



DOI: <https://doi.org/10.38035/gijlss.v4i3>
<https://creativecommons.org/licenses/by/4.0/>

Hybrid Criminal Liability Based on Risk-Based Approach in Artificial Intelligence-Based Crimes in Indonesia

Mohammad Erfan^{1*}, Nanik Sutarni², Ananda Megha Wiedhar Saputri³, Sri Budi Raharjo⁴

¹Universitas Boyolali, Jawa Tengah, Indonesia, irfanharajuku14@gmail.com

²Universitas Boyolali, Jawa Tengah, Indonesia, naniksutarni65@yahoo.com

³Universitas Boyolali, Jawa Tengah, Indonesia, wiedharsaputrianandamegha@gmail.com

⁴Universitas Boyolali, Jawa Tengah, Indonesia, 1000diraharjo@gmail.com

*Corresponding Author: irfanharajuku14@gmail.com¹

Abstract: The development of Artificial Intelligence (AI) has introduced new challenges in criminal law, particularly regarding criminal liability in technology-based crimes. AI systems capable of operating semi-autonomously blur the relationship between the perpetrator, the act, and culpability, rendering the conventional fault-based principle of *guilt without fault* increasingly inadequate. This study aims to analyze the existing Indonesian criminal law framework governing AI-related crimes and to formulate an ideal model of criminal liability in response to such developments. This research employs a normative juridical method with statutory and conceptual approaches. The findings indicate that current regulations in Indonesia remain limited and have not comprehensively addressed the complexity of AI-based crimes. Accordingly, this study proposes a hybrid criminal liability model based on a risk-based approach, integrating strict liability, vicarious liability, and risk-based accountability as mechanisms for proportional responsibility distribution without recognizing AI as a legal subject. This model is expected to contribute to the reform of criminal law toward a more adaptive, responsive, and just legal system in addressing AI-related crimes.

Keywords: Artificial Intelligence, criminal liability, hybrid liability, risk-based approach, criminal law.

INTRODUCTION

The development of digital technology, particularly Artificial Intelligence (AI), has brought significant transformations to various aspects of life, including the realm of criminal law. AI has transcended its role as a mere tool, evolving into a system with self-learning capabilities (machine learning) that can generate outputs resembling human actions (Haris & Tantimin, 2022). This situation poses serious challenges for criminal law, particularly in determining the form of accountability for crimes involving AI (Hutapea et al., 2025).

A fundamental distinction must therefore be drawn between Artificial Intelligence functioning merely as a tool under direct human control and AI systems possessing a certain

degree of operational autonomy. In the former situation, AI primarily serves as an instrument through which a human actor executes a predetermined intention: the user defines the objective, provides the relevant instructions or inputs, and exercises substantial control over the resulting output. In such circumstances, the use of AI does not necessarily alter the conventional attribution of criminal responsibility, because the causal and culpability relationship between the human actor and the criminal conduct remains relatively identifiable. By contrast, AI systems with a certain degree of autonomy may perform tasks, adapt their operations, generate outputs, or make system-level decisions with limited direct human intervention at each stage of the process. Autonomy in this context should not be understood as legal personhood or independent legal agency, but rather as a technological capacity to operate beyond continuous and specific human instruction. The distinction is therefore not binary but exists along a spectrum of human control and machine autonomy.

The degree of autonomy is particularly relevant to the construction of criminal liability because it affects the extent to which a human actor can reasonably predict, control, prevent, and attribute the consequences generated by an AI system. Where AI operates predominantly as a human-controlled tool, fault-based liability remains generally capable of addressing the resulting criminal conduct. However, as the system acquires greater operational autonomy, the possibility of unpredictable outputs, unintended system behavior, and reduced direct human control increases. This creates a potential gap between the occurrence of harmful consequences and the capacity to establish conventional *mens rea*. Accordingly, a risk-based liability model becomes relevant not merely because AI is technologically sophisticated, but because a combination of autonomy, potential harm, and the distribution of risk-control capacities among the actors involved may render conventional individual fault-based attribution insufficient.

This phenomenon is increasingly evident with the emergence of various AI-based crimes, one of which is the use of deepfake technology in fraud. In a case that occurred in East Java in 2025, the perpetrator utilized AI technology to create a video impersonating a regional head to convince the victim to hand over money, resulting in total losses reaching tens of millions of rupiah (Pamungkas et al., 2025). This case demonstrates that AI is no longer merely a passive tool but has become a medium that actively facilitates the commission of intricate crimes, making it difficult to identify the perpetrator and appropriate forms of criminal accountability (Eurike et al., 2024).

From a classical criminal law perspective, criminal liability is based on the principle of "geen straf zonder schuld," which asserts that there is no crime without fault. This principle requires a link between the act and the perpetrator's fault (*mens rea*), which generally can only be attributed to humans as legal subjects (Sudarto, 2024; Rohmatulloh, 2024). However, in the context of AI, which is capable of semi-autonomous action, the link between the perpetrator, the act, and the fault becomes less clear (Ramadhani et al., 2025). In many cases, actions produced by AI systems cannot be directly attributed to the will of a specific human, creating gaps in the application of conventional criminal liability concepts (Setiawan & Wijayanto, 2024).

In Indonesia, regulations regarding criminal liability for AI-based crimes are still limited. AI is still positioned as part of an electronic system within the framework of the Electronic Information and Transactions Law, so criminal liability remains the responsibility of humans as legal subjects (Rahman & Habibulah, 2019). However, this approach fails to fully address the complexities of the relationship between humans and AI, particularly in cases involving autonomous and self-learning systems (Hutapea et al., 2024). Consequently, there is a legal vacuum and uncertainty in determining who should be held responsible for crimes involving AI (Nawawi, 2026).

A review of relevant academic literature reveals three groups of studies that systematically map the development of thinking on criminal liability in the context of artificial intelligence. The first group of studies focuses on the adequacy of the doctrine of fault-based liability as an instrument of accountability in technology-based crimes. Haris and Tantimin (2022), in their analytical study of the criminal law dimensions of AI utilization in Indonesia, concluded that the prevailing normative constructions lack the capacity to encompass the complex relationship between human legal subjects and semi-autonomous AI systems. Similarly, Rahman and Habibulah (2019), through a critical examination of the Electronic Information and Transactions Law, assert that Indonesia's positive legal regime consistently positions AI as a mere instrument, thus shifting the burden of criminal liability entirely to humans as the sole recognized legal subjects. Rohmatulloh (2024) enriches this discourse by noting that although the 2023 Criminal Code has accommodated a number of contemporary legal developments, the paradigm of fault-based liability remains a normative foundation that has not undergone a paradigmatic shift. The second group of studies moves beyond the confines of the fault doctrine by exploring alternative liability frameworks that better align with the characteristics of AI technology. Anggraeny et al. (2024) argue for the application of strict liability to AI-based crimes, premise that reliance on *mens rea* creates structural impunity that cannot be justified normatively. Ihsan (2024), using a comparative perspective between Islamic law and Indonesian positive law, examines the relevance of vicarious liability in the context of AI-developing corporations and concludes that this approach has limited reach because it is only effective within strictly defined hierarchical relationships. Setiawan and Wijayanto (2024), in their comprehensive juridical study, identify the inability of any single approach to satisfactorily address the complexities of AI-based criminal liability. Hutapea et al. (2024) in their preliminary study emphasize the urgency of a comprehensive conceptual reconstruction of Indonesian criminal law, a recommendation reinforced by Hutapea et al. (2025), who specifically identify the absence of specific regulations as the primary source of the significant legal vacuum in the enforcement of AI-based criminal law.

The third group of studies introduces a risk-based approach as an alternative that is paradigmatically more responsive to the high-risk characteristics of AI technology. Sudarto (2024) explicitly argues that the hegemony of the fault-based approach in Indonesian criminal law is no longer epistemologically tenable in the face of the nonlinear and high-risk dynamics of digital technology. Nawawi (2026) strengthens this thesis by identifying the absence of an adequate criminal liability framework for AI as a normative deficit that demands a fundamental conceptual breakthrough. However, the studies in these three groups are generally partial and have not been able to present an integrative framework that systematically unites strict liability, vicarious liability, and risk-based accountability into a comprehensive and operationalizable accountability model (Hutapea et al., 2025). This is the research gap that this study aims to address.

Building on the research gaps identified above, this study constructs a multi-layered analytical framework based on the hierarchy of legal theories commonly used in normative legal research. This study bases its argument on Gustav Radbruch's three legal values, which postulate that a legitimate legal system must simultaneously realize three intertwined fundamental values: justice, legal certainty, and utility. Through the lens of Radbruch's theory, it can be argued that criminal law's reliance on a single accountability approach, particularly in addressing AI-based crimes, constitutes a form of normative reductionism that systematically ignores the values of distributive justice and social utility, as it fails to distribute the burden of responsibility proportionally across all actors in the technological ecosystem that have the potential to cause harm. As a philosophical complement, Satjipto Rahardjo asserts that law cannot and should not be confined within the confines of rigid

normative texts, but rather must proactively and emancipatorily respond to social change and technological developments to achieve substantive justice that favors the interests of the wider community (Radbruch, 2006; Rahardjo, 2009).

