International Journal of Law
www.lawjournals.org
ISSN: 2455-2194
Received: 06-04-2026, Accepted: 07-05-2026, Published: 22-05-2026
Volume 12, Issue 2, 2026, Page No. 288-294

# Artificial intelligence and criminal liability: Challenges to contemporary criminal law

**Vu Hai Anh**
Faculty of Law, Banking Academy of Vietnam, Dong Da District, Hanoi, Vietnam

**Abstract**
The rapid advancement of artificial intelligence is bringing about profound transformations in socio-economic life while simultaneously posing significant challenges to contemporary criminal law. As AI systems increasingly acquire capabilities of self-learning, autonomous decision-making, and a high degree of operational autonomy, the determination of criminal liability in cases where artificial intelligence causes harm has become a complex and highly contested legal issue. This article examines the difficulties associated with applying traditional doctrines of criminal liability to artificial intelligence, with particular emphasis on the elements of fault, the identification of the responsible legal subject, and the causal nexus between conduct and harmful consequence. Employing doctrinal legal analysis, comparative law methodology, and case study approaches, the study evaluates existing models of criminal liability related to artificial intelligence within contemporary international legal frameworks. The findings suggest that artificial intelligence does not, at present, satisfy the criteria necessary to be recognized as a subject of criminal liability in the traditional legal sense. Instead, the article argues for the development of a hybrid liability framework, encompassing individuals as well as organizations involved in the design, development, deployment, and operation of artificial intelligence systems.

**Keywords:** Artificial intelligence, criminal liability, criminal law, autonomous vehicles

## Introduction

In recent years, the rapid development of artificial intelligence (AI) has brought about profound transformations across almost all domains of socio-economic life. With capabilities for big data processing, self-learning, autonomous decision-making, and an increasing degree of independence from direct human control, AI has become one of the core technologies of the Fourth Industrial Revolution. No longer confined to a purely technical support role, AI systems are now widely deployed in areas that have a direct impact on human life, health, property, and security. In the field of transportation, autonomous vehicles have been developed with the ability to perceive their environment, analyze situations, and make driving decisions without the need for constant human intervention [1]. In the field of healthcare, AI is used to support disease diagnosis, recommend treatment regimens, and analyze health data with increasingly high levels of accuracy [2]. In the financial sector, AI algorithms are involved in credit scoring, automated trading, risk management, and fraud detection [3]. At the same time, in the field of cybersecurity, AI is both used as a tool to protect information systems and can also be exploited to carry out cyberattacks, distribute malicious software, or generate highly sophisticated fake content [4].

However, alongside its significant benefits, AI also poses serious legal risks. In recent years, there have been numerous cases in which AI systems have caused dangerous consequences for society, such as autonomous vehicles involved in fatal accidents, financial algorithms contributing to market manipulation, medical AI systems producing incorrect diagnoses that harm patients, and deepfake technologies being used for fraud, defamation, or the commission of criminal acts in cyberspace [4, 5]. In particular, as AI operates with an increasingly high degree of autonomy and is capable of self-learning from new data, predicting and controlling the behavior of such systems becomes increasingly difficult [1, 6]. This gives rise to major challenges for the traditional principles governing the attribution of criminal liability, particularly with regard to the element of fault (mens rea), the identification of the responsible legal subject, and the causal relationship in criminal liability [7].

One of the most intensely debated issues today is the determination of criminal liability when AI causes harm. In cases where socially dangerous consequences arise from the operation of AI, the question is who should bear criminal responsibility: the programmer, the manufacturer, the operator, the organization deploying the AI system, or the AI system itself [5]. This issue becomes particularly complex in cases where AI systems are capable of self-learning and making decisions beyond the scope of human predictability [1, 6]. In addition, the development of AI has also sparked academic debate as to whether AI can be recognized as a subject of criminal liability. While some views maintain that AI is merely a tool created by humans and cannot bear independent legal responsibility, other scholars have proposed the possibility of recognizing "electronic personhood" for highly autonomous AI systems [7, 8]. These debates reveal an increasingly apparent tension between artificial intelligence and traditional doctrines of criminal liability. Criminal responsibility is typically constructed on the basis of constituent elements, including free will, cognitive capacity, the ability to control one's conduct, and the element of fault attributable to human beings [9, 10]. However, AI lacks consciousness, moral agency, and cognitive capacity in the traditional biological sense. This makes the application of the constitutive elements of criminal liability to AI particularly challenging, especially with regard to the determination of fault, the identification of the responsible legal subject, and the causal relationship between conduct and consequence [11].

