ISSN: 2688-8653

The Criminal Risks of Generative AI Content Dissemination: A Case Study of the First Domestic Case of ChatGPT Fabricating False Information

Peichen Huang*

School of Media, Nanchang Institute of Technology, Nanchang, 330044, China

*Corresponding author's e-mail: siuwhkkx@outlook.com

Abstract

Currently, generative AI is deeply integrated into the field of information dissemination. While its ability to generate automated content enhances efficiency, it also gives rise to legal and criminal risks such as the spread of false information. The first domestic case of ChatGPT fabricating false information has further pushed such risks to the forefront of judicial practice. However, the current legal framework is not adapted to the characteristics of AI technology. There are gaps in the determination of responsibility for AI false information and the application of legal provisions, which leads to ambiguous rights and responsibilities in practice and compliance difficulties in case handling. This article sorts out the core facts and disputes of the case, analyzes the difficulties in the application of the law through qualitative analysis, and finds that the current supervision lags behind in regulating AI-generated content and lacks targeted provisions. It proposes that legislative research should be strengthened and the integration of law and AI should be promoted to achieve collaborative governance. This study offers a new perspective on the legal liability determination of generative AI, fills the theoretical gap in judicial practice, builds a complete regulatory framework, and also provides empirical references and practical guidance for policymakers to improve AI supervision.

Keywords

Generative AI; The Spread of False Information; Risk of Criminal Involvement; The ChatGPT False Information Case; Application of Law.

1. Introduction

With the rapid iteration of generative AI technology, it has been deeply embedded in diverse information fields such as news dissemination, social interaction, and knowledge production, and is reshaping the information dissemination landscape with its automated and efficient content generation capabilities[1]. However, behind the technological dividends, the "black box nature" of generative AI and the autonomy of content generation have made it a new type of carrier for the spread of false information - such false information not only spreads rapidly and has a wide coverage, but also is more likely to cause public cognitive biases, social trust crises, and even give rise to criminal risks such as provoking trouble and disrupting public order[2,3].

The emergence of the first case in China where ChatGPT fabricated false information has pushed the criminal risks associated with generative AI from theoretical exploration to the forefront of judicial practice: In this case, the false information generated by AI spread through the Internet, causing a bad social impact. However, the judicial authorities found themselves in a predicament in determining responsibility and applying legal provisions, exposing the insufficient compatibility between the legal framework of the industry and the characteristics of AI technology[4,5].

ISSN: 2688-8653

Although existing research has focused on the false information risk and legal regulation of generative AI, there are still two core limitations: Firstly, most studies focus on macro risk analysis or generalized regulatory suggestions, lacking in-depth analysis of specific cases, making it difficult to address the practical challenges of "guilt and innocence" and "this crime and that crime" in judicial practice. Secondly, the definition of the responsible subjects for AI-related criminal risks and the analysis of the adaptability of legal provisions are not detailed enough, and a systematic solution for the coordinated governance of "technology - law - regulation" has not yet been formed[6,7].

Based on this, this study takes the first domestic case of ChatGPT fabricating false information as the core entry point. By sorting out the facts of the case and analyzing the disputes over the application of law, it clarifies the core manifestations and causes of criminal risks related to generative AI, and then puts forward suggestions for legal improvement and regulatory optimization. The theoretical significance of the research lies in filling the gap in case studies of criminal risks related to generative AI and providing a new perspective for the determination of AI legal liability. The practical significance lies in providing references for judicial authorities in handling similar cases and empirical evidence for policymakers to improve the AI regulatory system.

2. Literature Review

Existing research on generative AI and false information mainly focuses on three dimensions: "risk identification - legal regulation - criminal-related application", laying a foundation for this study, but there are also obvious areas that need to be supplemented.

2.1. Research on Risk Identification of Generative AI False Information

The academic community generally believes that the technical characteristics of generative AI make its risk of false information "special" and "complex". The "realism" and "low threshold" of AI-generated content make it easier for false information to break through the traditional information review mechanism, and it is difficult to trace the source after dissemination[1]. The "autonomous learning ability" of AI may cause it to generate biased content under the influence of training data, which in turn turns into false information and exacerbates social cognitive fragmentation[5]. Such research clearly defines the risk manifestations of false information in generative AI, but its shortcomings lie in: mostly focusing on macro risk characteristics, without analyzing how risks are transformed into "criminal-related behaviors" in combination with specific judicial cases, and lacking practical discussions on the boundary between risks and crimes.

