



AI and Criminal Liability: Theoretical Dilemmas in Applying Criminal Law to Artificial Intelligence

Dr. Danish Bashir Mangi¹, Ihsanullah Butro² & Tahseen Akhtar Memon³

¹Lecturer/In-charge Chairman, Department of Law, University of Sindh, Jamshoro,

Email: Danish.mangi@usindh.edu.pk

²Visiting Faculty Member at Department of Law, University of Sindh, Jamshoro, Email: ahsanbutro212@gmail.com

³Law Student (batch 2k24) at Department of Law, University of Sindh, Jamshoro,

Email: memontahseen744@gmail.com

ARTICLE INFO			ABSTRACT
Article History:			<i>The rapid growth of artificial intelligence (AI) shows undetermined challenges for criminal law, particularly in determining liability for autonomous or semi-autonomous AI actions as traditional legal doctrines based on human intent (Mens Rea) and action (Actus Reus) are not sufficient when applied to Artificial intelligence which lacks consciousness or moral agency, this paper conducts analysis of these pre-existing elements and discusses international and domestic legal frameworks—focusing on Pakistan, the European Union, the United States, and others—to explore how AI-related criminal liability is perceived globally. The study show key gaps in legal areas, liability and enforcement mechanisms, it demonstrate that current legal systems must be reformed or expanded to address the complex legal status of AI for future regulatory development and legal areas.</i>
Received:	April	24, 2025	
Revised:	June	03, 2025	
Accepted:	June	08, 2025	
Available Online:	June	12, 2025	
Keywords:			
Artificial Intelligence, Criminal Liability, AI Personhood, AI Criminal Liability			
Corresponding Author:			
Dr. Danish Bashir Mangi			
Email:			
Danish.mangi@usindh.edu.pk			



Introduction

The regulation of emerging technologies poses considerable challenges across various legal domains, including criminal, civil, and administrative law. As society progresses, contentious debates emerge in areas such as genomics, robotics, and digital technology. While these scientific and technological innovations offer numerous advantages to humanity they also introduce new challenges and risks. We can see that human activities driven by technological advancements have already contributed to severe climate change and environmental degradation (Atif, 2021).

In today's rapidly transforming technological landscape, a previously unfamiliar subject such as AI has quickly found its area of application in everything, including technology and law, and creates

challenging and significant issues. Among these challenges, the most complicated issue is the criminal responsibility for autonomous actions by AI systems (Arora et al., 2024). Artificial intelligence plays a crucial role from self-driving cars to autonomous weapons in the armed forces which can be lethal and can decide their target on their own, analyze various methods of operation almost instantaneously, and can kill people without any human intervention (Jani et al., 2024).

The phrase "Artificial Intelligence" was coined during the prestigious Dartmouth College conference in 1956 by John McCarthy. Today, everyone knows about this term which encompasses a broad field of research that incorporates various techniques and technologies, including theorem proving, heuristic search, game playing, expert systems, neural networks, Bayesian networks, data mining, agents, and, more recently, deep learning (Giannini, A. 2023; Shiwani et al., 2023).

AI is typically classified into three levels: Artificial Narrow Intelligence focuses on specific tasks like voice recognition and chatbots, which are also examples of Narrow AI. Artificial General Intelligence (AGI) refers to a system capable of performing any intellectual task on par with human cognitive abilities; and finally, Artificial Super Intelligence (ASI), where it is believed to far surpass the capabilities of humans. AGI is not yet achieved, and ASI is purely hypothetical, thus also requiring a careful and speculative stance in terms of implications (Abdelaziz, D.K.A., 2025). Current AI systems should not be regarded as conscious entities, as they do not possess self-awareness, subjective experiences, or an understanding of their own existence. Instead, AI operates based on pre-programmed instructions provided by humans and functions through a mechanism of input and output.

Although the legal and criminal liability of AI might appear confusing, particularly some actions taken solely by AI or in conjunction with humans, it is poised to emerge as a significant legal issue in the near future given the ongoing advancements in technology and the increasing interaction between AI and humans. In this regard, the possible criminal accountability of AI, which is created and programmed by humans, should not be overlooked (Kan, C.H., 2024).