Building on these issues, this article focuses on analyzing the challenges that artificial intelligence poses to the

attribution of criminal liability under traditional legal doctrines, while also evaluating liability models currently discussed in legal scholarship and international legislative practice. On that basis, the article proposes directions for legal reform aimed at ensuring effective control of legal risks arising from AI in the context of digital transformation and contemporary technological development. To achieve these research objectives, the study employs a combination of methodological approaches, including doctrinal legal analysis to clarify theoretical issues concerning criminal liability and the legal subject of crime; comparative law methods to examine the approaches adopted by selected countries and international organizations regarding AI-related liability; and case study analysis to assess practical incidents involving harm caused by AI systems. Through these approaches, the article seeks to contribute to both the theoretical and practical foundations for the development of an appropriate criminal liability framework in the age of artificial intelligence.

## Artificial Intelligence and Challenges to the Doctrine of Criminal Liability

### 1. Legal Characteristics of Artificial Intelligence

The development of artificial intelligence has significantly transformed human perceptions of the role of technology in social life. Unlike traditional software systems that operate strictly on pre-programmed instructions, many modern AI systems are capable of data analysis, self-learning, and making decisions with an increasingly high degree of autonomy. These characteristics have made AI a particularly distinctive subject in legal research in general, and in criminal law in particular

The first notable characteristic of AI is its operational autonomy. Modern AI systems are capable of receiving data from their environment, analyzing information, and independently determining courses of action without the need for direct or continuous human intervention [1]. For example, autonomous vehicles can detect obstacles, calculate speed, select driving paths, and make real-time decisions in response to traffic situations [12]. In the financial sector, many algorithmic trading systems are capable of automatically executing buy and sell orders based on market data without requiring human approval for each individual transaction [6]. This autonomy means that AI is no longer merely a "passive tool," but is increasingly capable of directly participating in processes that generate legal consequences [8].

The second characteristic is the capacity for self-learning (machine learning). Modern AI systems can improve their performance by analyzing data and autonomously adjusting their algorithms without the need for complete system reprogramming [1]. This capability enables AI to adapt to new environments and handle complex situations more effectively than humans in certain domains [2]. However, AI's self-learning ability also means that the system's behavior may evolve over time, going beyond the original expectations of programmers or manufacturers. This creates significant challenges in controlling and predicting legal risks [6, 13].

The third characteristic is the unpredictability of AI behavior. Many AI systems, particularly deep learning models, operate through complex data-processing mechanisms that even their developers cannot fully explain in terms of how specific decisions are made [1]. This phenomenon is commonly referred to as the "algorithmic black box" (black box AI) [6]. In practice, AI may produce unusual or unexpected decisions due to the influence of input data, the self-learning process, or algorithmic biases [3]. This unpredictability significantly increases the risk that AI may cause harmful consequences to society, while also posing substantial challenges to traditional mechanisms of legal liability attribution [5, 11].

From a legal perspective, the three characteristics outlined above demonstrate that AI is no longer a simple technological object but has become a factor capable of directly affecting legal order and social safety. This compels criminal law to confront entirely new theoretical and practical challenges.

### 2. Traditional Elements of Criminal Liability

In criminal law scholarship, criminal liability arises when a socially dangerous act is committed by a subject possessing criminal capacity, with a certain degree of fault, and such conduct is prescribed as an offence under criminal law. The traditional doctrine of criminal liability is constructed on a human-centered foundation; consequently, all constitutive elements of criminal liability are closely linked to an individual's capacity for cognition and control of conduct.

The first element is the criminal act. Traditional criminal law defines a crime as being manifested through a concrete human conduct, including an act or omission that poses a danger to society. The criminal act reflects the external manifestation of the subject's will and serves as the basis for assessing the social harmfulness of the offence. In the traditional model, tools or means used in the commission of a crime are not regarded as subjects of criminal liability; instead, they are considered merely instrumentalities that facilitate human conduct.