2.2. Research on the Legal Regulation Path of Generative AI False Information

In response to the regulation of AI false information, scholars have proposed multi-dimensional solutions such as "technical regulation", "legal improvement", and "industry self-discipline". From the perspective of administrative supervision, it is suggested to establish a "filing system" for AI-generated content, requiring developers to conduct pre-approval for AI-generated content[7]. From a legal perspective, it is necessary to expand the scope of application of the existing information supervision laws and include AI developers and platform providers in the category of responsible entities[4]. Moreover, the cross-border characteristics of ChatGPT-like AI still require enhanced international regulatory collaboration[2]. These studies provide directions for regulatory paths, but the limitation lies in that most of the suggestions are inclined towards a "macro framework" and do not design specific provisions for "criminal risks", such as the standards for responsibility division and the scale of criminal penalty application after AI is involved in crimes, which are difficult to directly guide judicial practice.

ISSN: 2688-8653

2.3. Research on the Criminal Risks and Application of Charges of Generative AI False Information

In the field of criminal risks, research mainly focuses on the application of the crime of "fabricating and intentionally spreading false information" and the crime of "provoking trouble". By comparing the constituent elements of the two crimes, it is pointed out that the application of the crime of "fabricating and intentionally spreading false information" should be limited to the field of "disrupting public order", and the element of "subjective intent" must be met[6]. After further analysis of the criteria for distinguishing the two crimes in cases of online false information, it was found that "whether the information content is related to public interests" and "whether it causes actual social order chaos" are the key points[8]. Some scholars have specifically examined the current judicial application status of the crime of "fabricating and intentionally spreading false information", pointing out that there are problems such as ambiguous boundaries of the application of criminal charges and inconsistent sentencing standards in practice, and proposing to optimize the path from the perspectives of interpreting the constituent elements and improving the case guidance system[9].However, the deficiency of the existing research lies in the fact that it has not combined the technical characteristics of generative AI, such as the "non-subjective intent" and "multi-subject participation" of the information generated by AI, which makes it difficult to directly apply the constituent elements of traditional criminal charges and has not formed a unified view on the responsible subjects of AI-related crimes.

In conclusion, the existing research has covered the core areas of false information in generative AI, but there are still gaps in "in-depth case analysis", "detailed criminal responsibility", and "practical adaptation of legal provisions". This article takes the first domestic case of ChatGPT fabricating false information as a breakthrough point, precisely to make up for these deficiencies and achieve the connection between "case - theory - regulation".

3. Method

3.1. Research Ideas

This study takes "case analysis - problem focus - countermeasure proposal" as its core logic. It first clarifies the specific manifestations of criminal risks associated with generative AI by sorting out the facts and controversies of the first domestic case of ChatGPT fabricating false information. Combining legal norms with existing research, analyze the legal application difficulties and regulatory deficiencies exposed in the case; Finally, based on the problem analysis, specific paths for the coordination of legal improvement and supervision are proposed, forming a closed-loop research of "practice - theory - practice".

3.2. Research Contents

3.2.1. Review of the Core Facts and Controversies of the First Domestic Case of ChatGPT fabricating False Information

By publicly disclosing judicial documents, media reports and other materials, key information of the case is extracted, including: the content and dissemination path of AI-generated false information, the subjects involved in the case, and the focus of judicial authorities' trial, such as the determination of responsible subjects, the selection of charges, disputes over criminal penalty discretion, etc., laying a factual foundation for subsequent legal analysis.

3.2.2. Analysis of Legal Application Challenges in Criminal Risks Associated with Generative AI

Based on current legal norms such as the Criminal Law and the Cybersecurity Law, a qualitative analysis is conducted on the disputed points of the case: The first is the difficulty in defining the

ISSN: 2688-8653

responsible subject, that is, whether AI developers have the obligation to review the content generated by AI, whether users have the subjective intention of "intentionally using AI to spread false information", and whether the platform needs to bear the responsibility of information filtering. The second issue is the difficulty in applying the crime name, that is, whether the case meets the elements of "subjective intent" and "disrupting public order" of the crime of "fabricating and intentionally spreading false information", and whether there is a conflict in application with the crime of "provoking trouble". The third challenge is the adaptation of legal provisions, that is, whether the current laws cover the regulation of AI-generated content and whether there are any blank provisions.

3.2.3. Regulatory Deficiencies and Improvement Path Construction for Criminal Risks Associated with Generative AI

Based on case analysis and existing research, this paper summarizes the core deficiencies in the current supervision of generative AI, and then proposes improvement paths from the legislative, judicial, and administrative supervision levels: at the legislative level, it is necessary to supplement the responsibility provisions for AI-generated content; at the judicial level, it is necessary to clarify the judgment standards for AI-related criminal cases; at the administrative supervision level, it is necessary to establish a full-process traceability mechanism for AI content.

3.3. Research Methods

3.3.1. Case Analysis Method

Taking the first domestic case of ChatGPT fabricating false information as the core case, by dissecting the facts of the case, the focus of the dispute and the trial process, the specific forms of criminal risks related to generative AI are presented intuitively, providing a practical sample for the analysis of legal application and ensuring that the research conclusion is in line with judicial reality.

3.3.2. Literature Research Method

Systematically sort out the academic literature at home and abroad on the risk of false information in generative AI and legal regulations, as well as the current legal norms such as the Criminal Law and the Cybersecurity Law, absorb the reasonable viewpoints of existing research, clarify the research starting point and the problems to be solved, and ensure the theoretical rigor and legal compliance of the research.

3.3.3. Qualitative Analysis Method

For the legal application difficulties in cases, such as the definition of the responsible subject and the selection of charges, qualitative analysis is carried out in combination with legal provisions and legal theories. The compatibility conflicts between current laws and the characteristics of AI technology are dissected, and then the improvement direction of regulatory provisions is derived. Subjective comments are avoided, and only objective analysis based on facts and laws is conducted.

4. Result

This study takes the first domestic case of ChatGPT fabricating false information as the core analysis sample. Through the research approach of case dissection, interpretation of legal norms, and cross-verification of literature, it systematically refines three core practical results for the governance of criminal risks related to the dissemination of generative AI content.

ISSN: 2688-8653

4.1. The Division of Rights and Responsibilities among Multiple Subjects Lacks a Basis, and the Determination of Responsibilities Has Fallen into Practical Difficulties

The research, through the review of the focus of case hearings, found that in the scenario of false information dissemination by generative AI, the boundaries of rights and responsibilities of the responsible subjects show obvious ambiguity: There are no clear standards for the content review obligations of AI developers, and it is difficult to prove the subjective element of "intentional dissemination" by users. The distinction between the information filtering responsibility of online platforms and the traditional information review responsibility lacks legal basis, which directly leads to a deadlock in the determination of the objects of accountability by judicial authorities[5]. Existing research has confirmed that the core crux of the legal risks of generative AI lies in the unclear division of responsibilities among multiple parties. The "autonomous generation feature" of AI precisely breaks the linear responsibility chain of "producer - disseminator - platform" in traditional information dissemination, rendering the original logic of rights and responsibilities division ineffective[5]. Although some studies have proposed that AI developers and platform providers should be included in the category of information supervision responsibility subjects, the boundaries of the "reasonable review obligations" of developers in AI-generated scenarios have not been clearly defined, nor have the intensity of the "filtering obligations" of platforms for AI-generated content been defined. This has led to the relevant theoretical suggestions being difficult to respond to specific disputes in judicial practice. Meanwhile, the issue of the black box nature of AI technology exacerbating the difficulty of liability traceability has been fully confirmed in this case. The platform evaded responsibility by claiming that "AI-generated content is unpredictable", further highlighting the gap in the current legal definition of the rights and responsibilities of multiple subjects, and also confirming the practical pertinence of the research results.

