ISSN: 2710-0170

# Research on the Attribution and Liability Path of Artificial Intelligence Entities Infringing on Legal Interests

Ye Wang

School of Law, Anhui University of Finance & Economics, Bengbu Anhui 233030, China.

## Abstract

**The deep application of artificial intelligence (AI) technology has gradually permeated all aspects of social life, moving beyond a mere technological tool. The infringement of legal interests caused by AI entities is becoming increasingly prominent, posing a crucial issue for criminal law to address. This paper focuses on the attribution and liability for infringements of legal interests by AI entities. It defines the legal connotation of AI entities, distinguishes the legal attributes of weak and strong AI, categorizes the forms of infringement, and summarizes their core characteristics. Based on this, it analyzes the current challenges in criminal liability attribution, including disputes over the determination of the subject's qualifications, difficulties in determining causality, obstacles to the application of traditional attribution principles, and lagging criminal legislation. To address these issues, this paper proposes constructing a hierarchical and progressive system of responsible parties, clarifying the responsibilities of developers, users, and regulators; establishing a causal proof mechanism centered on algorithm tracing, and improving the rules for liability sharing in cases of multiple causes leading to a single effect; and constructing an internal governance path through optimized criminal legislation and improved judicial application, combined with the prioritization of civil liability and supplementary insurance and fund systems to build a multi-governance system. The aim is to provide theoretical reference and practical pathways for the criminal regulation of infringements of legal interests by AI entities, achieving a balance between the development of AI technology and the protection of criminal law.**

## Keywords

**Artificial intelligence entities that infringe upon legal interests, criminal liability, algorithm tracing, and multi-party collaborative governance.**

## 1. The basic theories of artificial intelligence entities infringing upon legal interests

The iterative upgrades of artificial intelligence (AI) technology have driven changes in production and lifestyles. However, while empowering social development, its increased complexity and autonomy have also led to a series of legal infringements. Research on this issue requires first clarifying the legal boundaries, types of legal infringements, and core characteristics of AI entities from a fundamental theoretical perspective, laying the foundation for subsequent research on attribution and liability[1].

### 1.1. Legal Definition of Artificial Intelligence Agents

Artificial intelligence (AI) agents and AI are related as technological entities and technological carriers. AI is a technological system that simulates, extends, and expands human intelligence, with its core being the simulation of human thought and behavior[2]. AI agents are the materialized carriers of AI technology, formed by the combination of hardware and algorithms, capable of independently performing certain actions[3]. Their scope ranges from small smart speakers to large-scale autonomous vehicles and industrial intelligent robots. Legally defining

ISSN: 2710-0170

AI agents requires focusing on their behavioral capabilities and social impact, namely, whether they can perform legally significant actions and whether they may infringe upon legally protected interests[4].

Based on their level of intelligence and autonomous decision-making ability, artificial intelligence (AI) can be divided into weak AI and strong AI, which is the core basis for defining the legal attributes of AI entities[5]. Weak AI, currently the mainstream form, is specialized and instrumental, capable only of performing pre-set tasks in specific domains. It lacks independent consciousness and subjective judgment; its behavior is essentially the result of human programming and instructions. Its legal attributes are those of an object or tool, dependent on a human subject[6]. Strong AI, on the other hand, possesses general intelligence and autonomy, capable of independent thinking and decision-making, having cross-domain learning capabilities, and even independent consciousness. Currently still in the theoretical research and development stage, if it becomes a reality, it will transcend the traditional scope of a tool, becoming a fictional subject with a special legal status. This classification also lays the foundation for determining the subject of liability for infringement of legal interests[7].

## 1.2. Typological Analysis of the Infringement of Legal Interests by Artificial Intelligence Entities

Based on the current state of technology and future trends, the infringement of legal interests by artificial intelligence entities can be divided into two categories: as tools of crime and as subjects of crime. The former is the main form in current judicial practice, while the latter is a forward-looking discussion in the era of strong artificial intelligence.

### 1.2.1. Infringement of Legal Interests When Used as Tools of Crime

The instrumental nature of weak artificial intelligence (AI) agents makes them a new type of tool for humans to commit crimes. Compared with traditional tools, they are characterized by high efficiency, strong concealment, and a wide range of impact, amplifying the harmful consequences of crimes. The essence of this type of harm is that humans use AI technology to commit crimes; the AI agent is merely a carrier, lacking independent criminal intent or negligence. Specific manifestations include using intelligent algorithms to commit telecommunications fraud and infringe on citizens' personal information, using industrial intelligent robots to disrupt production and business operations, and exploiting vulnerabilities in autonomous driving algorithms to commit intentional harm. In this type, the causal chain of the harm to legally protected interests clearly points to the underlying human subject, conforming to the logic of the elements of traditional crimes.

### 1.2.2. Infringement of Legal Interests by the Subject of Crime

This type targets strong AI agents whose actions that infringe upon legal interests are not based on preset human programs or instructions, but rather on independent actions based on autonomous judgment. They are outside of direct human control and become independent perpetrators of the infringement of legal interests. Examples include strong AI robots autonomously harming humans and intelligent systems autonomously committing financial fraud.

Currently, there is debate in academic circles regarding whether strong AI agents can be considered criminal subjects. Those who deny this argue that criminal subjects must possess criminal responsibility; strong AI agents lack human moral and legal awareness, cannot commit subjective culpability, and cannot bear the punitive or educational functions of criminal punishment. Those who affirm this view argue that the scope of legal subjects expands with social development, and legal persons and entities have already been included in the category of criminal subjects. If strong AI agents possess the independent capacity to infringe upon legal interests, they can be considered criminal subjects and appropriate penalties can be established

ISSN: 2710-0170

to protect these interests. This article argues that while the criminal subject status of strong AI agents is not a current practical issue, theoretical discussions and institutional pre-planning are necessary to address the legal challenges brought about by future technologies.

## 1.3. Summary of the characteristics of artificial intelligence entities infringing upon legal interests

Compared with traditional acts of infringing upon legal interests, the infringing behavior of artificial intelligence entities exhibits distinct characteristics due to their technological nature, which is also a significant contributing factor to the subsequent difficulties in attributing liability. These characteristics can be summarized in three points:

### 1.3.1. The Separability and Concealment of the Actor

In traditional cases of infringement of legal interests, the perpetrator and the actor are the same entity. However, in cases of infringement by AI entities, the direct perpetrator is the AI entity itself, while the responsible party may be multiple entities such as developers, users, and regulators. The degree of connection between each entity and the infringing act varies, making it difficult to accurately identify the responsible party. Furthermore, the behavior of AI entities is based on algorithms, which are technically sophisticated and covert. Human entities behind these algorithms can commit crimes through remote control and algorithm implantation, making it difficult to trace their identities and behavioral patterns using traditional investigative methods, further complicating the identification of the perpetrator.

### 1.3.2. Nonlinearity and Black Box Nature of Causal Processes

Traditional legal infringements exhibit a linear causal relationship with a clearly identifiable chain. However, the harm caused by AI agents is often the result of multiple interacting factors, including algorithmic vulnerabilities, improper operation, and external environmental interference. These factors are interconnected, making it difficult to distinguish a single cause of harm, resulting in a non-linear causal process. Furthermore, AI algorithms such as deep learning are inherently black-box; their operation and decision-making logic cannot be fully understood or explained by humans. Even developers cannot accurately trace the reasons behind specific decisions made by the algorithm. This technological black box blocks the path to factual attribution, making causal determination a significant challenge.

### 1.3.3. The Diffusion and Systemic Nature of Harmful Consequences

The networking and large-scale application of artificial intelligence (AI) technology allows its harmful activities to spread rapidly through the internet, overcoming the limitations of time and space and leading to widespread consequences. For example, a single algorithmic vulnerability could paralyze an entire network system, and a false AI recommendation could cause widespread financial losses for citizens. Furthermore, AI is widely used in critical sectors such as finance, transportation, and healthcare. Its harmful activities not only cause individual personal or property damage but can also trigger systemic risks. For instance, a malfunction in an autonomous driving system could lead to traffic chaos, vulnerabilities in a financial intelligent risk control system could cause market fluctuations, and even threaten public safety and national interests.

# 2. The Real Dilemma of Criminal Liability for Infringement of Legal Interests by Artificial Intelligence Entities

The core of criminal liability is to address the questions of "who should bear the responsibility, why should they bear the responsibility, and how should they bear the responsibility." The traditional criminal liability system is based on human subject behavior and is adapted to traditional forms of infringement of legal interests. However, the aforementioned

ISSN: 2710-0170

characteristics of artificial intelligence entities infringing upon legal interests create conflict with the traditional liability system, leading to multiple dilemmas in areas such as the identification of criminal liability subjects, the determination of causation, the application of liability principles, and criminal legislation.

## 2.1. Dilemmas in Identifying Subjects of Criminal Liability

The identification of the subject of criminal liability is a prerequisite for criminal liability. In traditional criminal law, the subjects of liability are natural persons and legal entities, both of which possess independent will and criminal capacity. However, in cases where artificial intelligence entities infringe upon legally protected interests, the separation of subjects and the controversy surrounding the legal attributes of artificial intelligence entities lead to difficulties in identifying the subject of liability.

### 2.1.1. The dispute over the subjectivity of artificial intelligence entities themselves

It is generally agreed that weak AI agents, lacking independent will and capacity for action, cannot be criminally liable, and responsibility should lie with the human entities behind them. However, the legal standing of strong AI agents is a focal point of theoretical debate, with the core disagreement revolving around the determination of criminal responsibility and the appropriateness of punishment. On one hand, the autonomous decision-making ability of strong AI agents is the result of algorithmic simulation, fundamentally different from the physiological and psychological abilities of humans to recognize and control themselves; therefore, they lack subjective culpability . On the other hand, traditional punishments, including capital punishment and imprisonment, are directed at humans and organizations, offering no real punitive effect on strong AI agents. If they were to be recognized as criminally liable, entirely new forms of punishment would be necessary, for which current theories and regulations are lacking, resulting in a lack of legal basis for determining criminal responsibility.

### 2.1.2. The Dispersion and Obscuration of Underlying Human Responsibility

Even if we deny the legal standing of AI agents and attribute responsibility to the human entities behind them, we still face the dilemma of diffused responsibility and a lack of accountability. The research, development, production, use, and regulation of AI agents constitute a complete chain involving multiple stakeholders. The occurrence of a harmful outcome may be related to the actions of multiple stakeholders, such as algorithmic vulnerabilities of the developers, improper operation by users, and the absence of regulators. These stakeholders may not share a common criminal intent, or even have any connection whatsoever, and may only have contributed to the harmful outcome through negligence, making it difficult to determine responsibility under the joint criminal liability system. Furthermore, the degree of correlation between the actions of each stakeholder and the harmful outcome is difficult to quantify, making it impossible to distinguish primary and secondary responsibilities, ultimately leading to a dilemma of "collective responsibility without any truly responsible party." For example, in the event of an accident involving an autonomous vehicle, the developers, users, and manufacturers may shift blame, making it impossible to accurately identify the responsible party.

## 2.2. The Dilemma of Causal Relationship Determination

Causation is the core element of criminal liability. Traditional criminal law theories of causation are based on the linear relationship between behavior and result, which is difficult to apply to the non-linear and black-box causal process of artificial intelligence infringing on legal interests, leading to a series of dilemmas in determination.

### 2.2.1. Technical black box obstructs fact attribution

Attribution of facts is a prerequisite for determining causation in criminal law, but the black box nature of algorithms makes it difficult to achieve. Deep learning algorithms learn

ISSN: 2710-0170

autonomously from massive amounts of data; their decision-making process involves complex data computation rather than pre-defined logical rules. Humans cannot accurately trace the reasons for their decisions or reconstruct the operational process. For example, if an erroneous approval by an intelligent risk control system leads to losses for a financial institution, the developers cannot explain the reasons for the algorithm's decisions or prove the causative factors. Furthermore, the technical investigative capabilities of judicial organs are not keeping pace with the development of artificial intelligence technology, creating technical barriers to the evidence collection and identification of complex algorithms, further exacerbating the difficulty of attribution of facts.

#### 2.2.2. Ambiguity in responsibility share due to multiple interactions

The harmful consequences of AI-mediated actions are often multi-cause-driven or multi-cause-driven, forming complex causal networks through the combined effects of multiple actors' actions, technological factors, and the external environment. Traditional theories of causality determination are applicable to situations where a single cause leads to a single result, but they cannot accurately distinguish the relative importance of each factor, leading to ambiguity in liability. For example, harm to humans by intelligent robots may be caused by multiple factors such as vulnerabilities in the development algorithm, user negligence, or tampering by third-party hackers. The proportion of each factor's influence is difficult to quantify, making it impossible to determine the criminal liability of the developer, user, and hacker. Some factors may also fall under the category of force majeure or accidents, requiring exclusion from the causal network, further increasing the difficulty of determination.

### 2.3. The Dilemma of Applying Traditional Principles of Liability

The principles of personal responsibility for crimes and the unity of subjective and objective elements in my country's criminal law are the basic principles of criminal liability. The acts of artificial intelligence entities that infringe upon legal interests break through the traditional forms of crime, making the application of the above principles face difficulties and also giving rise to theoretical debates on whether to include them in the principle of no-fault liability.

#### 2.3.1. The Dilemma of Applying the Principle of Individual Responsibility for Crime

The core of the principle of individual responsibility for wrongdoing is the identity of the act and the responsibility. However, in cases involving harm caused by AI entities, the AI entity that directly commits the harmful act cannot be held responsible. Multiple human entities are involved in the harmful act, but no single entity commits a complete crime, thus breaking the identity of the act and the responsibility. Strictly applying this principle might result in no one being held responsible due to the relatively small roles played by each entity, violating the goal of protecting legal interests. Conversely, requiring all related entities to be held responsible might implicate innocent people, contradicting the core meaning of individual responsibility for wrongdoing.

#### 2.3.2. The Dilemma of Applying the Principle of Unity of Subjectivity and Objectivity

The principle of unity of objective and subjective elements requires that the determination of criminal liability requires both objective criminal act and subjective culpability; neither can be lacking. However, in cases of harm caused by AI agents, the technical black box and multi-factor interactions present difficulties in determining subjective culpability: on the one hand, proving subjective culpability is challenging, as the human actors behind these acts often use technological means, making it difficult to prove their intent or negligence through traditional evidence, and often requiring inference through circumstantial evidence, resulting in uncertainty; on the other hand, in some cases, the human actor lacks clear subjective culpability, but their actions lead to harm through the AI agent. For example, the developer may have exercised reasonable care, but due to technological limitations and algorithmic vulnerabilities,

ISSN: 2710-0170

the harm occurred. In such cases, according to the principle of unity of objective and subjective elements, the developer is not liable, but the harmful result has already occurred, and the goal of protecting legal interests cannot be achieved.

#### 2.3.3. Disputes Regarding the Application of the No-Fault Liability Principle

In response to the dilemma of attribution of liability, some scholars have proposed introducing the principle of no-fault liability from the civil field into criminal liability. This means that regardless of whether a human subject is at fault, if an AI entity commits an infringing act, the relevant subject is liable. Opponents argue that the principle of no-fault liability violates the principles of restraint and unity of subjective and objective elements in criminal law. Criminal law punishes acts with subjective culpability, and introducing this principle would broaden the scope of criminal punishment, leading to its overgeneralization. This article argues that the principle of no-fault liability is essentially a principle of risk-sharing, more applicable to compensation for losses in civil torts. The core of criminal liability is punishment and condemnation, and it is not suitable for direct introduction, but it can be applied in civil liability, complementing criminal liability.

### 2.4. The Dilemma of Lagging Criminal Legislation

Although my country's current criminal law has been amended many times, it has all been aimed at traditional forms of crime and has not fully considered the infringement of legal interests brought about by artificial intelligence technology. As a result, the current legislation is lagging behind in dealing with this issue, which is specifically manifested in the insufficient coverage of the crime system, the inadequacy of the appropriateness of the penalty settings, and the existence of gaps in criminal regulation.

#### 2.4.1. The existing crime classification system is insufficiently comprehensive.

The current criminal law system, based on traditional crime patterns, struggles to cover new forms of infringement upon legal interests by artificial intelligence agents. Some acts of infringement cannot be covered by existing charges, such as acts of harm autonomously committed by strong AI agents, infringements of legal interests caused by algorithmic discrimination and manipulation, and systemic cybersecurity risks posed by AI agents. While some acts could be included in existing charges, there are issues with inaccurate characterization. For example, telecommunications fraud using AI algorithms differs in its social harm from traditional fraud, making it difficult to accurately assess its harm by directly classifying it as fraud.

#### 2.4.2. Insufficient Appropriateness of Penalties

The current penal system, including capital punishment, imprisonment, property penalties, and disqualification penalties, targets both humans and legal entities, and is ill-suited to the emerging forms of crime involving AI-generated entities that infringe upon legal interests. On the one hand, current sentencing for human perpetrators using AI technology fails to adequately consider technological factors, resulting in limited punitive effectiveness. On the other hand, if strong AI entities are recognized as responsible parties in the future, the current penal system lacks appropriate methods to achieve both punitive and preventative functions. Furthermore, for crimes committed by legal entities, the fines under the current dual-penalty system lack clear standards, making it difficult to align with the entity's illegal gains and social harm, resulting in ineffective punishment.

#### 2.4.3. Gaps in Criminal Regulation

Current criminal law focuses primarily on crimes committed during the use of AI agents, leaving gaps in regulations governing actions during the research, development, production, and regulatory phases. For example, actions such as developers intentionally creating algorithmic vulnerabilities that go unexploited, manufacturers producing AI agents with quality defects that

ISSN: 2710-0170

do not cause actual harm, and regulators failing to oversee the abuse of AI agents are all difficult to regulate under current criminal law. This leaves relevant parties without legal constraints and prevents infringements on legal interests from the outset.

## 3. Construction of Criminal Liability Path for Artificial Intelligence Entities Infringing on Legal Interests

To address the dilemma of criminal liability for infringements of legal interests by artificial intelligence entities, it is necessary to base our approach on the current state and future trends of technological development, and combine it with the basic principles of criminal law and the goals of protecting legal interests. We should construct a scientific and reasonable path for criminal liability from three core dimensions: identification of the responsible party, determination of causation, and improvement of the criminal governance system.

### 3.1. Construct a hierarchical and progressive system of responsible entities

To address the dilemma of fragmented and vaguely defined responsible parties, a single-entity liability model should be abandoned in favor of a hierarchical and progressive system of responsible parties. This system should be based on the intelligence level of the AI agent, the position, role, and duty of care of each entity within the industry chain, defining responsibility levels and clarifying the standards and scope for liability determination at each level to achieve precise identification of responsible parties. The core responsible parties are divided into three levels: developers, users, and regulators. Responsibility at each level is progressive; that is, if a higher-level entity cannot be identified or has no responsibility, the next higher-level entity assumes the corresponding responsibility.

#### 3.1.1. Researcher Level

Developers are the source of AI agents, and their actions determine the security performance and behavioral boundaries of these agents. They bear a primary duty of care and responsibility, with liability determined according to the principle of fault-based liability. A developer's intent includes intentionally creating algorithmic vulnerabilities or implanting malicious programs; negligence includes failing to fulfill reasonable technical care, with algorithmic vulnerabilities becoming the primary cause of the harm. For weak AI agents, developers primarily bear liability for negligence; for strong AI agents, their liability is appropriately increased, including not only liability for negligence but also a supervisory duty. If a lack of supervision leads to harm, the developer is liable. If a developer fulfills reasonable care, but the harm is caused by technical limitations or external factors, they are not liable.

#### 3.1.2. User Level

The user is the actual controller of the AI entity and the core responsible party in infringement cases. They bear direct duty of care and responsibility, and the determination of responsibility follows the principle of unity of subjective and objective elements. If the user intentionally uses the AI entity to commit a crime, they constitute a direct perpetrator and bear full criminal responsibility. If the user's negligence in operation, maintenance, or supervision leads to the infringement, they bear negligent criminal responsibility. If the user has fulfilled their reasonable duty of care, and the infringement is caused by external factors such as development vulnerabilities or third-party interference, they are not liable.

#### 3.1.3. Regulatory Levels

Regulators, including administrative regulatory departments and industry associations, bear the public responsibility of standardizing the development of artificial intelligence technology and preventing infringement on legal interests. They should bear regulatory obligations and responsibilities, and the determination of liability follows the principle of dereliction of duty.

ISSN: 2710-0170

Criminal liability is only incurred when there is dereliction of duty such as abuse of power or negligence, and the act has a causal relationship with the infringement under criminal law. The responsibilities of regulators include failing to formulate technical and safety standards, failing to conduct effective supervision and inspection, and failing to address risks in a timely manner. Their liability is supplementary, meaning that they bear corresponding dereliction of duty liability when the aforementioned developers and users have no liability or their liability is insufficient to compensate for the infringement.