The second element is fault (mens rea). This is regarded as the core foundation of criminal liability, embodying the principle that "there is no criminal liability without fault." Fault reflects the psychological attitude of the subject toward their conduct and the consequences caused thereby, including intent and negligence. The determination of fault requires an assessment of the subject's cognitive capacity, their ability to foresee consequences, and the volitional choice in engaging in the conduct. Accordingly, the element of fault is inherently and closely linked to the psychological and conscious characteristics of human beings.

The third element is criminal capacity. A subject can only bear criminal liability if they are capable of recognizing the dangerous nature of their conduct and of controlling their behavior. This requirement reflects the demand for both cognitive and volitional capacity on the part of the subject. In traditional criminal law, only human beings or, in certain specific cases, legal persons (corporate entities) may be recognized as subjects of criminal liability.

It can be observed that the entire doctrine of criminal liability is constructed on the premise that the acting subject possesses will, cognition, and the capacity for free choice of conduct. However, as AI becomes increasingly integrated into social activities with a high degree of autonomy, these traditional assumptions begin to reveal significant limitations.

### 3. Difficulties in Applying Criminal Liability to Artificial Intelligence

**First, the difficulty lies in determining fault**

The greatest difficulty in applying criminal liability to artificial intelligence lies in determining the element of

fault. As analyzed above, fault in criminal law reflects the psychological state and volitional attitude of the subject toward the socially dangerous conduct and its resulting consequences [9, 10]. However, AI does not possess consciousness, emotions, or moral awareness in the traditional biological sense. Although it is capable of processing data and making complex decisions, AI still operates solely on the basis of algorithms and computational models [1, 8].

This raises an important question: can fault be attributed to an entity that lacks free will? If an AI system causes serious harm as a result of self-learning and autonomous decision-making beyond human expectations, determining whether the AI acted with "intent" or "negligence" becomes virtually impossible under the traditional approach of criminal law. Even in cases where AI makes erroneous decisions that lead to significant damage, such conduct does not stem from motives, purposes, or moral choice in the way it does for human beings. Accordingly, many scholars argue that AI currently does not satisfy the fundamental conditions required to be recognized as a subject capable of fault in criminal liability [7, 11].

**Second, it is difficult to determine the responsible legal subject**

Another difficulty is identifying the subject who should bear responsibility when AI causes harm. In practice, the development and operation of AI systems typically involve multiple actors, including programmers, manufacturers, data providers, operators, and enterprises that deploy AI. When socially dangerous consequences occur, attributing responsibility to each of these actors becomes highly complex.

Programmers may be held liable if the error originates from algorithm design flaws or technical mistakes in the programming process. Manufacturers may bear responsibility if the AI system contains defects or fails to meet safety standards. Operators may be held accountable if they use the AI system improperly or fail to provide necessary supervision. Meanwhile, enterprises deploying AI may be considered responsible for inadequate risk control mechanisms or for implementing AI in hazardous environments [14, 15].

However, in many cases, the resulting harm does not stem from the direct fault of a specific individual but rather from the complex interaction between algorithms, data, and the operational environment. This undermines the effectiveness of the traditional individual-based attribution model in criminal law [11, 13].

**Third, it is difficult to establish causation**

Alongside the issues of fault and the identification of the responsible subject, proving the causal relationship between conduct and consequence in AI-related cases also presents significant difficulties. Modern AI systems often operate on machine learning mechanisms and complex data-processing structures, making the decision-making process of AI difficult to clearly explain. In particular, the phenomenon of "black box AI" means that even developers may not be able to fully determine why an AI system arrives at a specific decision [6, 13]. In cases where serious harm occurs, investigative and prosecutorial authorities may face difficulties in establishing a direct link between human conduct and the consequences caused by the AI system [5, 11].

In addition, AI behavior may be influenced by various factors, such as input data, the operational environment, or interactions with other technical systems [1, 6]. This makes the causal chain far more complex than in traditional criminal cases [5, 11]. Therefore, AI poses significant challenges to the fundamental principles of modern criminal law, particularly in the determination of fault, the identification of the responsible legal subject, and the causal relationship between conduct and socially dangerous consequences.