4.2. The Elements of Traditional Criminal Charges Conflict with the Characteristics of AI, and There are Compatibility Obstacles in the Application of Criminal Charges

The research, through the analysis of the application of law in the case, confirmed that the constituent element system of the current Criminal Law, which includes the crime of "fabricating and intentionally spreading false information" and the crime of "provoking trouble", is difficult to adapt to the technical characteristics of generative AI. Specifically, the establishment of the crime of "fabricating and intentionally spreading false information" requires the simultaneous satisfaction of the dual elements of "subjective intent" and "disrupting public order"[6]. However, in this case, the act of AI independently generating false information does not involve the "fabricating behavior" driven by "human subjective will" as stipulated in the traditional elements of criminal composition. The uncertainty of its dissemination range and influence is also difficult to meet the quantitative determination standard of "actual social disorder"[8]. What is more worthy of attention is that this charge has already faced the problems of "ambiguous application boundaries and inconsistent sentencing standards" in traditional cases of false information , and the characteristics of generative AI such as "non-human subjective drive" and "content generation autonomy" further magnify this legislative defect[9]. This has put judicial authorities in a dilemma when choosing between this crime and that crime - if the crime of fabricating and intentionally spreading false information is applied, the subjective elements and the subject of the act will be difficult to determine. If the crime of provoking trouble is applied, it conflicts with the definition of this crime that "the core field must be 'disrupting social order'"[6]. This contradiction also confirms the core finding of this study regarding the "obstacle in matching traditional criminal charges".

ISSN: 2688-8653

4.3. The Regulatory Framework Lacks a Full-process Mechanism, and Risk Prevention and Control Show a Significant Lag

Research findings reveal that the current regulatory system for the dissemination of generative AI content still adheres to the traditional model of "post-event accountability", lacking a full-process prevention and control mechanism tailored to the technical characteristics of AI, namely "pre-event review - in-event traceability - post-event assessment"[3]. Existing research has pointed out that the core of compliance risks for generative AI like ChatGPT-like lies in the lack of pre-control measures[3]. This judgment is fully confirmed in this case - the regulatory authorities only initiated intervention procedures after the false information generated by the AI had spread through the Internet and caused adverse social impacts, failing to achieve source control and process intervention of risks. Although some studies advocate for the establishment of a collaborative governance framework of "technology - law - regulation" , in current regulatory practices, administrative regulatory authorities have not yet established a filing system and technical traceability standards for AI-generated content, and industry self-discipline also lacks unified AI content review norms, making it difficult for regulatory measures to keep pace with the iteration speed of AI technology[1]. Furthermore, the rapid dissemination nature of AI-generated content has exacerbated the difficulty of supervision in this case. False information spread across multiple social platforms within 24 hours, and the traditional supervision model of "manual review + complaint handling" is difficult to cope with the challenge of such technology-driven information dissemination. This further confirms the practical judgment of this study regarding the "lag in regulatory regulation".

5. Research Conclusion and Future Prospects

5.1. Core Conclusion

This study provides an in-depth analysis of the first domestic case involving ChatGPT's fabrication of false information. It elucidates that the primary issues surrounding criminal risk governance in the dissemination of generative AI content centre on three key aspects. Firstly, there is an absence of a clear legal framework delineating the responsibilities of various parties. The rights and obligations of AI developers, users, and platform providers remain ambiguous, which complicates the accountability process for judicial authorities. Secondly, a fundamental conflict exists between the elements of traditional criminal charges and the intrinsic characteristics of AI, which are "self-generated" and "not driven by human subjectivity." The existing legal frameworks struggle to accommodate the behavioural innovations introduced by technological advancements, resulting in significant challenges in applying criminal charges. Thirdly, the regulatory system predominantly emphasises "post-event accountability" and lacks a comprehensive prevention and control mechanism that aligns with the unique characteristics of AI technology. This shortcoming hampers efforts to achieve proactive governance and effective risk mitigation at the source.

The crux of the aforementioned issues resides in the current legal and regulatory framework, which positions "human behaviour" as the primary object of regulation. Consequently, its rule design and accountability logic are predominantly centred around traditional, human-centric behavioural patterns. However, the "black box characteristics," "multi-subject participation," and "rapid dissemination" associated with generative AI fundamentally undermine the core assumptions of conventional regulatory scenarios. This results in a disjunction between the legal and regulatory framework and the practical application of technology.Simultaneously, the swift advancement of technology, coupled with the delay in legislative updates and the enhancement of regulatory frameworks, engenders a temporal contradiction that exacerbates the challenges associated with the governance of criminal-related risks. This study, by conducting a thorough analysis of specific cases, translates abstract technical risks into tangible

ISSN: 2688-8653

issues within judicial practice, thereby addressing a gap in the current literature that "emphasises macro risk categorisation while overlooking the practical analysis of individual cases." It offers robust empirical evidence to support the judicial assessment and regulation of criminal risks associated with generative AI.

