

Analysis of the Criminal Subject Qualification of Artificial Intelligence

Ziqi Yang*

Guizhou University, Guiyang550025, Guizhou, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the rapid development of autonomous decision-making systems, generative artificial intelligence, embodied robots, and algorithmic agents, criminal law is increasingly confronted with the question of whether artificial intelligence can or should be recognized as a criminal subject. Traditional criminal law is built upon human-centered assumptions: a criminal subject must possess legal capacity, behavioral capacity, culpability, and the ability to bear punishment. However, advanced AI systems may independently generate harmful outputs, manipulate information, control physical devices, or participate in decision-making processes that lead to legally significant harm. This paper adopts a systematic review method to examine the criminal subject qualification of artificial intelligence from the perspectives of legal personhood, actus reus, mens rea, culpability, punishment, and accountability allocation. It argues that current AI systems should not be recognized as independent criminal subjects in the full sense because they lack consciousness, moral self-reflection, normative understanding, and punishment responsiveness. Nevertheless, AI challenges the traditional binary structure of “human subject–technical object” and requires a reconstructed accountability framework. The paper proposes a layered model of criminal responsibility in which developers, deployers, operators, corporate entities, and AI systems are functionally distinguished. AI should be treated as a risk-generating autonomous instrument rather than a punishable offender, while criminal law should strengthen duties of design, supervision, risk assessment, explainability, and post-incident traceability. This approach can avoid both excessive anthropomorphism and responsibility gaps, thereby preserving the normative foundation of criminal law while responding to the practical risks created by artificial intelligence.

Keywords: Artificial intelligence; Criminal subject; Criminal liability; Legal personhood; Autonomous systems; Mens rea; Culpability; AI governance; Criminal law

Online publication: July 15, 2026

1. Introduction

Artificial intelligence has moved from narrow computational assistance to complex systems capable of learning, prediction, generation, and autonomous interaction with social environments. In areas such as

autonomous vehicles, medical diagnosis, financial trading, policing technologies, smart manufacturing, and generative content production, AI systems may produce outcomes that are difficult to attribute exclusively to a single human actor^[1-3]. When these outcomes cause death, injury, property loss, fraud, discrimination, cyber intrusion, or public-security risks, criminal law faces a difficult question: should AI remain merely an object or tool of human conduct, or can it become a criminal subject with independent liability?

The issue of AI criminal subject qualification differs from general civil liability or administrative regulation. Criminal law is not merely a mechanism of compensation or risk distribution; it is a system of moral condemnation, culpability judgment, punishment, and social blame. Therefore, recognizing AI as a criminal subject would require more than acknowledging that AI can cause harm. It would require proof that AI can perform legally relevant conduct, possess a blameworthy mental state, understand normative commands, and respond meaningfully to punishment. These requirements are deeply connected with the legitimacy of criminal punishment and the principle that there should be no crime or punishment without culpability^[4].

This paper selects the systematic review method because the topic is primarily theoretical, doctrinal, and normative rather than based on original statistical data. It reviews representative literature on AI legal personhood, robot criminal liability, moral agency, autonomous systems, and regulatory governance. The central argument is that AI currently lacks full criminal subject qualification, but its increasing autonomy requires criminal law to develop a more precise responsibility-allocation framework. The paper is divided into three main sections: the doctrinal basis of criminal subject qualification, the arguments for and against AI criminal subject status, and the realization path for reconstructing criminal accountability in the AI era.

2. Doctrinal basis of criminal subject qualification in the context of AI

Criminal subject qualification refers to the legal status of an entity that can commit a crime and bear criminal responsibility. In traditional criminal law, natural persons are the primary criminal subjects because they possess biological existence, consciousness, will, rationality, and moral agency. Many legal systems also recognize corporate criminal liability, but corporate liability is usually justified through organizational decision-making, collective action, and attribution from human agents. Whether the subject is a natural person or a corporation, criminal law still depends on human-oriented concepts such as intention, negligence, fault, culpability, deterrence, and punishment^[5].

The criminal subject must normally satisfy at least four conditions. First, it must have conduct capacity, meaning that it can perform an act or omission that produces legally relevant consequences. Second, it must have normative capacity, meaning that it can understand or be expected to understand legal norms. Third, it must have culpability capacity, meaning that blame can be meaningfully attached to its conduct^[6-7]. Fourth, it must have punishment capacity, meaning that criminal sanctions can produce retribution, deterrence, rehabilitation, incapacitation, or expressive condemnation. These conditions show why criminal subject qualification is narrower than causal responsibility. An entity may cause harm without being a criminal offender.

AI systems complicate these requirements because they can increasingly perform autonomous or semi-autonomous actions. For example, an autonomous vehicle may choose a driving path, a trading algorithm may execute market transactions, and a generative AI system may create deceptive content. These actions

may appear functionally similar to human acts. However, functional similarity does not automatically establish criminal subject status. AI systems process data according to models, parameters, training objectives, and environmental inputs, but current systems do not possess consciousness, self-understanding, moral remorse, or genuine comprehension of legal obligation. Therefore, AI can satisfy a limited form of conduct capacity but not full culpability or punishment capacity (**Table 1**)^[8].

Table 1. Criteria for criminal subject qualification and their application to AI

Criterion	Meaning in Criminal Law	Application to Natural Persons	Application to Corporations	Application to Current AI
Conduct capacity	Ability to perform legally relevant acts or omissions	Fully recognized	Recognized through organizational acts	Partially recognized through autonomous outputs
Normative capacity	Ability to understand legal rules and obligations	Fully recognized, except in special incapacity cases	Constructed through compliance systems	Not genuinely recognized
Culpability capacity	Ability to be blamed for intentional or negligent wrongdoing	Core basis of liability	Attributed through corporate fault or human agents	Largely absent
Punishment capacity	Ability to receive punishment with deterrent or expressive meaning	Fully recognized	Recognized through fines, sanctions, compliance orders	Weak; deletion or shutdown is risk control, not punishment
Legal-policy justification	Social need to impose criminal responsibility	Strong		

3. Arguments for and against recognizing AI as a criminal subject

Supporters of AI criminal subject qualification usually rely on three arguments. The first is the autonomy argument: when AI systems make decisions not directly controlled by humans, it becomes difficult to treat them merely as passive tools. The second is the responsibility-gap argument: if harm cannot be fully attributed to programmers, manufacturers, users, or corporations, recognizing AI as a criminal subject may prevent impunity. The third is the analogy argument: because corporations are artificial legal persons, AI systems might also be granted limited legal personality for purposes of liability. These arguments are important because they reveal genuine pressure on traditional criminal-law categories^[9].

However, each argument has limits. Autonomy in technical operation is not equivalent to autonomy in moral or legal agency. An AI system may produce outputs that were not specifically predicted by its developer, but unpredictability does not prove that the system understands norms or chooses to violate them. The responsibility-gap argument is also incomplete because criminal law should first examine whether human actors failed in design, training, testing, deployment, monitoring, or risk control. If the law too quickly blames AI, human institutions may escape responsibility. The corporate analogy is also problematic because corporate criminal liability ultimately depends on human organization, governance structures, compliance duties, and institutional benefit, whereas AI itself lacks social membership and normative consciousness.

The strongest objection to AI criminal subject qualification concerns punishment. Punishment is not only the imposition of cost; it is a normative response to blameworthy wrongdoing. A fine imposed on a corporation can influence shareholders, managers, and compliance structures. Imprisonment imposed on a person restricts liberty and expresses condemnation. By contrast, “punishing” an AI system through deletion, shutdown, retraining, or access restriction is closer to risk control than criminal punishment. Such measures

may be necessary, but they do not prove that AI is a proper criminal subject. Therefore, current AI should not be recognized as an independent criminal offender; instead, it should be regulated as an autonomous risk source embedded in human and organizational systems (**Table 2**).

Table 2. Proposed responsibility allocation model for AI-related criminal harm

Actor	Possible Source of Liability	Typical Risk Scenario	Recommended Legal Response
Developer	Defective design, unsafe training, intentional harmful coding	AI system designed to deceive, attack, or manipulate	Criminal liability for intent, recklessness, or gross negligence
Deployer	Unsafe implementation, lack of supervision, use in prohibited contexts	High-risk AI used without human oversight	Duty-based negligence or omission liability
Operator/User	Misuse, illegal instruction, disabling safeguards	AI used for fraud, cybercrime, or harmful automation	Direct criminal liability as principal or accomplice
Corporation	Organizational failure, profit-driven risk tolerance, weak compliance	Company releases unsafe AI despite known risks	Corporate criminal liability and compliance sanctions
AI system	Autonomous output or system behavior causing harm	Unpredictable harmful decision or generated illegal content	

4. Reconstruction of criminal accountability for AI-related harm

Although AI should not currently be treated as a full criminal subject, criminal law must still respond to AI-related harm. The first path is to clarify human responsibility across the AI lifecycle. Developers may bear responsibility when they intentionally design harmful systems, ignore obvious safety risks, manipulate training data, or fail to implement reasonable safeguards. Deployers may bear responsibility when they use AI in prohibited or high-risk contexts without adequate supervision. Operators may bear responsibility when they misuse AI, disable safety functions, or knowingly rely on unreliable outputs. Corporate entities may bear responsibility when organizational policies prioritize profit over safety or create incentives for unlawful AI deployment ^[10].