**Models for Approaching Criminal Liability in Relation to Artificial Intelligence**

**1. The model that regards AI as merely a tool**

One of the most common approaches in contemporary criminal law scholarship is to regard AI merely as a tool created and used by human beings. According to this view, despite AI's high level of automation and data-processing capability, the essence of AI systems remains that of a technical instrument serving human activity. Therefore, only human beings can be recognized as subjects of criminal liability, while AI cannot be considered a subject of crime [7, 9]. This view is grounded in the traditional doctrine of criminal liability, according to which the subject of a crime must be an entity capable of cognition, controlling its conduct, and possessing fault with respect to socially dangerous behavior [10]. Meanwhile, although AI is capable of processing information and making autonomous decisions, it does not possess consciousness, moral agency, or free will in the traditional legal sense [8]. AI lacks the ability to recognize the social value of its conduct and is also incapable of choosing actions based on moral motives or purposes. Therefore, many scholars argue that AI should only be regarded as a technical instrument similar to other tools used in the commission of crimes, rather than an independent subject of criminal liability [11]. If an AI system is used to carry out socially dangerous conduct, criminal responsibility still lies with the human actors behind its design, control, or deployment [5, 9]. For example, if an individual uses AI technology to generate deepfake content in order to commit fraud, the AI merely serves as a supporting tool for the criminal conduct of the human being [4]. Similarly, if an AI system is programmed to carry out cyberattacks or data theft, the subject bearing criminal liability remains the programmer or the user of the system [8, 11].

The greatest advantage of this model is that it ensures consistency with the traditional theoretical foundations of criminal law. Maintaining the principle that "only human beings can be subjects of criminal liability" helps avoid altering the philosophical basis of modern criminal law, which is grounded in human free will and the element of fault [9, 10]. At the same time, this model also helps to limit the risk of "dehumanizing" criminal responsibility by shifting legal accountability from humans to machines [8]. In addition, the approach that regards AI as merely a tool helps maintain the deterrent effect on actors who develop and use AI systems. If criminal liability were attributed solely to AI while disregarding human involvement, it could lead to a situation in which individuals and organizations evade responsibility [5]. Attributing responsibility to human actors creates pressure on those involved in the development and operation of AI to comply with standards of safety, transparency, and technological risk control [16, 17].

However, this model also reveals several limitations in the context of increasingly autonomous AI systems. In practice, many modern AI systems are capable of self-learning, autonomously adjusting their algorithms, and making decisions beyond the original expectations of their developers [1, 13]. In such cases, treating AI merely as a simple tool may no longer accurately reflect the nature of the process that leads to socially harmful consequences [8].

For example, in the field of autonomous vehicles, AI systems can independently analyze the traffic environment and make real-time driving decisions without direct human intervention [12]. If a serious accident occurs due to an AI decision, determining the criminal liability of a specific individual may become very difficult, especially when the AI's behavior does not entirely result from initial programming errors [5, 11]. This indicates that the model which regards AI merely as a tool does not fully address the legal risks arising from highly autonomous AI systems.

## 2. Indirect Liability Model

In light of the limitations of the traditional model, many scholars and lawmakers have proposed an indirect liability approach. Under this model, AI is still not recognized as a subject of criminal liability; however, legal responsibility may be imposed on individuals or organizations involved in the development, deployment, and operation of AI systems.

Potentially liable subjects include software developers, AI system manufacturers, operating entities, and businesses that deploy AI in practice. The basis of liability does not lie in the direct commission of socially dangerous conduct, but rather in the failure to supervise, the breach of management duties, or the inadequate fulfillment of obligations to control technological risksờ. [5, 16, 17].

For developers, liability may arise where an AI system is designed with insufficient safety safeguards, contains algorithmic errors, or is not adequately tested before deployment. In such cases, the resulting socially harmful consequences may be regarded as stemming from negligence or a lack of due diligence in the design and programming of the AI system [14, 15].

For operators or enterprises using AI, liability may arise when such entities deploy AI systems in high-risk environments without establishing appropriate monitoring mechanisms. For example, if a company puts autonomous vehicles into operation while the system has not yet achieved the required level of safety, or fails to maintain human oversight mechanisms in emergency situations, the company may be held responsible for the resulting consequences [17, 18].

The indirect liability model has the advantage of being consistent with the current operational reality of AI, as AI functioning is typically the result of coordination among multiple actors. Rather than attempting to attribute responsibility directly to AI, this approach focuses on the risk management obligations of the humans and organizations involved. This helps maintain human control over technology while also providing a legal basis for addressing negligent conduct in the development and use of AI systems [19, 20].