Law defines both the entitlements and duties of individuals in society. To follow the legal norms is thus to fulfill duties and to be accorded rights, providing these same things to AI in shape of legal personhood is consequently a question of whether AIs should have rights and obligations in alignment with the law, the solution may be forward-looking and innovative, yet a sufficient analysis should involve a preliminary examination about legal personhood for AIs, as it would make the AIs responsible for their own actions (Chaudhary, 2024).

This paper aims to explore both national and international legal frameworks to assess the accountability of artificial intelligence and to investigate whether current liability doctrines applicable to humans can be extended to AI, as we know these AI systems and bots are ultimately created or developed by humans.

Research Questions

- a. How can artificial intelligence be held accountable under criminal law frameworks?
- b. What are the key challenges in assigning criminal liability for acts committed through or by AI systems?
- c. How do international legal instruments address the issue of AI and criminal responsibility?
- d. What is the current status of national laws in regulating AI-related crimes, and how can they be improved?

Research Methodology and Framework

This research employs a doctrinal methodology to analyze and evaluate the legal framework related to artificial intelligence (AI) and criminal liability, both at the international and national levels. This approach allows for a comprehensive examination of statutory law, case law, and legal literature, helping to identify the deficiencies and obstacles in holding AI systems or their developers legally responsible when crimes occur through or by AI.

The research is structured into two key parts. The first part investigates international legal instruments, such as the Tallinn Manual, UN reports, and relevant academic debates, to understand how the global legal system is addressing AI-related criminal acts. The second part concentrates on national laws of different states including Pakistan, United States, China, Indonesia and Hungary

Literature Review

In order to hold AI accountable for any criminal actions, it must first be recognized as a legal person as this necessitates a re-evaluation of the jurisprudential definition of personhood which can expose some Key legal questions regarding whether AI should be granted citizenship, what fundamental, statutory, and legal rights would be afforded to it, and what obligations it would have. In this context, the European Parliament, in 2017, urged the European Commission to develop a legislative framework addressing the 'Civil Liability of Robots', which proposed the concept of electronic personhood for robots to facilitate the imposition of tort liability. Although this was an initial suggestion, it encountered significant resistance from various stakeholder. Over 150 AI experts signed an open letter to the European Commission, advocating against the conferral of legal personhood to robots on both ethical and legal grounds (Lagioia et al., 2019; Pagallo, 2013; Solum, 2021).

This viewpoint is grounded in the belief that humans, as superior and dominant beings, are the sole entities capable of possessing moral personhood. The ethical personhood perspective contends that only humans, with their advanced consciousness, reasoning abilities, programming skills, natural intelligence, emotions, and physical capabilities, are eligible for personhood (Wolfgang & Ansay, 2022; Kan C.H., 2024).

Certain experts argue that weak AI merits recognition as a legal entity endowed with rights. Nonetheless, weak AI lacks subjective intent akin to human consciousness and therefore cannot be held criminally accountable. Even if it causes harm, it doesn't have the intention or carelessness needed for crime. Its actions are based on human input and training, not its own will. Therefore, harmful acts by weak AI are seen as accidents or unforeseeable circumstances, not crimes (Na J and Xin T., 2025).

AI Related Incidents

The incorporation of AI into contemporary society is rapidly advancing which offers significant benefits across various fields. However, this progress also presents the potential for harm Furthermore AI-related incidents include situations where these systems malfunction, generate biased outputs, or result in unintended consequences, adversely affecting individuals and society. These incidents can vary from minor accidents and inconveniences to serious safety threats and privacy infringements. The ramifications of AI-related incidents can be intricate and multifaceted (Dharm et al., 2024).

There can be two aspects to discuss the incidents related to AI, either on individual basis or societal level.