In Criminal Responsibility, as systematically constructed by Moeljatno and further developed by Roeslan Saleh (Moeljatno, 2008; Saleh, 1983), criminal responsibility is normatively fulfilled if three elements are present: the existence of a criminal act that can be attributed, the existence of fault in the form of intent or negligence, and the absence of an excuse that eliminates the element of responsibility. In AI-based crimes, this theory operates as an analytical instrument to demonstrate why conventional liability frameworks suffer structural dysfunction, since semi-autonomous AI systems disrupt the causal nexus between human intent and outcomes, thereby rendering the *mens rea* element legally indistinct and practically unprovable. In addition, the Integrative Theory of Punishment is used as a normative framework to explain the objectives of punishment that the proposed hybrid model aims to realize. Different from retributive punishment theory which is solely oriented towards punishing the perpetrator, integrative punishment theory emphasizes that punishment must synergistically realize retributive, preventive, and restorative objectives. This orientation is normatively relevant in the context of controlling the systemic risks of the use of AI technology. These two intermediate theories serve as an epistemological bridge between the fundamental values established at the grand theory level and the technical-operational construction of the accountability model, namely the integration of strict liability, vicarious liability, and risk-based accountability as a mechanism for proportional distribution of responsibility in the AI ecosystem (Muladi, 1985; Muladi & Arief, 2010).

This normative situation necessitates a paradigmatic shift in the construction of criminal accountability, from a fault-based liability paradigm to a risk-based approach. In this paradigm, determining accountability depends not solely on proof of subjective fault but also on the extent to which a party has the capacity to control, mitigate, and prevent the risks inherent in the use of AI technology, as reflected in the duty of care that can be normatively imposed on developers, service providers, and users of AI systems (Ramadhani et al., 2025).

The scientific novelty of this research lies in the construction of a hybrid criminal liability model based on a risk-based approach that synergistically and systematically integrates the three operational concepts of strict liability, vicarious liability, and risk-based accountability into a comprehensive and legally operationalized accountability framework. The fundamental differentiation of this research from previous studies lies in its rejection of the monistic approach: instead of advocating for a single, exclusive accountability mechanism as done by Anggraeny et al. (2024), Ihsan (2024), and Hutapea et al. (2025), this research builds an argument that the relational complexity in the AI ecosystem that simultaneously involves developers, service providers, and users can only be adequately addressed through an integrative model that distributes the burden of accountability proportionally based on the level of control and risk mitigation capacity of each party, without constructing AI as a separate legal subject (Wijanarko et al., 2026).

METHOD

This study adopts a prescriptive doctrinal legal approach, aiming not only to describe and analyze existing legal norms (*das sein*), but also to construct normative arguments on how the law should be reformed (*das sollen*) to address issues of criminal liability in Artificial Intelligence-based crimes. This prescriptive orientation positions normative legal research not as a merely descriptive activity, but as an intellectual endeavor to generate normative propositions that can serve as a foundation for reforming Indonesian criminal law (Soekanto & Mamudji, 2007; Hutchinson, 2013). This research employs four integrated operational approaches. First, the statute approach is conducted through the examination of

the KUHP (UU Nomor 1 Tahun 2023), UU Nomor 19 Tahun 2016 concerning Electronic Information and Transactions, and UU Nomor 27 Tahun 2022 concerning Personal Data Protection, as relevant positive legal instruments. Second, the conceptual approach is used to examine and construct relevant criminal law doctrines, including criminal responsibility, strict liability, vicarious liability, and risk-based accountability. Third, the philosophical approach is employed to analyze the value legitimacy of the proposed model based on Radbruch's Three Values Theory of Law and Satjipto Rahardjo's Progressive Law Theory. Fourth, a comparative approach is used to examine AI regulations in other jurisdictions, particularly the EU Artificial Intelligence Act 2024, in order to identify regulatory principles that may be adapted to the context of Indonesian criminal law reform.

