Judgment and Identification of Criminal Subjects in Generative Artificial Intelligence

Zhenghang Song *

School of Law, Hunan University, Changsha, 410000, China

* Corresponding Author Email: szhlyt@outlook.com

Abstract. In the era of artificial intelligence (AI) sweeping across the globe, AI-related crimes have sparked a debate over whether AI entities can be eligible subjects of criminal liability and how to determine the subject of the crime. Previous relevant studies have faced issues such as unclear standards for concept definition, ambiguous boundaries between objective attributes and normative evaluations, and complexity, which not only triggered disputes over standards and viewpoints but also easily led to the uncertainty of conclusions. Based on the elaboration of the essential attributes of artificial intelligence, this article will analyze the specific difficulties in identifying the subject of criminal responsibility, such as the difficulty in recognizing the subject and applying the constitutive elements of crime. It will seek specific identification and imputation paths for the subject of criminal responsibility of artificial intelligence, such as clarifying the scope of the responsible subject in judicial practice, constructing a hierarchical imputation framework, or adjusting the constitutive elements of crime to meet the requirements of technological development. Furthermore, it will draw on legislative experience from outside the region to improve some gaps in China's prevention and punishment of artificial intelligence crimes.

Keywords: Generative AI, criminal subject, criminal responsibility, free will.

1. Introduction

Generative artificial intelligence relies on algorithms, models and massive data training, which can independently generate original content such as text, images, audio and video. Its core is different from the "self-generation" ability of traditional AI, and it can imitate or even surpass human creativity and make independent decisions. Typical representatives include ChatGPT and DeepSeek. Its wide application has become a new criminal tool while improving efficiency and facilitating life, which has given birth to AI's illegal and criminal activities such as generating fake videos, voice fraud and copyright infringement.

China's "Interim Measures for the Management of Generative Artificial Intelligence Services" clarifies the regulatory obligations but does not stipulate relevant criminal responsibilities. Traditional criminal law takes natural persons and entities as criminal subjects, and the autonomy of generative AI breaks through the traditional standards of subject identification, which leads to disputes on subject qualification, developer, user and platform responsibility division.

Related crime risks are divided into four categories: tool type, object type, data type and autonomous type. The first three can be regulated by traditional criminal law, while autonomous crime brings new judicial challenges. At present, related cases are frequent in various places, and the judicial standards are different. Based on the theory and practice of criminal law, this paper systematically studies the subject identification of generative AI crime, which has both theoretical value and practical significance.

2. Core Controversies Regarding the Criminal Subject of Generative Artificial Intelligence and the Identification Types of Human Subjects

2.1. The Core Controversy Surrounding the Subject of Generative AI-Related Crimes

Whether generative artificial intelligence can become the subject of crime is the core dispute whether it has criminal responsibility ability and independent will in the sense of criminal law, and

the academic circles form three views: support, opposition and compromise. Supporters advocate giving AI a limited subject status and believe that advanced algorithms can form a sense of autonomy and take responsibility with reference to the legal person system [1,2]. However, this view lacks legal basis, AI has no real cognitive and self-control ability, and its behavior originates from preset procedures, so it cannot be punished independently, and it breaks through the traditional criminal law subject framework. The opposition view is mainstream, thinking that AI has no independent will and criminal intention, but is only a tool controlled by human beings, and the responsibility should be attributed to developers, users and other subjects [3]. The compromise point of view suggests that AI should not be identified as the main body for the time being, and strict liability should be introduced to expand the scope of human liability, but it is difficult to solve the problem of fault identification of AI's autonomous behavior and its practicality is insufficient.

2.2. A Typological Analysis of the Identification of Human Subjects in Crimes Involving Generative Artificial Intelligence

2.2.1. Developer

The developer of generative artificial intelligence is the technical source and creator of the system, who determines the operation of the system through algorithms, data, and models. When AI engages in harmful behavior, it should bear criminal responsibility according to fault.

