model outputs shall be conducted; for the identified systematic false information generation modes, model repairs or risk prompts shall be carried out in a timely manner. The performance standard for the performance of the aforementioned obligations shall be "reasonable technical level and industry practices", rather than requiring design and development entities to foresee and prevent all possible abuse scenarios. Criminal law shall tolerate the residual risks that cannot be completely eliminated under the current technical level. Criminal negligence liability shall only be

considered when the design and development entity seriously violates the aforementioned reasonable obligations and such violation has a causal relationship with the result of serious legal interest infringement.

4.2.2 Duty of care of service providers

As the intermediate link between technology application and user interaction, the duty of care of service providers shall focus on the "duty of process control and post-event response", that is, conducting reasonable supervision over the service operation process and promptly handling the identified illegal acts. Firstly, ensuring the performance of the identification obligation: in accordance with the provisions of the Measures for the Identification of Artificial Intelligence Generated and Synthetic Content and other relevant regulations, implement and ensure the effective operation of mandatory identification technical measures for AI-generated content, so as to ensure that users can identify the source of information. Secondly, user management obligation: establish a user registration and behavior monitoring mechanism, and take disposal measures such as warning, function restriction, and account closure in accordance with the law for users who use their services to engage in illegal activities on a large scale. Thirdly, report handling obligation: establish convenient user reporting and complaint channels, and promptly handle illegal and false information that has been effectively reported. The performance standard for the performance of the aforementioned obligations shall be "reasonable care" rather than "absolute control". When a service provider fails to establish a basic review mechanism or turns a blind eye to obviously illegal information, it may be determined that it has failed to fulfill its duty of care.

4.2.3 Duty of care of users

As the terminal link of technology application, the duty of care of users shall focus on the obligation of legal use, that is, not using the technology for illegal purposes and assuming the reasonable review liability for the content generated by themselves using the technology. Firstly, no malicious abuse: no use of generative artificial intelligence to commit illegal and criminal acts such as fabricating and spreading false information. Secondly, content review obligation: for the content generated by artificial intelligence and intended to be disseminated to the public, the obligation of reasonable authenticity review shall be fulfilled, and false content shall not be disseminated with knowledge of its falsity. Thirdly, obligation to maintain identification: no malicious deletion, tampering with, or circumvention of the legal identification of artificial intelligence-generated content; the circumstance of unintentional failure to identify shall be excluded to avoid absolute liability imputation. The performance standard for the user's duty of care requires users to fulfill the basic obligation of verifying the authenticity and source when disseminating artificial intelligence-generated content. For those who use artificial intelligence to commit obvious illegal acts, their intent may be directly determined; for those who disseminate false information negligently, criminal liability shall only be considered when it causes serious harmful consequences and the actor obviously fails to fulfill the reasonable duty of care.

5. Conclusion

The rapid development and widespread application of artificial intelligence have not only promoted the improvement of social production efficiency and innovated innovation models but also given rise to a new type of governance challenge characterized by the proliferation of false information. Focusing on the practical issue of false information generated by generative artificial intelligence, this paper systematically analyzes its technical causes, risk types, and the multi-dimensional impacts of such false information on the existing criminal law system, and attempts to propose responsive and constructive criminal law regulatory paths. The research shows that the generation of false information by generative artificial intelligence stems from the inherent "system-endogenous risks" of the technology itself and the "malicious abuse risks" of technical tools by external actors. This dual risk not only significantly lowers the threshold for the production and dissemination of false information but also makes its harmful consequences more concealed, diffusive, and severe. Against this backdrop, the traditional criminal law normative system constructed around natural persons faces systematic dilemmas in key links such as act regulation, liability attribution, subjective determination, and legal application.

This paper argues that on the premise of adhering to the principle of legality and the principle of criminal law modesty, the path of

teleological expansive interpretation should be adopted to include generative artificial intelligence service providers in the scope of "network service providers" and then apply the provisions on relevant omission crimes; at the same time, new types of false information whose harm degree is equivalent to that of the statutory false information related to "dangers, epidemics, disasters, and police situations" should be incorporated into the regulatory scope of the crime of fabricating and intentionally spreading false information. In addition, it is necessary to clearly construct the specific content and performance standards of the duty of care of liable subjects, refine and clarify the duty of care of system developers, service providers, and users, require all subjects to strictly perform the corresponding duty of care, and take "failure to fulfill the duty of care" as an important basis for inferring the actor's reckless mental state.

Based on the integration of existing academic research results, this study is a preliminary exploration on the criminal law regulation of false information generated by generative artificial intelligence, which still has certain limitations. Future research should further focus on the construction of cross-legal collaborative regulation and multi-subject collaborative governance mechanisms, while closely monitoring the risk evolution rules of false information in the process of technological iteration, continuously improving the criminal governance system of generative artificial intelligence, and providing more targeted theoretical support and practical guidance for criminal law regulation practice in the digital age.

References

[1] Gong Pengcheng, Fu Chenglu. Criminal Risks and Legal Responses to False Information Generated by Generative Artificial Intelligence. Journal of Jiangsu Police Institute, 2024, 39(3): 64-71.

[2] Liu Xianquan. A New Path for the Criminal Law Regulation of Data Crimes Involving Generative Artificial Intelligence. Contemporary Law Review, 2024, 38(6): 3-15.

[3] Pi Yong. Criminal Law Governance of False Information Generated by Artificial Intelligence——Drawing on the Safety Risk Prevention and Control Mechanism in the EU's Artificial Intelligence Act . Journal of Comparative Law, 2025, (1): 75-90.

[4] Deng Hongguang, Wang Xuefan. Risks and Responses in the Governance of False Information Generated by Generative Artificial Intelligence. Theory Monthly, 2024, (9): 115-129.DOI:10.14180/j.cnki.1004-0544.2024.09.012.

[5] Ma Haiming. Research on Criminal Risks and Legal Responses to False Information Generated by Generative Artificial Intelligence. Hebei Legal Vocational Education, 2025, 3(07): 85-90

[6] See Ma Kechang (Ed.). General Theory of Crimes. Wuhan University Press, 1999, pp. 546-548.

[7] Chen Xiaobiao, Yin Jiehui. Classification of Criminal Risks and Hierarchical Governance of False Information Crimes Involving Generative Artificial Intelligence. Journal of Kunming University of Science and Technology (Social Sciences), 2026, 26(01): 1-11. DOI: 10.16112/j.cnki.53-1160/c.2026.01. 211.

[8] Yang Jianmin. Challenges to the Sanction System of False Information Crimes Posed by Generative Artificial Intelligence and the Responses. Journal of Northeastern University (Social Sciences), 2025, 27(03): 98-105. DOI: 10.15936/j.cnki.1008-3758.2025.03.011.

[9] Jiang Tao, Guo Xinyi. Criminal Law Imputation of False Information Crimes Involving Generative Artificial Intelligence. Journal of Chongqing University (Social Sciences Edition), 2025, 31(06): 183-197.

[10] Huang Junya. Research on the Dilemma and Path of Criminal Law Regulation of Generative Artificial Intelligence. Legal Vision, 2025, (30): 40-42.

[11] Pang Liangcheng. Research on Criminal Regulation of False Information Generated by Generative Artificial Intelligence. Social Sciences in Guangdong, 1-13[2026-04-01]. https://link.cnki.net/urlid/44.1067.C.20260308.1454.032.