Legal sources are systematically classified into three categories. Primary legal materials include the 1945 Constitution of the Republic of Indonesia; UU Nomor 1 Tahun 2023 concerning the Criminal Code; UU Nomor 19 Tahun 2016 concerning Electronic Information and Transactions; UU Nomor 27 Tahun 2022 concerning Personal Data Protection; and other laws and regulations directly related to the research object. Secondary legal materials include scholarly literature in the form of criminal law textbooks, nationally and internationally indexed journal articles, previous research results, and policy documents discussing criminal liability, AI regulation, and criminal law reform. Tertiary legal materials include legal dictionaries, legal encyclopedias, and glossaries of terms used to clarify technical legal concepts and terminology. The following inclusion criteria were applied in selecting secondary legal materials: (1) substantial relevance to the research topic; (2) indexed in verified academic databases; and (3) published within a timeframe relevant to developments in global AI regulations, namely between 2019 and 2026. The legal materials were collected through systematic library research of sources with verifiable scientific validity.

The analysis was conducted qualitatively and normatively using deductive legal reasoning as the primary method of inference. The reasoning process proceeds from general principles and doctrines of criminal responsibility to the specific legal problems arising from AI-based crimes. Legal syllogism is employed by positioning applicable criminal law principles and liability doctrines as the major premise, the characteristics of AI systems and the respective roles, degrees of autonomy, and risk-control capacities of developers, platform or service providers, users, and corporations as the minor premise, and the appropriate form and distribution of criminal liability as the resulting legal proposition. The construction of the proposed model is further developed through proportionality-based legal reasoning by considering the degree of AI autonomy, the level of risk generated, the actor's capacity to control and mitigate such risk, and the corresponding duty of care. Accordingly, the model does not impose a uniform form of liability on all actors but determines the appropriate application of fault-based liability, strict liability, or vicarious liability within the broader framework of risk-based accountability.

Several legal interpretation techniques are applied to support the legal reasoning process. First, grammatical interpretation is used to identify the textual meaning of relevant provisions. Second, systematic interpretation is applied by examining the interrelationship between the relevant provisions within the broader Indonesian criminal law system, particularly concerning criminal responsibility, attribution, and corporate liability. Third, teleological interpretation is employed to identify the purposes underlying the relevant legal norms, particularly legal certainty, justice, prevention, and protection against risks arising from AI technology. Fourth, conceptual interpretation is used to clarify and integrate the doctrines of fault-based liability, strict liability, vicarious liability, and risk-based accountability. The comparative analysis of the EU Artificial Intelligence Act 2024 is subsequently used to test and refine the normative coherence of the proposed model and identify principles that may be adapted to the Indonesian legal context. This study's

limitations lie in its focus on the normative dimension of criminal liability; therefore, empirical aspects concerning law enforcement implementation and stakeholder perceptions are beyond its scope and may constitute an agenda for further empirical-sociological research.

RESULTS AND DISCUSSION

Indonesia's Positive Legal Regulations on Artificial Intelligence-Based Crimes

A critical examination of Indonesia's positive criminal law system reveals a fundamental normative paradox: on the one hand, the development of Artificial Intelligence technology has introduced a new category of crime that is qualitatively different from conventional crimes; on the other hand, available criminal law instruments are still built on normative assumptions rooted in the conventional crime paradigm that positions humans as the sole relevant criminal actors. Indonesia's positive legal provisions relevant to AI-related crimes remain dispersed across various legislative instruments, none of which are specifically crafted to address the complex legal relationships arising from autonomous or semi-autonomous AI systems (Nawawi, 2026).

The Criminal Code (UU Nomor 1 Tahun 2023), as a general criminal law instrument, contains several significant conceptual reforms, but has not yet formulated explicit provisions governing Artificial Intelligence as a criminally relevant entity. The construction of accountability in the 2023 Criminal Code still fundamentally rests on the doctrine of individual culpability, which requires free will (*vrije wil*) as a prerequisite for accountability, a prerequisite that AI systems cannot ontologically fulfill. UU Nomor 19 Tahun 2016 concerning Electronic Information and Transactions, although specifically regulating information technology-based crimes, consistently positions technology as a mere instrument of crime, rather than as an actor contributing to the determination of the quality of the criminal act. As a result, accountability is consistently shifted to humans as formal legal subjects, without considering the extent to which the intervention of AI systems has altered the causality and attribution of criminal acts (Rahman & Habibulah, 2019). Meanwhile, UU Nomor 27 Tahun 2022 concerning Personal Data Protection, while relevant in the context of data misuse by AI systems, is more oriented toward protecting victims' rights and does not construct an adequate framework for criminal accountability for actors within the AI ecosystem.