## 3.2. Clarify the causal relationship standards for determining the infringement of legal interests by artificial intelligence entities

To address the challenges in determining causal relationships, we should break through the traditional linear determination theory, combine the characteristics of artificial intelligence technology, establish a causal relationship proof mechanism with algorithmic tracing as its core, improve the responsibility sharing rules for multiple causes and one effect, and achieve accurate determination of causal relationships.

### 3.2.1. Establish a causal relationship proof mechanism with "algorithm tracing" as its core.

Algorithms are the core of artificial intelligence agents, and the causal attribution of harmful acts is essentially algorithm attribution—that is, reconstructing the algorithm's operation process through technical means to determine the reasons and influencing factors behind the decisions. This mechanism requires simultaneous efforts from both technical and legal perspectives: Technically, it necessitates the development of algorithm attribution standards and norms, requiring developers to establish "algorithm logs" to ensure full traceability of the operation; developing algorithm identification technologies and tools, and establishing third-party identification institutions to provide technical support for factual attribution. Legally, it requires clarifying the obligations of developers and users to preserve and provide algorithm logs, with those who refuse to comply bearing corresponding responsibilities; including algorithm identification opinions as a legally recognized type of evidence, clarifying their evidentiary value; and implementing a reverse burden of proof, requiring developers and users to prove that their actions and the harmful results are not causally related, with the inability to prove otherwise presumed to be causally related, thus reducing the burden of proof for judicial organs.

### 3.2.2. Constructing a responsibility-sharing rule for multiple causes and a single effect

To address the ambiguity in liability share caused by multiple causal interactions, a liability sharing rule for multiple causes leading to a single effect is constructed. Based on the magnitude of the effect and degree of fault of each factor, the criminal liability share of the responsible parties is determined, adhering to the principles of unity of subjective and objective elements and proportionality between crime and punishment. The process involves three steps: First, screening for causal relationships under criminal law, adopting the theory of adequate causation, and excluding non-criminal causal factors such as force majeure and accidents; second, determining the degree of fault of each party, classifying it into three levels based on the subjective form of culpability: intent, gross negligence, and ordinary negligence, and determining the standards for determination in conjunction with each party's duty of care; third, determining the liability share and the range of punishment, with greater effects and higher degrees of fault resulting in a larger liability share and heavier punishment. If a joint crime is constituted, the responsible parties are identified as principals and accomplices and bear responsibility accordingly.

ISSN: 2710-0170

## 3.3. Multiple Approaches to Criminal Governance of Crimes Infringing on Legal Interests by Artificial Intelligence Bodies

The infringement of legal interests by artificial intelligence entities is a comprehensive issue involving technology, law, and society. Criminal governance cannot rely solely on internal optimization of legislation and the judiciary; it should also combine civil and administrative means to build a governance system that integrates internal criminal governance with multi-party co-governance, thereby achieving full-chain regulation.

### 3.3.1. Internal Optimization of Criminal Proceedings

Optimizing the internal mechanisms of criminal justice is the core of governance. By improving criminal legislation and applying it precisely in criminal justice, we can compensate for legislative lags and achieve effective regulation.

(1) Improvement of criminal legislation: Adhering to the principles of restraint and foresight, firstly, supplement and improve the existing crime system, add new crimes such as algorithm manipulation and artificial intelligence entities endangering public safety, revise the constituent elements of existing crimes and clarify sentencing standards; secondly, optimize the setting of penalties, increase the application ratio of property penalties and qualification penalties to relevant human subjects, and reserve space for the criminal system for strong artificial intelligence entities; thirdly, improve the criminal regulation chain, extend the scope of regulation to the research and development and production stages, add crimes such as negligence in the design of artificial intelligence algorithms, and improve the identification standards for dereliction of duty crimes by regulators.

(2) Precise application of criminal justice: First, strengthen the technical skills of judicial personnel, carry out professional training, and establish a judicial technical expert database to provide professional support for case handling; Second, unify judicial judgment standards, with the Supreme People's Court promptly issuing guiding cases and judicial interpretations to clarify the standards for fact-finding, causal determination, liability determination, and sentencing; Third, pay attention to the individualized application of penalties, and achieve proportionality between crime and punishment based on factors such as the technical level, degree of fault, and social harm of the responsible party.

### 3.3.2. Construction of a Multi-Governance System

Criminal governance is the last line of defense, but it lacks preventative and compensatory functions. A multi-governance system should be built to leverage the compensatory function of civil liability, the regulatory function of administrative liability, and the punitive function of criminal liability to achieve multi-dimensional regulation.