However, this model still faces difficulties in cases where AI systems are capable of self-learning and developing behaviors beyond reasonable human foresight. In such situations, proving fault on the part of developers or enterprises may become complex, particularly if they have fully complied with applicable technical standards and ordinary supervisory obligations [6, 13]. Therefore, while indirect liability represents a more feasible approach at the present stage, it still cannot fully resolve the challenges posed by autonomous AI systems [11].

## 3. The "Electronic Personhood" Model

Along with the rapid development of AI, some scholars have proposed recognizing AI as a special legal subject through the "electronic personhood" model. Under this approach, highly autonomous AI systems may be granted a certain legal status, similar to that of corporate legal persons in modern law [7, 8].

This view is based on the fact that many modern AI systems are capable of self-learning, autonomous decision-making, and operating with an increasingly high degree of independence, making it no longer entirely appropriate to regard AI merely as a traditional technical tool [1]. Some scholars argue that if the law can recognize corporate legal entities as subjects of legal responsibility despite not being biological persons, then, in theory, it may also be possible to consider granting a limited legal status to AI in the future [8].

This proposal has also been discussed in European Union documents referring to the possible establishment of an "electronic person" status for robots and highly autonomous AI systems, in order to facilitate the determination of civil liability and compensation for damages [21].

However, the "electronic personhood" model remains highly controversial in legal scholarship. The main obstacle is that AI lacks free will, moral awareness, and the capacity to bear responsibility in the traditional sense of criminal law [9, 10]. Meanwhile, a commercial legal entity is still operated through human will, whereas AI is merely a product of algorithms and data [5, 8].

Many scholars argue that recognizing AI as a subject of criminal liability could undermine the fundamental principle of modern criminal law, which links legal responsibility to the human capacity for cognition and behavioral control [11]. Moreover, if AI were granted legal subject status, the application of criminal penalties would become difficult to implement, as AI lacks the capacity to perceive punishment, rehabilitation, or education in the same way as humans or commercial legal entities [6].

Therefore, although the concept of "electronic personhood" offers certain value in suggesting new approaches to autonomous AI, at the present stage there is still insufficient theoretical and practical basis to recognize AI as a subject of criminal liability [6, 11].

Most modern legal systems continue to maintain the view that legal responsibility related to AI must ultimately be attributed to the human beings or organizations that control such technology [17, 19]. This approach is clearly reflected in the AI governance policy of the European Union, which emphasizes the principle of "human oversight" and the responsibility for risk management borne by entities involved in the development, deployment, and operation of AI systems [16, 18].

## 4. The author's viewpoint

From the perspective of modern criminal law, the author argues that, at the present stage, there is still insufficient theoretical and practical basis to recognize artificial intelligence as a subject of criminal liability. Although AI is increasingly capable of self-learning, autonomous decision-

making, and operating with a high degree of autonomy, in essence, AI is still not an entity possessing free will, moral awareness, and the capacity to bear responsibility in the traditional legal sense. Meanwhile, the doctrine of criminal liability is built upon the principle that fault constitutes the basis of legal responsibility, and fault is inherently linked to the psychological state and cognitive capacity of the subject committing the act.

At present, AI operates solely on the basis of algorithms, data, and information-processing mechanisms established by humans. Although such systems may generate outcomes beyond the original expectations of their developers, this does not mean that AI possesses independent legal consciousness. AI lacks the capacity to recognize the social danger of its conduct, has no criminal motive, and is incapable of choosing its behavior on the basis of moral or legal norms. Therefore, the direct application of doctrines concerning fault, motive, or criminal intent to AI is inconsistent with the nature of criminal liability.

Moreover, recognizing AI as a subject of criminal liability would also make it difficult to preserve the fundamental purposes of criminal punishment. The purpose of punishment is not only to impose retribution, but also to educate, rehabilitate, and prevent crime. However, AI is incapable of experiencing punishment, lacks any sense of remorse, and cannot be educated or rehabilitated through the traditional approaches of criminal law. If AI were regarded as a subject of criminal liability, many existing sanctions would become impracticable and lose their inherent legal significance.

However, the author also argues that continuing to maintain the view that AI is merely a simple tool is no longer appropriate in the context of the rapid development of autonomous AI. Current practice shows that many AI systems are capable of self-learning and making decisions beyond the direct control of humans. In such cases, attributing all responsibility to a specific individual may not accurately reflect the true nature of the process that causes socially dangerous consequences. Therefore, the author argues that a more appropriate approach at the present stage is the application of a hybrid liability model. Under this model, legal responsibility is allocated among the actors involved throughout the entire lifecycle of the AI system, including developers, manufacturers, data providers, operators, and enterprises deploying AI. The determination of liability should be based on each actor's level of involvement, capacity to control the technology, and obligation to manage risks.

The hybrid liability model has the advantage of being consistent with the multi-layered and complex nature of today's AI ecosystem. Rather than attempting to personify AI as an independent subject, this approach focuses on the responsibility of the humans and organizations behind the design, operation, and exploitation of the technology [5, 11].

At the same time, this model also contributes to promoting obligations relating to algorithmic transparency, risk control, and safety oversight in the development of AI [16, 19]. This risk-governance approach is also consistent with international principles on trustworthy AI and human-centered AI, which emphasize the role of human oversight over high-risk AI systems [17, 20].

From a legislative perspective, the application of a hybrid liability model also helps ensure the flexibility of criminal law in response to the rapid development of technology. Rather than fundamentally altering the traditional theory regarding the subject of criminal offenses, the law may instead focus on improving liability mechanisms relating to risk-governance obligations, the liability of commercial legal entities, and obligations to control high-risk AI systems. This is considered a cautious approach that is nevertheless more suitable to current technological developments and legal practice.

**International Experience and Implications for Vietnam**

**1. The Experience of the European Union**

The European Union is currently regarded as one of the leading regions in establishing a legal framework for artificial intelligence. Rather than approaching AI primarily from the perspective of promoting technological innovation, the EU has adopted an approach that balances technological development with the protection of human rights, social safety, and legal responsibility [16]. A notable feature of the EU's legal policy is the enactment of the AI Act - the world's first comprehensive legal instrument governing the development and use of AI [19].

The AI Act does not recognize AI as an independent legal subject, but instead continues to place humans and organizations at the center of the legal liability framework. This instrument approaches AI through a risk-based model of governance, under which AI systems are classified according to their level of danger to society. High-risk AI systems, such as those used in transportation, healthcare, employment recruitment, or law enforcement, are subject to stricter requirements regarding transparency, data control, human oversight, and system safety assurance [18, 22].

One of the notable aspects of the EU's approach is its emphasis on the principle of "human oversight" — human supervision over AI systems. Accordingly, regardless of how highly automated an AI system may be, ultimate responsibility must still rest with the humans or organizations deploying the technology. Developers and enterprises using AI are obliged to assess risks, test the safety of the system, and establish control mechanisms to prevent consequences that may be dangerous to society [17, 18].

In addition, the EU also places particular emphasis on algorithmic transparency and the explainability of AI systems. For high-risk AI systems, enterprises are required to ensure data traceability, maintain records of operational activities, and provide mechanisms for accountability in the event of incidents. This constitutes an important basis for determining legal liability when AI causes harm [23, 24].

It can be seen that the EU's approach is more preventive and risk-governance oriented rather than focused on directly criminalizing the conduct of AI. Instead of attempting to recognize AI as a subject of criminal liability, the EU has chosen to strengthen obligations relating to the management, supervision, and control of technology imposed on individuals and organizations involved in AI [16, 20].

**2. The Experience of the United States**

Unlike the EU, the United States has not yet established a unified federal law on AI, but instead mainly approaches the issue through sector-specific regulations, case law, and product liability mechanisms. The U.S. approach reflects the typical characteristics of a common law system, in which the role of courts and judicial practice is particularly important in determining legal liability related to AI [14, 15].

In the field of legal liability, the United States places greater emphasis on the responsibility of enterprises developing and deploying AI rather than considering AI as an independent subject. When an AI system causes harm, liability is usually determined based on doctrines relating to negligence, product liability, or corporate liability. Manufacturers and enterprises may be held liable if the AI system contains design defects, lacks adequate risk warnings, or fails to ensure reasonable safety standards before being put into use [25, 26].