Intentional crimes can be categorized into two types. The first type involves specifically developing and guiding the generation of harmful content for criminal purposes. For instance, in the Alien Chat case in Shanghai, the developer lifted restrictions through prompt words and output obscene content [4]. The second type involves knowingly allowing harm to occur for personal gain despite the risks involved. Intentionality can be corroborated through algorithm design, training data, internal documents, and other evidence.

Negligent crimes manifest as a failure to fulfill reasonable duty of care, including algorithmic oversights, lax data auditing, and insufficient risk testing before launch. Developers bear obligations such as data cleaning, security testing, and vulnerability remediation. Violating these obligations and causing harmful consequences constitutes a negligent crime. Developers who have dominance over AI and intentionally commit crimes are directly responsible as perpetrators, while those who are negligent bear the responsibility for negligent crimes.

2.2.2. User

Users of generative artificial intelligence refer to individuals or entities who input instructions, obtain, and utilize content generated through dissemination. As direct operational entities, their actions determine the usage of the content. Those who utilize AI to harm society should bear criminal responsibility and are also the most common direct perpetrators in related crimes [5]. Users who actively operate the system know that the instructions will generate harmful content and commit crimes based on it constitute direct intentional perpetrators. There are mainly two types of cases in judicial practice: one is the generation of false information to commit defamation and fraud, such as the case of a 4.3 million yuan scam in Inner Mongolia within 10 minutes using "AI face-changing"; the other is the generation of infringing content to commit copyright infringement and other crimes, such as Beijing's first AI copyright infringement case, where the defendant was convicted for profiting over 270,000 yuan by fine-tuning others' artworks through AI.

For ordinary users, if they accidentally generate illegal content by inputting neutral instructions but do not disseminate it, they are usually not held criminally liable; however, those who generate and disseminate illegal content in batches for profit should be severely punished, such as the "AI pornographic novel case" in Hubei.

In summary, if the user actively inputs illegal instructions and uses the generated content to commit a crime, they constitute a direct perpetrator; if neutral instructions accidentally generate illegal content and it is not disseminated, it does not constitute a crime.

2.2.3. Service provider

The service provider refers to the party providing generative artificial intelligence services, generally including generative artificial intelligence platforms, operators, custodians, and other entities, also known as providers. Regarding servers, they bear criminal responsibility based on their control and supervisory obligations. According to the "Interim Regulations on the Management of Generative Artificial Intelligence Services", providers shall fulfill the obligations of "optimizing training, maintaining information confidentiality, implementing warning restrictions, and conducting audit supervision". If a platform is aware that its model is being used for illegal or criminal activities but fails to take measures such as stopping services, deleting content, or restricting access, thereby allowing harmful consequences to occur, it may be considered as a crime of omission or complicity. When a company or enterprise, as a service provider, commits a crime with the will of the unit and the illegal proceeds belong to the unit, it constitutes a unit crime, and both the unit and the directly responsible personnel shall be held accountable. A typical case is the aforementioned "Shanghai AC pornography case".

3. The Judicial Dilemma of Identifying the Subject of Crimes Committed by Generative Artificial Intelligence

3.1. Difficulty in Subject Recognition

The generation behavior of artificial intelligence is characterized by anonymity, cross-regionality, and long chains. The black box of algorithms makes it difficult to trace the source of instructions and decision-making nodes, and it is challenging to distinguish the roles of developers, users, and operation platforms, especially in tracing users and distinguishing between developers and operation platforms. In AI tool-based crimes, such as the aforementioned "pornographic novel case" in Hubei Province, where users generate content and sell it on external networks, not only is it difficult to reach the scope of regulation, but it is also challenging to trace the source accurately. The case was accidentally discovered by the Mianyang police while investigating the sale of obscene comics and was reported to the Hubei police, who then initiated an investigation. The difficulty of regulation and tracing the source is evident. In AI system flaw-based crimes, if users use neutral instructions but generate illegal content, whether it is due to the platform's inadequate regulation or the developer's negligence in the model, there is often controversy when assigning responsibility.

3.2. It is Difficult to Prove the Application of the Constitutive Elements of a Crime

The autonomous generation characteristic of generative artificial intelligence weakens the direct domination of human will, making the criteria for determining intent and negligence tend to blur. In terms of intentional determination, it is difficult to judge whether the technology provider in instrumental crimes "knowingly" allows others to use AI to commit crimes, such as whether the developer of AI voice synthesis is aware that the tool may be used for fraud. For technological products with both legal and illegal purposes, the current judicial interpretation adopts a comprehensive judgment standard, taking into account factors such as the main purpose and safety precautions. However, due to the lack of specific rules, there are differences in judicial determination results [6]. In terms of negligence determination, crimes involving AI system defects face the dilemma of limited foresight. The complexity of technology makes it difficult for developers to fully predict issues such as algorithmic bias. Judicial practice often relies on industry standards for judgment, but industry norms are not yet well-established, making it difficult to provide objective evidence. In scenarios where AI autonomously engages in harmful behavior, if the technological evolution exceeds the expectations of developers, it is often difficult to identify negligence based on the "should have foreseen" requirement of negligence [7]. In addition, when users generate and disseminate illegal content through neutral instructions, they can argue that they have no subjective fault. In judicial practice, it is difficult to distinguish whether the content is generated due to

instruction guidance, spontaneous model generation, or data bias, further exacerbating the difficulty in determining subjective fault.

4. Improving the Response Path of Criminal Law for AI-Related Crimes

4.1. Clarifying the Scope of Responsible Entities and Establishing a Hierarchical Framework for Assigning Responsibility

The principle of "impact magnitude + subjective knowledge" can be established to delineate responsibilities. The user, who directly controls the behavior, bears the primary responsibility; the developer, who creates the risk, bears the fault liability; the service provider, who fails to fulfill its supervisory obligations, bears the secondary responsibility; if the provider knows that others are using the tool for criminal purposes but still provides it, they are identified as an accomplice; if the provider does not know but fails to take measures such as "risk warning" or "usage restrictions," they are identified as a negligent accomplice; the assistant bears supplementary liability and is only held accountable when their behavior plays a significant role in the commission of the crime. For example, in the Zhejiang AI crawler infringement case, the crawler user bears the primary responsibility, the developer bears the secondary responsibility if they knowingly provide the tool for infringement, and the provider bears supplementary liability if they provide a large amount of services.

4.2. Adjusting the Constitutive Elements of Crime to Align with Technological Development Requirements

It should be clarified that the subjects involved in AI-related crimes are both natural people and entities, encompassing developers, users, and providers. The boundaries of responsibility and the standards for imputation should be clarified to avoid different judgments for the same case.

It is suggested to introduce industry standards and risk foreseeability obligations into subjective determination. At the intentional level, if the provider fails to take general risk prevention measures in the industry, it can be presumed that they "should have known". At the negligence level, if a developer fails to meet industry technical standards (such as not testing in accordance with regulations) or fails to fulfill the obligation of risk assessment, it is presumed that negligence exists.

From a comparative perspective, some scholars advocate for addressing the accountability gap by pursuing criminal liability for AI providers and adjusting the negligence system, to prevent providers from using AI autonomy as a "shield of responsibility". This approach is of reference significance to China [8].

4.3. Drawing on Experience from Outside the Region

Italy is the first country in the European Union to regulate the use of artificial intelligence. In 2025, it enacted the "Artificial Intelligence Regulation and Government Authorization Act", which is consistent with the EU's "Artificial Intelligence Act" and aims to promote "people-oriented, transparent, and safe" AI applications [9]. The Act specifies that the illegal dissemination of AI-generated content causing harm to individuals is punishable by imprisonment for 1 to 5 years, and severe penalties are imposed for AI fraud and identity theft; transparency and human supervision must be strengthened in the workplace, healthcare, education, and other fields; the use of AI by children under 14 requires the consent of their guardians; AI-assisted creative works can be copyrighted, and text and data mining are limited to copyright-free content or authorized scientific research scenarios [9].