The absence of specific legal provisions becomes even more evident when contrasted with the more advanced regulatory frameworks on Artificial Intelligence adopted in other jurisdictions. The European Union, through the EU Artificial Intelligence Act (EU Regulation 2024/1689), which came into full force in 2026, has adopted a systematic risk-based approach by classifying AI systems into four risk categories: unacceptable risk, high risk, limited risk, and minimal risk. The regulation explicitly requires developers and providers of high-risk AI systems to meet strict standards of transparency, accountability, and a duty of care, and establishes proportionate sanctions for violations of these obligations. This normative comparison confirms that the fragmented and reactive approach that dominates Indonesian positive law is argumentatively inferior to a systematically and proactively developed risk-based regulatory framework. It can be conclusively concluded that the absence of a specific criminal law regime governing AI-based crimes in Indonesia not only creates a legal vacuum that threatens legal certainty but also places Indonesia in a vulnerable position in the face of the escalation of AI-based crimes globally (Hutapea et al., 2025; European Parliament and of the Council, 2024).

Problems of Criminal Liability in Artificial Intelligence-Based Crimes

The fundamental problem in enforcing criminal liability for AI-based crimes stems from the structural dysfunction that arises when the principle of "*geen straf zonder schuld*," which postulates the absence of crime without fault, is confronted with the ontological characteristics of AI systems that operate semi-autonomously. In classical criminal law construction, as developed by Moeljatno, liability requires the fulfillment of a verifiable chain of causality between the actus reus and the mens rea of the legal subject in question. However, in AI-based crimes, systems capable of autonomously generating decisions and actions through machine learning mechanisms effectively break this chain of causality: objectively unlawful acts (actus reus) can no longer be directly and convincingly attributed to the subjective will of any human (Sudarto, 2024).

The deepfake fraud case in East Java in 2025 provides a concrete empirical illustration of this problem. In this case, the perpetrator utilized generative AI technology to produce an audiovisual representation resembling a regional head to manipulate the victim's perception. Constructively, human actors can still be identified and held accountable for fraudulent acts. However, the involvement of AI systems in producing representations that are technically indistinguishable from reality introduces a new layer of complexity into causal analysis: AI, in this case, does not function as a passive tool completely under the perpetrator's control, but rather as a technical actor autonomously determining the quality and effectiveness of the fraud. This situation makes the doctrine of intervening cause relevant to consider: to what extent the autonomous contribution of an AI system can qualify as a causal factor that weakens or breaks the link of responsibility between the human actor and the resulting consequences (Eurike et al., 2024; Fletcher, 1998).

This complexity is further compounded when analyzed within the context of an AI ecosystem involving multiple actors. AI system developers may be unable to prove malicious intent (*dolus malus*) when designing the system, as the system is generally designed for legitimate purposes. Platform providers may not be held directly liable for lack of actual knowledge of the specific misuse that occurs. End users may not fully understand the technical capabilities and legal consequences of the technology they operate. Meanwhile, AI systems, as technical entities, lack the legal capacity to bear criminal responsibility. Consequently, a fault-based liability approach that requires the identification of specific mens rea has the potential to create a liability gap situation where the actual harm suffered by victims cannot be adequately addressed by the existing criminal law system (Ramadhani et al., 2025).

The alternatives available in criminal law doctrine, namely strict liability and vicarious liability, each offer partial contributions and substantial limitations. Strict liability allows for the imposition of responsibility without proving mens rea, effectively closing the gap for impunity in high-risk situations. However, philosophically, it potentially conflicts with the principle of individual justice because it imposes responsibility on parties who may be substantively innocent (Anggraeny et al., 2024). In contrast, vicarious liability allows the distribution of responsibility to parties who have a hierarchical relationship with the direct perpetrator, so that the corporation that develops or provides the AI platform can be burdened with responsibility; However, this approach is limited in its effectiveness to strictly defined relational structures and cannot encompass the complexity of horizontal actor networks in the modern AI ecosystem (Ihsan, 2024). This dual limitation normatively demands the construction of an accountability model that transcends this dichotomy.

Hybrid Criminal Liability Based on Risk-Based Approach in Artificial Intelligence Crimes

Building on the normative diagnosis developed in the previous sub-discussion, this study constructs a hybrid criminal liability model based on a risk-based approach as a normative response to the structural limitations of conventional approaches. This model is built on two basic, foundational assumptions. First, Artificial Intelligence cannot and should not be constructed as an independent subject of criminal law, as it does not meet the ontological prerequisites of legal awareness and free will. Second, the relational complexity within the AI ecosystem demands a proportional distribution of responsibility based on levels of control, rather than a single attribution mechanism to a single party. Its philosophical foundation, as postulated by Radbruch, is that the liability system must simultaneously realize distributive justice, legal certainty, and social benefit in controlling technological risks (Radbruch, 2006; Wijanarko et al., 2026).

Operationally, this hybrid model integrates three liability concepts, operationalized hierarchically and contextually. Strict liability applies to developers of high-risk AI systems proven to have design defects or fail to meet required security standards, on the basis that the party creating the systemic risk is obligated to bear the consequences without relying on proof of malicious intent. Vicarious liability applies to corporations or platform providers that have a legal relationship with the direct perpetrator and are proven negligent in supervising the use of AI systems within their scope of responsibility. Risk-based accountability serves as the primary distributive framework that determines the proportionality of the burden of responsibility based on each party's capacity to control, mitigate, and prevent the risk, not solely based on hierarchical position or causal proximity to the crime.

The applicability of the hybrid model is determined through a set of cumulative normative parameters rather than through the mere involvement of Artificial Intelligence in the commission of a crime. The first parameter is the degree of AI autonomy, which distinguishes AI functioning primarily as a human-controlled tool from AI systems capable of performing operational functions with limited direct human intervention. The second parameter is the level of risk generated by the AI system, particularly the potential magnitude, foreseeability, and severity of harm. The third parameter is the distribution of risk-control capacity, namely the extent to which each actor is technically and legally capable of controlling, preventing, detecting, or mitigating the relevant risk. The fourth parameter is the existence of a corresponding duty of care, which determines the specific preventive or supervisory obligations imposed upon each actor. The fifth parameter is the existence of a breach of such duty or other legally relevant conduct, followed by the establishment of a causal or risk nexus between that conduct and the resulting criminal harm.

These parameters function as a sequential normative test for determining the applicable form of liability. First, the legal analysis must establish whether the AI merely functions as a tool under substantial human control or possesses a degree of operational autonomy that limits direct human intervention and predictability. Second, the level and nature of the risk associated with the AI system must be identified. Third, the actors who possess effective control or substantial capacity to mitigate the relevant risk must be identified. Fourth, the applicable duty of care and the extent to which that duty has been breached must be established. Fifth, the analysis must determine whether the breach, risk creation, or failure of risk mitigation has a sufficient causal or normative risk nexus with the criminal harm. Only after these elements have been established can the appropriate liability mechanism be selected.

The resulting liability is therefore determined proportionally rather than uniformly. Fault-based liability remains applicable to users or other direct actors where intentional or negligent criminal conduct can be established. Strict liability may be applied to developers of

high-risk AI systems where a legally recognized duty of care concerning system design or safety is breached through a proven design defect or failure to satisfy prescribed safety requirements, without requiring proof of malicious intent. Vicarious liability may be applied to platform or service providers where a legally relevant relationship and supervisory or risk-control responsibility exist, and where negligent failure to implement reasonable detection, prevention, or mitigation measures contributes to the criminal harm. Risk-based accountability operates across these mechanisms as the distributive principle for determining the proportionality of responsibility according to autonomy, risk, control capacity, duty of care, breach, and risk nexus. Accordingly, the hybrid character of the model lies not in imposing all forms of liability simultaneously on every actor, but in combining different liability mechanisms within a single risk-based framework according to the legally relevant position and responsibility of each actor.

The operationalization matrix of this model can be seen in Table 1 below:

Table 1. Hybrid Criminal Liability Matrix Based on Risk-Based Approach

Parties	Basis of Liability	Conditions/Scenarios	Legal Mechanisms	Model Components
AI Developers	Strict Liabilities	The AI system being developed poses a high risk to society; there is a proven design flaw (design defect);	Imposition of liability without proving mens rea; based on the duty of care inherent in developers of high-risk systems	Strict Liability + Risk-Based Accountability
AI Platform/Service Providers	Vicarious Liability + Risk-Based Accountability	The platform is proven to be negligent in controlling the risk of misuse; it does not implement adequate detection and prevention mechanisms;	Liability based on the legal relationship with users and risk control capacity; proportional to the platform's level of control	Vicarious Liability + Risk-Based Accountability
AI Users (Direct Actors)	Fault-Based Liability (Primary)	There is evidence of knowingly and maliciously exploiting the AI system to commit a crime (e.g., deepfake fraud);	Full liability based on fault (dolus/culpa); AI acts as a means of strengthening liability, not eliminating it	Fault-Based + Risk-Based Accountability
Corporations (Legal Entities)	Vicarious Liabilities	The corporation benefits from the use of AI by employees/agents, resulting in losses to third parties.	Corporate liability based on the doctrine of respondeat superior; corporate sanctions complement individual sanctions	Vicarious Liability

The relevance of this model is tested through its application to a deepfake fraud case in East Java in 2025. In this scenario: a human perpetrator who knowingly uses AI technology to deceive a victim bears full fault-based liability; the deepfake technology platform provider can be held proportionally liable based on vicarious liability and risk-based accountability if proven negligent in implementing misuse detection and prevention mechanisms; while the AI system developer can be subject to strict liability if the system developed is proven to have design weaknesses that foreseeably facilitate criminal misuse. This model is normatively consistent with the EU AI Act framework, which also distributes obligations in a tiered

manner based on the role and risk-control capacity of each actor in the AI value chain. Thus, this hybrid model is not only normatively responsive to the evolving realities of AI technology but also compatible with the direction of global AI regulation development, which has moved beyond the single-liability paradigm towards an ecosystem-based and proportional accountability system (Ramadhani et al., 2025).

On this basis, the human perpetrator who knowingly uses generative AI to deceive the victim remains subject to fault-based criminal liability because the criminal intention and unlawful objective originate from the user. The platform or service provider may incur vicarious liability within the legally recognized scope of such liability where it has a relevant relationship and control capacity over the use of the system, together with a proven failure to implement reasonable or legally required mechanisms for detecting and preventing foreseeable misuse. The developer may be subject to strict liability only where the applicable legal framework establishes such liability and where a high-risk system is proven to contain a relevant design defect or failure to satisfy prescribed safety requirements that has a sufficient causal or normative risk nexus with the criminal harm. Thus, the model does not automatically impose liability on every actor involved in the AI value chain; rather, each actor's responsibility is determined through the sequential assessment of autonomy, risk, control capacity, duty of care, breach, and risk nexus. This demonstrates the prescriptive character of the proposed hybrid model by transforming the abstract combination of liability concepts into a structured normative process for determining responsibility.

CONCLUSION

This study concludes that Indonesia's existing criminal law framework remains fragmented in addressing AI-based crimes, particularly where AI systems possess a degree of operational autonomy that may reduce direct human control and predictability. The conventional fault-based liability framework remains applicable where AI functions primarily as a tool and the criminal intention and conduct can be attributed to a human actor. However, where AI autonomy, technological risk, and the distribution of control among multiple actors complicate conventional attribution, a risk-based framework becomes necessary to complement conventional criminal liability. The study further finds that strict liability, vicarious liability, and fault-based liability should not be applied uniformly to all actors in the AI ecosystem. Their application must be determined proportionally by assessing the degree of AI autonomy, level of risk, actor's control and risk-mitigation capacity, applicable duty of care, breach of that duty, and the causal or normative risk nexus between the relevant conduct and the criminal harm.

The principal novelty of this study lies in the construction of a prescriptive hybrid criminal liability model that transforms these liability mechanisms into a sequential and operational framework for determining responsibility among AI developers, platform or service providers, users, and corporations. Rather than treating hybrid liability as the simultaneous imposition of multiple forms of liability, the proposed model establishes a structured decision process consisting of autonomy assessment, risk assessment, actor-control assessment, duty-of-care assessment, breach assessment, risk-nexus assessment, and proportionality assessment. Through this construction, risk-based accountability functions as the overarching distributive framework, while fault-based, strict, and vicarious liability operate as actor-specific mechanisms according to the legally relevant conditions. The novelty of this research therefore lies in providing an operational normative mechanism for determining when and how different forms of criminal liability should be applied in AI-based crimes without recognizing AI itself as an independent subject of criminal law.