The second path is to establish a duty-based model of criminal negligence and omission. AI-related harm often results not from a single intentional act but from accumulated failures in risk assessment, testing, documentation, monitoring, and emergency response. Criminal law can develop clearer duties for high-risk AI systems, including duties to evaluate foreseeable harm, preserve technical logs, maintain human oversight, conduct impact assessments, update safety mechanisms, and withdraw unsafe systems. Where serious harm results from a gross violation of these duties, criminal liability should attach to responsible human or organizational actors. This model avoids blaming AI itself while still preventing accountability gaps ^[11].

The third path is to create a functional legal status for AI systems without granting full criminal subject qualification. AI systems can be treated as legally relevant autonomous instruments that trigger special duties, evidence rules, and risk-control obligations. For example, law may require registration of high-risk AI systems, traceability of model decisions, auditability of training data, explainability of critical outputs, and emergency shutdown mechanisms. In criminal proceedings, AI-generated decisions should be examined as evidence of system behavior, human oversight, and organizational control. This functional approach recognizes the special role of AI in causing harm while preserving the principle that criminal punishment must ultimately be directed at morally and legally responsible subjects ^[12].

5. Conclusion

The criminal subject qualification of artificial intelligence is one of the most challenging issues in contemporary criminal-law theory. AI systems can already generate harmful consequences with a degree of autonomy that complicates traditional attribution models. Nevertheless, causal autonomy should not be confused with criminal subjectivity. Criminal subject qualification requires conduct capacity, normative understanding, culpability, and punishment responsiveness. Current AI systems may possess functional decision-making capacity, but they do not possess consciousness, moral self-reflection, legal understanding, or the capacity to experience criminal punishment as blame. Therefore, recognizing AI as an independent criminal subject would weaken the culpability principle and risk allowing human actors to transfer responsibility to machines.

A more appropriate solution is to reconstruct criminal accountability around the AI lifecycle. Developers, deployers, users, operators, and corporations should bear differentiated responsibility according to their control, knowledge, benefit, and duty of care. Criminal law should focus on intentional misuse of AI, reckless deployment of dangerous systems, gross negligence in supervision, and organizational failure to manage foreseeable risks. At the same time, AI systems should be given a functional legal status as autonomous risk-generating instruments, which allows the law to impose registration, audit, explainability, traceability, and safety-control obligations without anthropomorphizing AI as a punishable offender.

In the long term, if future AI systems develop consciousness, self-understanding, normative reasoning, and genuine moral agency, the question of criminal subject qualification may need to be reopened. For the present stage, however, criminal law should resist premature recognition of AI personhood. The priority should be to close responsibility gaps through human-centered accountability, corporate compliance, risk-based regulation, and technical governance. In this way, criminal law can respond effectively to AI-related harm while maintaining its core principles of legality, culpability, proportionality, and human dignity.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Abbott R, 2016, I Think, Therefore I Invent: Creative Computers and the Future of Patent Law. *Boston College Law Review*, 57(4): 1079–1126.
- [2] Bathae Y, 2018, The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, 31(2): 889–938.
- [3] Chopra S, White LF, 2011, A Legal Theory for Autonomous Artificial Agents. University of Michigan Press, Michigan.
- [4] European Parliament, 2017, European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
- [5] European Parliament and Council of the European Union, 2024, Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
- [6] Gless S, Silverman E, Weigend T, 2016, If Robots Cause Harm, Who Is to Blame? Self-driving Cars and Criminal Liability. *New Criminal Law Review*, 19(3): 412–436. <https://doi.org/10.1525/nclr.2016.19.3.412>

- [7] Hallevy G 2010, The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control. *Akron Intellectual Property Journal*, 4(2): 171–201.
- [8] King TC, Aggarwal N, Taddeo M, et al., 2020, Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 2020(26): 89–120. <https://doi.org/10.1007/s11948-018-00081-0>
- [9] Matthias A, 2004, The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology*, 2004(6): 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- [10] Simmler M, Markwalder N, 2019, Guilty Robots? Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence. *Criminal Law Forum*, 2019(30): 1–31. <https://doi.org/10.1007/s10609-018-9360-0>
- [11] Solum LB, 1992, Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, 70(4): 1231–1287.
- [12] Sullins JP, 2006, When is a Robot a Moral Agent? *International Review of Information Ethics*, 2006(6): 23–30.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.