The individual person can face physical harm by AI through malfunction in autonomous vehicles or robots controlled by AI, financial loss can also be an example due to biased algorithmic decision in loan approvals or insurance premiums, the most trending feature of AI which is of generating content which can be offensive and create either psychological distress or poor mental health.

On Societal level, repeated incidents such as accidents of autonomous vehicles can weaken public trust in AI. Apart from this such features of AI which causes misinformation or damages reputation by AI-generated deep fakes can lead to ethical dilemmas throughout the society domain.

Determination of Criminal Liability

Criminal law functions to mediate interpersonal conflicts and ensure the stability of societal structures (Bayraktar, 1977; Kan C.H., 2024), an individual's criminal responsibility is triggered upon committing and offense another, thereby subjecting them to the legal framework of that jurisdiction. When it comes to artificial intelligence, crimes committed through its use may not fit the traditional definition of a crime, as they are executed by software or robots that are fundamentally separated from the human creators of that software, program, or robot. Consequently, to assess the criminal liability for actions carried out by artificial intelligence, it is crucial to determine whether AI itself qualifies as a legal entity. Additionally, we must address the key issues that arise in identifying and establishing the Actus Reus and Mens Rea—namely, the physical act and the mental (intent) element—which are essential for determining the commission of a crime (United Nations, 1998; Hallevy, 2016; Kumar, P., 2023).

Guilty Act (Actus Reus)

The problem of establishing causality between the AI system and the harmful event arises because this type of explanation is linked to the possibility for humans to fully control a certain event. The Reason is: even if we know what input we gave to the AI and what output it produced, we might not understand how it made that decision or why it did it that way, As in many cases, it is unclear which specific element of the input led to the generated outcome. This is especially a problem in AI-related crimes because the connection between a human's action and the AI's harmful result can be very ambiguous and complicated (Giannini, A., 2023).

This study identifies three primary factors contributing to failures in establishing: (1) the problem of many hands, (2) the black box problem, and (3) shortcuts. These factors are introduced briefly here, with further analysis presented in subsequent sections.

AI operations typically involve combinations of actions by numerous individuals, whether it is the user, developer or the programmer, consequently, when harms occur it become difficult to identify the individual cause, and therefore the liable individual, behind the even.

On the other hand, the black box phenomenon is more complex, as it encompasses calculations or actions determined by the core of AI that are difficult to analyze, such as Deep Neural Networks (DNNs), which address problems in an opaque manner. This can be defined as "an inability to fully comprehend an AI's decision-making process and the inability to predict the AI's decisions or outputs." This implies that it may be impossible for the designer or user of the program to anticipate AI behavior once the system is deployed, as well as to reconstruct the reasons behind its specific actions after the fact.

The last but not the least A shortcut in machine learning, especially in self-supervised learning, happens when the AI model uses an easy clue (like green grass) instead of truly understanding the data. This leads to inaccurate results, even though the output may seem correct. These shortcuts are hard to detect. In legal cases, this raises a problem: how can a judge assign responsibility if the AI's decision-making path can't be clearly traced or explained (Giannini, A. (2023).

Guilty Mind (Mens Rea)

However, AI systems lack *mens rea* since they are not sentient, do not experience emotions, and do not have a moral framework. While more advanced AI can imitate a reasoning process, such responses are the product of algorithmic programming rather than genuine intent. For example, autonomous vehicles do not 'choose' to run a red light or injure a pedestrian in the way a human driver might; instead, their behavior results from input data and real-time processing. The capacity of AI to generate *actus reus* without *mens rea* introduces a significant complication when trying to apply conventional principles of criminal liability (Arora et al., 2024; Jani et al., 2024).

Proving intent is difficult because intent involves an internal "state of mind." To assess whether an AI system possesses such a state, attention must be directed toward its internal framework, particularly how it functions. This includes evaluating whether the AI holds internal cognitive states like beliefs, and motivational states such as goals, desires, or intentions—and whether it processes these states to adapt its behavior (Sartor, 2009). However, the existence of black-box systems like deep neural networks (DNNs) makes it extremely challenging to access or interpret the decision-making processes within the AI, unless the system discloses that data itself or the developer is able to extract it (Lagioia et al., 2019).