With respect to autonomous vehicles, numerous debates have emerged in the United States concerning the liability of technology companies when AI causes traffic accidents. In such cases, the legal focus is generally not on whether AI itself should be considered a liable subject, but rather on the obligations of enterprises in designing, testing, and supervising AI systems. This demonstrates that U.S. law continues to uphold the traditional principle that ultimate legal responsibility must rest with human beings or commercial legal entities [`12, 14].

In addition, the United States also emphasizes the role of internal compliance mechanisms and corporate governance in AI development activities. Major technology companies are often required to establish risk-assessment systems, algorithm-testing procedures, and safety oversight mechanisms in order to minimize the risk of legal liability. In many cases, corporate liability arises not only from conduct that directly causes harm, but also from the failure to implement appropriate control mechanisms for AI systems [27].

Overall, the experience of the United States demonstrates a trend toward expanding the liability of commercial legal entities and strengthening technology governance obligations, rather than fundamentally altering the doctrine of criminal liability subjects [25].

## 3. Implications for Vietnam

The rapid development of AI is creating an urgent need for Vietnam to establish and refine an appropriate legal framework to manage the risks arising from this technology. Although AI is increasingly being applied across various sectors such as finance, transportation, healthcare, and public administration, the current legal system of Vietnam still lacks specific regulations governing legal liability related to AI, particularly criminal liability.

First, Vietnam needs to develop a comprehensive legal framework for AI that balances the promotion of innovation with the control of technological risks. This framework should clearly define the fundamental principles governing the development and use of AI, especially the principles of safety, transparency, accountability, and human oversight over high-risk AI systems. Establishing a risk-governance mechanism based on the level of danger posed by AI may be an appropriate approach given Vietnam's current stage of technological development.

In addition, Vietnam needs to further improve the legal framework governing the criminal liability of commercial legal entities. In the context where AI is often developed and deployed by technology enterprises, expanding risk-management obligations and technological control responsibilities for commercial legal entities is a necessary requirement. The law should clearly stipulate the responsibilities of enterprises in testing, monitoring, and ensuring the safety of AI systems that may cause serious consequences to society.

Another important issue is the introduction of obligations relating to algorithmic control and transparency. Enterprises developing AI should be required to retain operational data and ensure the traceability and explainability of AI system decisions in the event of incidents. This serves as an important basis for supporting investigations, proving fault, and determining legal liability when AI causes damage.

In addition, Vietnam also needs to strengthen its capacity for digital investigation and technological forensics in support of criminal proceedings. Cases involving AI are often technically complex, requiring prosecuting authorities to possess the capability to analyze digital data, assess algorithms, and determine the operational mechanisms of AI systems. Therefore, training specialized human resources in digital investigation, cybersecurity, and algorithmic forensics will be of particular importance in the context of the current digital transformation.

International experience shows that the prevailing trend among countries today is not to recognize AI as a subject of criminal liability, but rather to strengthen the responsibility of individuals and organizations in the development, operation, and control of technology. This is also an appropriate approach for Vietnam at the present stage, in order to ensure a balance between technological development and the protection of legal security and social order.

## Conclusion

The rapid development of artificial intelligence is bringing about profound changes in social life while simultaneously posing unprecedented challenges to modern criminal law theory. Unlike traditional technical tools, AI is increasingly capable of self-learning, autonomous decision-making, and operating with a high degree of autonomy in many critical sectors such as transportation, healthcare, finance, and cybersecurity. This creates the risk of harmful consequences to society that traditional mechanisms of criminal liability are not yet able to address fully and effectively.

The analysis shows that AI is directly challenging the fundamental theoretical foundations of criminal law, particularly the doctrines of fault, the subject of crime, and causation. Traditional criminal law is built upon the assumption that the actor possesses free will, cognitive capacity, and the ability to control their behavior [9, 10]. However, although contemporary AI systems are capable of processing information and making complex decisions, they still lack moral consciousness, motives, or will in the traditional legal sense [8]. Therefore, the direct application of the doctrine of fault to AI encounters numerous difficulties in both theoretical and practical terms [11].

The article also demonstrates that there is currently insufficient basis to recognize AI as a subject of criminal liability. The personification of AI as an independent legal subject not only lacks a foundation in terms of will and criminal responsibility capacity, but also undermines the significance of criminal punishment, which is fundamentally intended to educate, rehabilitate, and deter human beings. Meanwhile, AI lacks the capacity to perceive punishment or reform its behavior in the same way as humans or commercial legal entities.