REFERENCES

- Anggraeny, K. D., Khakim, M., & Sirojudin, M. R. (2024). The urgency of cybercrime law reform in Indonesia: Resolving artificial intelligence criminal liability. JUSTISI.
- Eurike, H., Gresmelian, A., & Afifah, W. (2024). Criminal responsibility of artificial intelligence committing deepfake crimes in Indonesia. *Asian Journal of Social and Humanities*.
- European Parliament and of the Council. (2024). 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*, L 2024/1689.
- Fletcher, G. P. (1998). *Basic Concepts of Criminal Law*. Oxford University Press.
- Haris, M. T. A. R., & Tantimin. (2022). Analisis pertanggungjawaban hukum pidana terhadap pemanfaatan artificial intelligence di Indonesia. *Jurnal Komunikasi Hukum*, 8(1).
- Hutapea, N. M. S., et al. (2024). Artificial intelligence and criminal liability: A preliminary study within the Indonesian legal system. *Jurnal Ilmu Hukum Kyadiren*, 7(2).
- Hutapea, N. M. S., et al. (2025). Artificial intelligence and criminal liability within Indonesian legal system. *Jurnal Ilmu Hukum Kyadiren*, 7(2).
- Hutchinson, T. (2013). *Researching and Writing in Law* (3rd ed.). Thomson Reuters.
- Ihsan, R. M. (2024). Pertanggungjawaban hukum tindak pidana ekonomi melalui artificial intelligence perspektif hukum Islam dan hukum positif Indonesia. *Journal of Islamic*, 8(1).
- Istiqomah, et al. (2025). Kedudukan hukum deepfake AI sebagai subjek hukum dan pertanggungjawaban pidananya berdasarkan teori dualistik. *Jurnal Publikasi Ilmu Hukum*, 3(2).
- Moeljatno. (2008). *Asas-Asas Hukum Pidana*. Rineka Cipta.
- Muladi, & Arief, B. N. (2010). *Teori-Teori dan Kebijakan Pidana*. Alumni.
- Muladi. (1985). *Lembaga Pidana Bersyarat*. Alumni.
- Nawawi, W. (2026). Kedudukan hukum artificial intelligence dan konstruksi pertanggungjawaban pidana. *Jurnal Ilmu Sosial dan Hukum*, 4(2).
- Pamungkas, A. A., Sutarni, N., & Pranawa, B. (2025). Legal analysis of deepfake technology in Indonesia from the perspective of fair and civilized humanity. *Pena Justisia*, 24(1).
- Radbruch, G. (2006). Statutory lawlessness and supra-statutory law. *Oxford Journal of Legal Studies*, 26(1), 1–11.
- Rahardjo, S. (2009). *Hukum Progresif: Sebuah Sintesa Hukum Indonesia*. Genta Publishing.
- Rahman, R. A., & Habibulah, R. (2019). The criminal liability of artificial intelligence in Indonesian criminal system. *Legality Journal*, 27(2).
- Ramadhani, M., Hidayat, R., & Hosnah, A. U. (2025). Pertanggungjawaban pidana atas tindak pidana yang dihasilkan oleh sistem AI otonom. *Jurnal Review Pendidikan dan Pengajaran*.
- Rohmatulloh. (2024). Pertanggungjawaban pidana pasca KUHP baru. *Jurnal Ilmiah Hukum dan Keadilan*.
- Saleh, R. (1983). *Perbuatan Pidana dan Pertanggungjawaban Pidana: Dua Pengertian Dasar dalam Hukum Pidana*. Aksara Baru.
- Setiawan, A. D., & Wijayanto, I. (2024). Tinjauan yuridis pertanggungjawaban pidana dalam tindak pidana yang melibatkan artificial intelligence. *YUSTISI Jurnal Hukum & Islam*, 12(2).
- Soekanto, S., & Mamudji, S. (2007). *Penelitian Hukum Normatif: Suatu Tinjauan Singkat*. Raja Grafindo Persada.

- Sudarto. (2024). The urgency of cybercrime law reform in Indonesia: Resolving artificial intelligence criminal liability. *JUSTISI*, 11(1).
- Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (The 1945 Constitution of the Republic of Indonesia)
- Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana (UU Nomor 1 Tahun 2023 concerning the Criminal Code)
- Undang-Undang Nomor 19 Tahun 2016 tentang Informasi dan Transaksi Elektronik (UU Nomor 19 Tahun 2016 concerning Electronic Information and Transactions).
- Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU Nomor 27 Tahun 2022 concerning Personal Data Protection).
- Wijanarko, S., et al. (2026). Konstruksi pertanggungjawaban pidana korporasi pengembang AI terhadap tindak pidana pemalsuan dokumen elektronik. *Lex Journal*, 10(2).