Strict Liability

Another approach to holding AI accountable is through strict liability, where the individual or company whose AI has executed an action resulting in harm—such as an autonomous vehicle striking a pedestrian after misinterpreting a traffic signal due to a software glitch—can be held responsible. However, the development of a single AI system often involves thousands of researchers, programmers, and developers who may have dedicated several years to research and development. Consequently, attributing criminal liability to these individuals for the actions of AI becomes exceedingly difficult. Determining the specific contribution of each person is a significant challenge. According to criminal law, punishment should be proportional to the crime, but it is nearly impossible to ascertain the proportionate contribution of these individuals (Jani et al., 2024).

Criminal Liability for Artificial Super Intelligence

Scholarly opinions regarding criminal liability for artificial super intelligence (ASI) are diverse. One viewpoint argues for the preservation of traditional legal principles, confining liability to humans—such as programmers or operators—since ASI lacks consciousness and will, and therefore cannot be held criminally accountable (Chalmers et al., 2023). Others advocate for shared liability, suggesting that responsibility should be distributed between the ASI system and its creators or operators (Rubinstein, 2022). An alternative perspective posits that highly advanced ASI systems should be recognized as legal entities capable of independent criminal liability, given their ability to make complex decisions autonomously (Smith, 2022). Lastly, some scholars support the notion of regulatory or social accountability, placing responsibility on the organizations that own and control ASI systems, with a focus on implementing stringent regulations (Jones and Walker, 2023).

Specified Modes to held AI Liable

There are three possible scenarios in which AI can be used to commit a crime or where AI itself commits the crime. Since Actus Reus and Mens Rea, which are primary requirements to hold anyone liable, face opposition when applied to AI, to address these misunderstandings, there are three situations in which AI can be prosecuted to determine who is liable: the user, the producer, or AI itself (Hallevy, 2016;Jani et al., 2024;Nirma University Law Journal 2019).

When AI is acting as innocent agent

In the first possible scenario, the AI entity is regarded as an innocent agent operating under the instructions of the user. In this context, criminal liability may emerge from the intentional programming by the producer to facilitate an offense, or from the misuse of the AI entity by the user to perpetrate the crime.

Fictive illustration for the first case: A programmer designed malfunctioned software for traffic signals which confuses the drivers and pedestrian resulting in various accidents. The software commits the offense, but the programmer is deemed to be the perpetrator.

Fictive illustration for the second case: The user buys a robot and instructs it to kill a third person. Here, the robot does not apply its intelligence or experience and simply follows the user's command.

In the first case, only the producer would be liable. In the second case, only the end user would be liable because the robot is a mere innocent intermediary.

When AI is acting as semi-innocent agent

The second possible scenario revolves around the foreseeability of the producer/programmer or end user concerning the potential commission of offenses. In this situation, both the producer and the user engage closely with the AI entity, even if they did not intend for a specific offense to occur. Criminal liability may arise in two ways: first, from the negligence or recklessness of the producer in programming the AI entity, and second, as a natural and probable consequence of the actions directed by the user.

Fictive illustration for the first case: B launches an AI-powered drone for deliveries. Due to a programming flaw, it crashes into a crowd, causing injuries. The fault lies in the software code formatted by the producer or manufacturer.

Fictive illustration for the second case: A buys a specific robot and instructs it to hack a drone. In an attempt to hack a drone, the robot hacked a AI powered Car which result in accident and killed a person and property therein. Although A did not intend to hack the or kill anyone, such outcomes can be seen as natural and probable consequences of his actions, which he could have Reasonably foreseen.

In the first case, the producer would be liable. In the second case, the end user would be held liable. In the second case, A may not be liable for murder but for the offense of negligence.

When AI is acting an independent entity/fully autonomous

The third scenario envisions a future where AI entities operate in a fully independent and autonomous manner, transcending mere reliance on algorithms to learn from their experiences and

observations. In this context, such an AI entity would exhibit cognitive capabilities, allowing it to select from various potential solutions to a given problem. Should this advanced AI entity engage in criminal activity, it could be held criminally liable for its actions.