(1) Prioritizing Civil Liability: In infringement cases, the relevant parties are first held liable for infringement through civil means to compensate the victims for their losses. Civil liability is based on the principle of no-fault liability, reducing the difficulty for victims to provide evidence; the product liability and network service provider liability systems are improved, including artificial intelligence entities within the product category, and clarifying the platform operator's security obligations and supplementary compensation responsibilities. Prioritizing civil liability does not negate criminal liability, but rather achieves diversified protection of legal interests.

(2) Supplementary Insurance and Fund Systems: In response to the uncertainty and high risk of infringement by artificial intelligence entities, a mandatory insurance system for artificial intelligence and a compensation fund system for infringement of legal interests should be established to achieve socialized risk sharing. Developers and users should be required to purchase artificial intelligence liability insurance, with insurance companies compensating within the scope of liability; a compensation fund should be established, with sources including

ISSN: 2710-0170

industry taxes, entity contributions, and social donations, to provide a safety net for situations where the responsible party is unable to compensate or the responsible party cannot be identified, and the fund management institution can exercise the right of recourse against the actual responsible party.

## 4. Conclusion

The development of artificial intelligence technology is an inevitable trend in the progress of human society. While empowering economic and social development, it also brings unprecedented challenges to the traditional criminal law system. The attribution and liability for infringement of legal interests by artificial intelligence entities have become core issues in the field of criminal law.

This paper concludes that the legal attributes of artificial intelligence agents are closely related to their level of intelligence; weak AI agents are tools, while strong AI agents may be fictional subjects. Their characteristics of infringing upon legal interests lead to difficulties in criminal liability attribution in areas such as subject identification, causality determination, liability principles, and criminal legislation. To address this dilemma, it is necessary to construct a hierarchical and progressive system of responsible parties, clarifying the hierarchical responsibilities of developers, users, and regulators; establish a causal relationship proof mechanism centered on algorithm tracing; improve the rules for liability sharing among multiple causes and single effects; and build a governance system that combines internal criminal governance with multi-party co-governance.

The development of artificial intelligence (AI) technology is endless, and the legal interests it raises will continue to evolve in new forms. The response of the criminal justice system must be dynamic and continuous. Future research should combine technological development trends to deeply explore the legal status and penal suitability of strong AI entities, and improve attribution and liability theories. It should also strengthen interdisciplinary research, integrating knowledge from criminal law, AI technology, sociology, and other disciplines to construct a scientific legal system for AI, ensuring the healthy development of AI technology within the framework of the rule of law and achieving a balance between technological development and the protection of legal interests.

## Acknowledgements

This work is supported by Innovation and Entrepreneurship Training Project for College Students of Anhui University of Finance and Economics in 2024, Project number: 202410378183.

## References

[1] Zhu Liangyu. The paradigm shift and liability boundary of criminal causation in crimes caused by artificial intelligence [J]. Journal of Henan University of Economics and Law, 2026, 41(02): 118-132.

[2] Xu Shengzhang. The identification of copyright infringement of artificial intelligence training data and its criminal regulation from the perspective of the intersection of criminal and civil law [C]// Beijing Criminology Research Association. Criminology Research (Vol. 4). School of Intellectual Property, Nanjing University of Science and Technology; 2025: 573-583.

[3] Chen Lu. Criminal liability allocation for damage caused by artificial intelligence under the theory of negligence and competition [J]. Journal of Zhengzhou University (Philosophy and Social Sciences Edition), 2025, 58(06):73-79+171.

ISSN: 2710-0170

[4] Li Chuan. The dilemma of criminal law regulation of personal information and the way out of attribution of responsibility under the background of generative artificial intelligence [J]. Journal of East China University of Political Science and Law, 2025, 28(06):22-34.

[5] Ma Yongqiang. Understanding the Perspective of Generative Artificial Intelligence Crimes and Criminal Law Responses [J]. Journal of the National Prosecutors College, 2025, 33(06):43-64.

[6] Zhang Yiran. Algorithmic infringement and criminal liability of personal information in the field of generative artificial intelligence —with "risk control" as the core [J]. Data Law, 2025, 8(01): 54-80.

[7] Wang Hongqiu. A Study on the Criminal Liability Subject Qualification of Strong Artificial Intelligence Entities [D]. Lanzhou University, 2025.