In March 2024, the European Union officially adopted the "Artificial Intelligence Act", adopting a risk-based regulatory approach. AI systems are categorized into four levels: unacceptable risk, high risk, limited risk, and minimal risk, with differentiated regulations implemented accordingly. Specifically, AI posing unacceptable risks is explicitly prohibited, while high-risk AI requires providers to fulfill strict compliance obligations [10].

Overall, China's AI legislation is converging with international trends, supporting innovation, valuing talent cultivation, and cracking down severely on AI crimes. However, there is still a gap in terms of refined supervision, such as rules regarding restrictions on the use of AI by minors, which have not yet been clarified. China can draw on international advanced legislative experience to improve its AI legal system and enhance the pertinence and completeness of supervision.

5. Conclusion

The autonomous characteristics of generative artificial intelligence pose new challenges to the identification of criminal subjects in traditional criminal law. Although it possesses the ability to generate and make decisions autonomously, it does not have the freedom of will, cognitive control ability, or the capacity to bear punishment in the sense of criminal law, and therefore should not be considered a subject of criminal responsibility. The attribution of responsibility for AI-related crimes should be traced back to human entities such as developers, users, and service providers, with liability being apportioned based on their respective behavioral roles and subjective faults. Facing the current issues in judicial practice, such as difficulties in tracing the subject and ambiguity in determining subjective fault, China's judicial community needs to improve the supply of rules. On the basis of adhering to the principle that natural persons and units are the core subjects of crime, it can refine the statutory duty of care for each party, introduce industry standards to judge subjective fault, and draw on legislative experience from overseas in risk classification supervision and severe crackdown on AI crimes to fill the gaps in criminal regulation. By clarifying the boundaries of responsibility, optimizing the application of constituent elements, and improving the regulatory system, effective punishment and prevention of generative AI crimes can be achieved, while taking into account technological innovation and the protective function of criminal law.

References

[1] Chen Xingliang. Criminal Law Response to Artificial Intelligence-Related Crimes. Journal of the National Procuratorate College, 2025, 33(02): 3-9.

[2] Peng Wenhua. Halliwell's Perspective on Artificial Intelligence Crime and Its Implications. Journal of Shanghai University of Political Science and Law (Law Review), 2019, 34(04): 70-88.

[3] Russkevich Evgeniy A., Dmitrenko Andrey P., Kadnikov Nikolay G., et al. Crisis and alingenesis (rebirth) of criminal law in the context of digitalization. Vestnik of Saint Petersburg University. Law, 2022, 3: 585-598.

[4] King Thomas C, Aggarwal Nikita, Taddeo Mariarosaria, et al. Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. Science and Engineering Ethics, 2020, 26(1): 89-120.

[5] Blauth T. F., Gstrein O. J., Zwitter A., et al. Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. IEEE Access, 2022, 10: 77110-77122.

[6] Alyssa Chia. Exploring the Response Path of Criminal Law to AI-Related Crimes. Interdisciplinary Science Express, 2026, 10(1): 28-36.

[7] Sun Daocui. Deconstruction of Knowledge and Criminal Law Response to AI-Related Crimes. Juvenile Delinquency Issues, 2023, (02): 4-19.

[8] Elina Nerantzi, Giovanni Sartor. 'Hard AI Crime': The Deterrence Turn. Oxford Journal of Legal Studies, 2024, 44(3): 673-701.

[9] Repubblica Italiana. Legge 23 settembre 2025, n. 132: Disposizioni e deleghe al governo in materia di intelligenza artificiale. Gazzetta Ufficiale della Repubblica Italiana, 2025.

[10] Melina Anastasopoulou. Exploring algorithmic governance: The AI Act and new realities for criminal justice, and fundamental rights. New Journal of European Criminal Law, 2025, 16(2): 176-196.