Based on technological developments and international legal experience, it can be seen that a hybrid liability model is a more appropriate approach in the current context. Under this model, legal responsibility should be allocated among the parties involved throughout the entire process of

designing, developing, deploying, and operating AI systems, including developers, manufacturers, operators, and enterprises utilizing the technology. This approach both preserves the traditional principles of criminal law and meets the need for risk control in the digital technological environment.

In the coming years, alongside the development of the digital society and autonomous technologies, criminal law should continue to be refined toward strengthening technological risk-governance responsibilities, ensuring algorithmic transparency, and enhancing the digital investigative capacity of prosecuting authorities. At the same time, the development of an appropriate legal framework for AI must be pursued in a manner that balances the promotion of innovation with the protection of social safety, human rights, and legal order in the digital era.

**References**


1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed., Pearson Education, 2021, 1027–1058.
2. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, New York, 2019, 44–79.
3. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, New York, 2016, 128–156.
4. King TC, Aggarwal N, Taddeo M, Floridi L. Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. In: Dubber MD, Pasquale F, Das S (eds.), The Oxford Handbook of Ethics of AI. Oxford University Press, 2020, 775–798.
5. Hallevy G. Liability for Crimes Involving Artificial Intelligence Systems. Springer, Cham, 2015, 157–184.
6. Pasquale F. The Black Box Society: The Secr*et al*gorithms That Control Money and Information. Harvard University Press, Cambridge (Massachusetts), 2015, 187–220.
7. Hallevy G. The Criminal Liability of Artificial Intelligence Entities – From Science Fiction to Legal Social Control. Akron Intellectual Property Journal,2010:4(2):171–201.
8. Pagallo U. The Laws of Robots: Crimes, Contracts, and Torts. Springer, Dordrecht, 2013, 91–128.
9. Ashworth A. Principles of Criminal Law. 9th ed., Oxford University Press, Oxford, 2021, 83–137.
10. Alexander L, Ferzan KK. Crime and Culpability: A Theory of Criminal Law. Cambridge University Press, Cambridge, 2009, 69–118.
11. Lai L, Contissa G. Artificial Intelligence and Criminal Liability: Theoretical and Comparative Perspectives. Artificial Intelligence and Law,2021:29(4):567–593.
12. Goodall NJ. Machine Ethics and Automated Vehicles. Road Vehicle Automation, 2014, 2, 93–102.
13. Burri T. Machine Learning and the Law: Five Theses. Springer International Publishing, 2017, 2, 443–466.
14. Abbott R. The Reasonable Computer: Disrupting the Paradigm of Tort Liability. George Washington Law Review,2020:86(1):1–52.
15. Barfield W, Trong pha là đấy Pagallo U. Advanced Introduction to Law and Artificial Intelligence. Edward Elgar Publishing, 2018, 89–117.
16. European Commission. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. Brussels, 2020, 11–24.
17. OECD. OECD Principles on Artificial Intelligence. Paris, 2019.
18. Schuett J. Risk Management in the Artificial Intelligence Act. European Journal of Risk Regulation,2023:14(2):278–298.
19. European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, 2024.
20. UNESCO. Recommendation on the Ethics of Artificial Intelligence. Paris, 2021.
21. European Parliament. Civil Law Rules on Robotics, European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL), paras. 59(f)–60.
22. Ho-Dac M. First Analysis of the EU Artificial Intelligence Act: Towards a Global Standard for Trustworthy AI? European Papers,2024:9(3):1–24.
23. Silva NSE. The Artificial Intelligence Act: Critical Overview. UNIO – EU Law Journal,2024:10(2):45–68.
24. Fabiano N. Subject Roles in the EU AI Act: Mapping and Regulatory Implications. arXiv Preprint,2025:arXiv:2510.13591:1–31.
25. Calo R. Robotics and the Lessons of Cyberlaw. California Law Review,2015:103(3):513–563.
26. Marchant GE, Lindor RA. The Coming Collision Between Autonomous Vehicles and the Liability System. Santa Clara Law Review,2012:52(4):1321–1340.
27. National Institute of Standards and Technology (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce, Washington D.C., 2023, 1–54.