General Defenses

General defenses available to humans under criminal law include intoxication, necessity, and self-defense. These defenses may also be adapted for AI entities, albeit in slightly modified forms. For example, while intoxication can affect human behavior, a malware or virus could similarly impact a robot's functioning. Therefore, with some adjustments, the general defenses applicable to humans can be applied to AI entities as well.

Kinds of Punishments

Common sanctions under criminal law include the death penalty, imprisonment, and fines. These forms of punishment can be adapted for AI entities with certain modifications. For example, permanently deleting the software of an AI entity could be likened to the death penalty for humans. In a similar vein, temporarily deleting the software might be equated with imprisonment for human offenders. Additionally, community service could represent a comparable penalty for AI entities.

International Frameworks

Tallinn Manual 2.0

The Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations is a comprehensive academic guide that explains how existing international law applies to cyber activities, including those involving autonomous systems like artificial intelligence (AI). Though not legally binding, it reflects expert consensus and serves as a valuable reference point (Schmitt, 2017).

According to Rule 2, a state has full authority over all entities linked with cyber activities, which includes AI systems operating within its territory. This supports the argument that a state must regulate and monitor AI to prevent cybercrimes. However, when such cyber activity affects another state, international law becomes applicable, holding the originating state accountable for cross-border harm. If AI systems cause serious adverse consequences abroad, international responsibility arises, even without an explicit duty to cooperate unless a treaty requires it. Rule 14 reinforces this by confirming that a state will bear international responsibility for cyber operations attributable to it—even when AI is involved. This attribution can include state organs under Rule 15, highlighting direct state liability when public institutions develop or use harmful AI systems. Rule 31 adds that when AI is owned by a private organization, that organization can also be held liable, expanding criminal responsibility beyond the state to private actors. Under Rule 34, if an AI commits actions that violate basic human rights, the owner—whether a state or corporation—can still be held responsible. Furthermore, Rule 68 explains that using AI in cyber operations involving force is unlawful, and Rule 70 extends this prohibition to threats. Rule 80 adds that if AI is deployed in armed conflict, it must comply with international humanitarian law, including key principles like distinction and proportionality. This introduces a framework to evaluate the legality of AI actions in warfare contexts. Rule 84 makes it clear that individual—whether users or producers of AI—can be investigated or prosecuted for criminal actions committed through AI, including by direct commission or joint action. Rule 98 explicitly prohibits the use of AI in cyber operations that amount to terrorism and Rule 103 clarifies that cyber attacks may involve AI-driven systems just as much as human-operated ones. Collectively, these rules provide strong

support for recognizing AI-related conduct under international criminal liability and justify both state and individual responsibility.

UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)

Artificial Intelligence (AI) can be held responsible when its actions or results cause harm, especially if these harms come from design defects, misuse, or negligence during its creation or deployment. According to the UNESCO Recommendations, liability may be assigned either to users or developers depending on the specific situation—such as when an AI system causes damage because of a biased algorithm, lack of transparency, or failure to maintain accountability. In these instances, developers could face legal consequences for failing to implement proper safeguards or conduct thorough testing, while users might be held accountable if they deliberately misuse the system to cause harm. The principle of accountability emphasizes that responsibility must be clear, traceable, and enforceable, making both the creators and operators of AI potentially liable under the relevant laws.

European (EU) Artificial Intelligence Act, 2021

The AI Regulation sets out a detailed legal framework for managing artificial intelligence within the European Union, aiming to ensure that AI respects fundamental rights and supports democratic values. Section 1 stresses the need for a consistent legal standard throughout the EU to keep up with the fast pace of AI development and its cross-border effects. Section 5 underlines the EU's dedication to creating a human-centered, trustworthy AI environment that reflects European values and fundamental rights. Section 28 deals with categorizing AI systems by risk level, imposing stricter rules for high-risk AI, such as conformity checks and proper data management. