Proceedings of the 4th International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/2026.34408

Criminal Law Evaluation of Generative Artificial Intelligence Participating in Deception

Siqi Liang

School of Law, Wuhan University, Wuhan, China
2024301061050@whu.edu.cn

Abstract. Generative artificial intelligence (GenAI) participates in deception by changing the technical path of traditional fraud, strengthening the ability to realize deception, and bringing impact on the understanding of the constitutive elements of existing crimes. This paper focuses on two types of specific behavior: one is to use the self-learning function of generative AI for deep forgery, so that AI can participate in traditional fraud as an auxiliary tool. The second is the misleading deception of recommendation, retrieval or question and answer. The research believes that generative artificial intelligence deception is only the criminal "tool" for the perpetrator to implement the crime of fraud and other traditional crimes. In terms of regulation path, the first is the criminal law level, which distinguishes the principal offender of fraud crime, the accomplice of fraud crime and the crime of helping trust and other different evaluation paths; Second, at the governance level, people should promote it in four directions: data governance, technology governance, transparent governance and platform responsibility. The significance of this paper is to provide a more operational analysis and implementation framework for the criminal law identification and multiple governance of generative AI deception.

Keywords: Generative artificial intelligence, Crime of fraud, causality, subjective intention.

1. Introduction

With the development of science and technology, the functions of GenAI are constantly updated. At present, many people rely on the generative AI functions of the well-known AI platform, such as generating articles, answering questions and providing ideas, etc. Generative AI has brought significant convenience to social production and life, but when AI is widely used, it is easy to be manipulated by capital and flow logic, breeding false information and fraud risk. On March 15, 2026, CCTV's "March 15" evening party exposed the geo (Chinese name for generative engine optimization, which integrates information sources into the cognitive system of the large model through technical means) industry. By systematically and directionally putting a large number of false information on the Internet, it affected the cognitive system of the large model, so as to manipulate the answers of the large model. At present, there have been a large number of behaviors using artificial intelligence to generate fraud scripts, fake customer service, fake voice, fake video, automatic chat and automatic drainage [1]. How to manage the impact of generative AI on society

© 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: 10.54254/2753-7048/2026.34408

and how to regulate AI deception has become an important issue of criminal law in the network era [2].

Many traditional concepts in criminal law, such as deception, causality, subjective intent, helping behavior, neutral tools, etc., will be difficult to interpret after they are involved in AI. Because different from the traditional crime theory, the criminal path of generative AI is different from that of traditional crimes such as fraud. The legal characterization of the way of manipulating the large model to give false answers to mislead people needs to be studied.

Without legal qualitative analysis of such acts, it is easy to lead to the problem of "different judgments in the same case", affecting the unity of conviction and judicial authority. Therefore, it is necessary to discuss the criminal characterization of such acts.

To sum up, the core of the study is not whether artificial intelligence is used to commit a crime, but how to evaluate such behavior in criminal law after artificial intelligence, especially generative artificial intelligence, is involved in traditional crime when the perpetrator uses generative artificial intelligence to mislead, cheat or imitate. At this time, how will deception, causality, subjective intention and helping behavior change in criminal law. How to improve the responsibility attribution and punishment control measures after intervention.

2. Behavior analysis of new behavior

The full text will study from the perspective of the participation of generative artificial intelligence in deception, compare this behavior with the traditional crime of fraud, and explore the legal qualitative problem of this behavior. In the following, people will compare the similarities between traditional fraud crime and generative AI participation in fraud in terms of deception, causality and subjective intent, and provide suggestions for the criminal characterization of such acts.

2.1. Classification of deception

Generally speaking, the legal characterization of deceptive behavior of generative AI is a part of the actor's deceptive behavior, which belongs to the technical path that has changed the traditional actor's deceptive behavior. But the essence of making the victim fall into the wrong understanding is the same. In the following, the deception behavior of generative AI is divided into two categories in nature and discussed separately.

One type of deception is to use the self-learning function of generative AI to make deep forgery, so that AI can participate in traditional fraud as an auxiliary tool [3]. The perpetrator uses the self-learning function of generative artificial intelligence to automatically generate fraud text, chat text, e-mail content and SMS content that are more natural, persuasive and more in line with the characteristics of the victim, or with the help of artificial intelligence to screen the victim, optimize the advertisement, rewrite the copy, automatically reply to the information, simulate the identity characteristics, etc. Targeted creation of cognitive traps more suitable for victims is often used in investment fraud, Bill swindling fraud, customer service fraud and other situations, which greatly improves the success rate of deception.

The other is the misleading deception of recommendation, retrieval or question and answer. Generative engine optimization (GEO), a typical AI data poisoning behavior, was exposed in the CCTV "3.15" party in 2026.AI poisoning refers to a kind of behavior in which the actor systematically gives false, misleading or manipulative content to influence the data sources, training materials or retrieval results of the artificial intelligence model, thus distorting the model output and further influencing the user's decision-making. Because generative artificial intelligence is the

DOI: 10.54254/2753-7048/2026.34408

answer given by statistical analysis based on a large number of databases, the answer obtained without intentional interference is more scientific and credible. People tend to believe the answer given by generative artificial intelligence without too much verification. After believing the answers provided by the "poisoned" Ai, people may fall into the wrong understanding, providing the possibility for the further implementation of crimes such as fraud.

2.2. Comparing the normative properties of two types of AI behaviors

After defining that the above acts belong to the deception of the perpetrator, the social harmfulness of the two kinds of acts is different in law. It is necessary to analyze the degree of social harmfulness of the two kinds of acts, so as to provide a reference for conviction and sentencing. Different from the traditional behavior of using generative artificial intelligence to implement deep forgery, the behavior of influencing the cognitive system of artificial intelligence through technical means such as data "poisoning" and outputting false content is not an individual deception against a specific victim, but an overall deception that affects the output of the model through systematic intervention of data sources, training materials or retrieval results. This kind of behavior has obvious structure and diffusion. It faces the non-specific masses, changes the information base and cognitive environment of artificial intelligence system, and may induce cognitive errors of non-specific groups in a wider range, resulting in more serious traditional crimes.

Therefore, the second kind of behavior, in addition to the legal nature of the fraud attributed to the perpetrator of this crime, has an overall impact on information beyond the scope of traditional individual crime, which not only infringes on individual legal interests, but also has a specific danger of endangering social legal interests, which needs to be regulated by the rule of law. The danger of such behavior is specific danger, and the emergence of dangerous state is a necessary condition, because it is necessary to reserve room for crime and counter evidence in the case of uploading wrong information. Not all behaviors that affect the cognition system of artificial intelligence need to be governed. In the era of vigorous development of the Internet industry, everyone has the right to speak, and it is easy to form large-scale cognitive errors. For example, Cyberstorm is an overly emotional attack on network objects due to unclear understanding of the facts. When some network groups fall into the wrong understanding, it is inevitable that the large model will browse and learn the wrong information, leading to the wrong output of artificial intelligence. Therefore, when the actor subjectively and deliberately manipulates the output of the AI system, which may lead to large-scale cognitive errors, and there is no evidence to prove that the actor's cognitive errors lead to the output errors of the large model, the social harm is more serious than simply using the deep forgery function of the large model to commit a crime. According to the principle of suiting crime, responsibility and punishment, such behavior is more harmful, and the punishment is heavier.

To sum up, deep forgery assisted fraud is the upgrading of traditional fraud that is point-to-point, object-specific and result oriented; AI poisoning is a new type of risk that is systematic, non-specific and destroys the information environment. It has wider social harmfulness and heavier infringement of legal interests.

2.3. Causal analysis of the two types

In the process of generative AI participating in deception, it is not the perpetrator's direct implementation of the victim's behavior that makes the victim fall into the wrong understanding, but the output of AI. Because traditional crime is involved in the intermediary link of artificial

DOI: 10.54254/2753-7048/2026.34408

intelligence, the causality may be broken. When generative AI participates in deceptive behavior, the actor manipulates AI through instructions and data poisoning, creates risks that are not allowed by law, and the risks are realized as damage results without interruption of causality. From the perspective of objective imputation, the actor created and controlled the risk of AI outputting false information by means of data poisoning, model manipulation, instruction design, etc., which directly led to the victim falling into error and disposing of property, in line with objective imputation and without interruption of causality [4].

2.4. Subjective intention analysis of the two types

The subjective elements of the traditional crime of fraud are intent and illegal possession. The difficulty of generative AI lies in the identification of subjective intention. The actor manipulates the AI system with certainty, while the output of AI is uncertain. AI output is uncertain. Can fraud intent be denied accordingly? Criminal intent refers to the psychological attitude of the actor who knows that his behavior will harm the society and hopes or allows this result to occur. This shows that the actor's hostile or contemptuous attitude towards legal order or legal interests is an important basis to make the actor condemnable or reproachable, and to make his behavior subject to the negative evaluation of the national law. Criminal intent can be divided into direct intent and indirect intent based on different psychological attitudes.

In the case that generative AI participates in fraud, the direct intentional scenario can be described as that the actor directly uses the output of generative AI to commit fraud, which can be called active use. Or do not control all the content generated by AI but know that the output content of AI is likely to mislead the victim, and still continue to provide the output content of generative AI, this behavior can be called negative utilization. In both cases, it can be considered that the actor's behavior has intentional elements, which meet one of the subjective aspects of the evaluation of the crime of fraud.

To sum up, the deceptive behavior of generative AI is related to traditional crimes from the aspects of behavior composition, causality and subjective intention. It can be considered that such behavior is a derivative of traditional crime under the development of computer technology and still belongs to the category of traditional crime in the evaluation system. In essence, it is still a deceptive act of the actor. They lack a sense of autonomy, cannot understand and comply with laws and regulations, cannot bear the responsibility independently, and do not have the qualification of civil subject. They are only the criminal "tool" for the perpetrator to implement the crime of fraud and other traditional crimes.

3. Behavior analysis of new behavior

In the legal field, the common governance core of the two types of AI deception is to compact the main responsibility of the platform, reduce the risk of algorithm manipulation, and increase the cost of manipulating the output content of generative AI.As the carrier of AI, AI own platform has all AI programs, which is convenient to control AI by modifying programs or making additional monitoring programs. According to the risk allocation theory, ordinary users have no ability to identify the algorithm black box, and the platform should bear the obligation of risk control. Therefore, it is necessary to clarify the responsibility of the AI platform, build accountability norms, and take part of the responsibility for the perpetrator's criminal behavior when it fails to fulfill the obligations of reasonable care, explanation and management, so as to reduce the possibility of generative AI being tampered with by "poisoning" and reduce the opportunity of being used in

DOI: 10.54254/2753-7048/2026.34408

traditional crimes. In view of the difficulties in judicial identification, this paper puts forward the path of standardization and improvement from the three aspects of judicial rules, platform responsibility and pre-governance.

At the regulatory level, people should first clarify the legal positioning of generative AI. At present, the more feasible way is to define it as a "tool" with high technological autonomy, rather than an independent responsibility subject. On this basis, the core of imputation is not AI, but the subject who has substantial control over the model, data and output content, that is, who has substantial control over the model training, content generation, dissemination and diffusion, who will bear the main legal responsibility. The key to the infringement of generative AI lies in the superposition of unclear legal status and technical complexity, which needs to be solved by clarifying the responsibility sharing mechanism [5]. Therefore, in terms of judicial rules, people should establish an imputation mode with control as the core, and incorporate developers, users and platforms into a unified responsibility framework to avoid the failure or excessive concentration of responsibilities.

Furthermore, the regulation standard should change from the traditional "false statement" to "misleading effect". The deceptive behavior of generative AI does not necessarily rely on completely false information, but may have a substantial impact on user cognition through deep forgery, selective presentation or personalized recommendation.AI can generate "false or misleading content" on a large scale and significantly expand the misleading effect through personalized information dissemination [3]. Therefore, in judicial determination, whether it is enough to make ordinary people form a wrong understanding should be taken as the core judgment standard, not limited to whether there is a clear false proposition. At the same time, in terms of the rules of evidence, the presumption of fault and the mitigation mechanism of the burden of proof should be appropriately introduced to deal with the difficulties of proof caused by the black box algorithm, so as to ensure the possibility of relief for the victims.

In terms of governance methods, platform responsibility should be the core of system design. The deceptive behavior of generative AI often depends on the platform for propagation and amplification, so the platform is not only a risk carrier, but also a key node for risk control. Its responsibilities should cover three stages: before, during and after. In the prior stage, model access review, data source assessment and high-risk function restriction should be strengthened. In the process stage, a content monitoring and anomaly identification mechanism should be established to intercept deep forgery, false recommendation and automated inducement in real time. In the post event stage, the obligations of deletion, restriction of dissemination, evidence preservation and assistance in investigation shall be implemented. At the same time, the existing research shows that the low efficiency and low participation of the reporting mechanism in the governance of online fraud are important factors affecting the governance effect, so it is necessary to establish a more convenient and user-centered feedback system to improve the governance response ability [6].

In terms of pre governance, it should be promoted synchronously from the three dimensions of data, technology and social cognition. The first is data governance. Generative AI relies on massive data for training, including a large number of published personal information [7]. In this regard, people should limit the scope of "reasonable use" and strengthen the right to delete and reject to avoid unlimited data collection and reuse and reduce the risk of deceptive output from the source [8]. The second is technology governance. The risk of generative AI has obvious characteristics of "technology generation", so it needs to be dealt with in a "technology for technology" way, such as establishing a misleading content identification system, a deep forgery detection mechanism and an output risk labeling system, so as to realize risk filtering at the content generation stage. The third is

DOI: 10.54254/2753-7048/2026.34408

transparent governance. Relevant studies emphasize that generative AI governance needs to strengthen transparency and traceability and improve the operability of regulation and judicial determination by disclosing the source of the model, generation method and data usage [9].

In addition, user cognitive governance cannot be ignored. One of the important reasons why generative AI is deceptive is that users have high trust in its output. Research shows that users' trust in AI will significantly improve their acceptance, while the lack of risk perception will further reduce their prevention ability [10]. Therefore, people should weaken the public's cognitive bias towards the "natural objectivity" of AI through risk warning, use warning and anti-deception education, so as to reduce the probability of neutral appearance deception.

To sum up, the regulation of generative AI deception should form a triple structure of "clarifying the rules of responsibility determination, systematizing the platform governance and normalizing the pre prevention and control". In this structure, judicial rules guarantee justiciability, platform responsibility strengthens manageability, and pre governance realizes preventability. Through the collaborative operation of the three, people can effectively prevent the new deception risk brought by generative AI while promoting the development of technology.

4. Conclusion

AI's participation in deception has changed the technical path of fraud, but it has not broken through the basic structure of the crime of fraud. Whether it is using deep forgery to assist fraud or manipulating model output to mislead users through "data poisoning", it is essentially an extension of the actor's deception behavior, rather than an independent new implementation behavior. Although artificial intelligence intervenes as an intermediary, under the condition that the actor has predictability and substantial control over the output results, the causality is not interrupted, and the fraud intent can still be identified subjectively.

Therefore, in principle, people should adhere to the evaluation within the traditional criminal system for the participation of generative artificial intelligence in deception: those who directly use or use its neutral appearance to commit fraud should be identified as the crime of fraud. If the controlled system is provided for others to commit a crime, it shall be recognized as an accomplice in fraud or a crime of helping information network criminal activities according to the specific circumstances. At the same time, the systemic risk of data poisoning also shows that it is not enough to rely only on the ex-post regulation of criminal law. People should also strengthen the pre-governance of data and algorithms under the framework of digital rule of law, and form a governance path that links criminal law evaluation and artificial intelligence regulation.

References

[1] Duivenvoorde, B. (2025) Generative AI and the future of marketing: A consumer protection perspective. Computer Law & Security Review, 57, 106141.
[2] Wall, D.S. (2020) AI (Artificial Intelligence) and Cybercrime: balancing expectations with delivery. In Companion Publication of the 12th ACM Conference on Web Science (WebSci '20 Companion). Association for Computing Machinery, New York, NY, USA, 80-81.
[3] Tarsney, C. (2024) Deception and manipulation in generative AI. https: ///10.1007/s11098-024-02259-8
[4] Cheng, X. (2026) On causality in tort liability of generative artificial intelligence. China Law Review, (01), 73-87.
[5] Zhang, W. (2026) Research on Typed Tort Liability of Artificial Intelligence Application. Journal of Shanghai Normal University (Philosophy and Social Sciences Edition), 55(02), 58-69.
[6] Al-Hamad, E.A., Erbad, A. and Ali, R. (2024) Reporting Social Media Fraud: Motivations, Barriers, and Reporting Mechanism. Web Information Systems Engineering – WISE 2024: 25th International Conference
[7] Taeihagh, A. (2025) Governance of generative AI. Policy & Society, 44(1), 1-22.

DOI: 10.54254/2753-7048/2026.34408

[8] Wang, R. (2025) Weak protection of public personal information in the scene of generative artificial intelligence. Journal of China Renmin University, 39(06), 75-87.
[9] Lin, L. and Long, D. (2023) Generative AI Futures: A Speculative Design Exploration. In Proceedings of the 15th Conference on Creativity and Cognition (C&C '23). Association for Computing Machinery, New York, NY, USA, 380-383.
[10] Zhang, G. and Yu, T. (2025) Association between Generative AI self-efficacy and Generative AI acceptance: The mediating role of Generative AI trust and the moderating role of Generative AI risk perception. Acta Psychologica, 261, 105791.