5.2. Future Research Prospects

Based on the core findings of this study and the shortcomings of existing research, future exploration in related fields can be approached from three perspectives. Firstly, it is essential to refine the standards for the division of multi-subject responsibilities in criminal scenarios involving generative AI. This refinement should include the construction of a gradient responsibility system that aligns with the "degree of autonomy" of AI technology. It is necessary to delineate the review obligations of developers, the duty of care of users, and the filtering obligations of platform providers at various levels, thereby enhancing the practicality and adaptability of the standards for determining responsibility. Secondly, given the significant cross-border dissemination characteristics of ChatGPT-like AI, attention must be directed towards the governance challenges associated with cross-border cases of false information. This entails exploring judicial collaboration mechanisms and jurisdictional division rules that are compatible with the technical features of such AI, addressing the practical needs for international regulatory coordination, and resolving the conflicts of rights and responsibilities inherent in the governance of cross-border AI-generated misinformation.Finally, it is essential to advance research on the integration of technical means and legal regulations. By utilising technical tools such as AI content traceability, a collaborative supervision model of "technical filing + legal accountability" should be explored. This model must clarify the standards for traceability, data retention requirements, and judicial determination rules pertaining to AI-generated content, thereby facilitating the mutual reinforcement of technical prevention and control alongside legal regulation.

Future research may further refine the practical implications of research findings through cross-disciplinary collaboration among law, computer science, public administration, and other fields. This collaboration can enhance both the feasibility and systematic nature of the proposed plans, thereby offering more comprehensive theoretical support and practical guidance for the development of a legal-regulatory system tailored to the characteristics of generative AI technology.

References

[1] Zhang S.H., Li K. (2024) Research on the Risk and Governance of False Information in Generative Artificial Intelligence. Academic Exploration, (07): 129–140.

[2] Tang Y., Wang Y.T., Xu D.T. (2024) Legal Regulation of ChatGPT-like Artificial Intelligence from the Perspective of False Information Governance. Journal of Hubei University of Education, 41 (04): 34–39.

[3] Nie T. (2023) Legal Risks and Compliance of ChatGPT Generative AI. Internet World, (03): 29–33.

[4] Cheng L. (2023) Legal Regulation of Generative Artificial Intelligence: From the Perspective of ChatGPT. Political Science and Law Forum, (04): 69–80.

[5] He X. (2023) Legal Risks of Generative AI and Responses. Journal of Northwest University for Nationalities (Philosophy and Social Sciences Edition), (04): 89–98. DOI: https://doi.org/10.14084/j.cnki.cn62-1185/c.2023.04.008.

[6] Fu H., Fu L.R. (2023) On the Application Fields of the Crime of Fabricating and Intentionally Spreading False Information: From the Perspective of Distinguishing It from the Crime of "Network-based" Provoking Trouble. Journal of Sichuan Institute for Nationalities, 32 (03): 72–78, 97. DOI: https://doi.org/10.13934/j.cnki.cn51-1729/g4.2023.03.006.

ISSN: 2688-8653

[7] Yang J.W., Luo F.Y. (2024) The Operating Mechanism, Legal Risks and Regulatory Paths of ChatGPT-like Generative Artificial Intelligence. Administration and Law, (04): 101–115.

[8] Jia J. (2021) Research on the Application of Criminal Charges in Cases of Fabricating and Intentionally Spreading False Information Online: From the Perspective of Distinguishing between the Crime of Fabricating and Intentionally Spreading False Information and the Crime of Provoking Trouble. Theoretical Issue, (01): 146–153. DOI: https://doi.org/10.14180/j.cnki.1004-0544.2021.01.019.

[9] Zhang Y.G., An R. (2021) The Judicial Application of the Crime of Fabricating and Intentionally Spreading False Information: Current Situation Reflection and Path Optimization. Qilu Academic Journal, (03): 107–117.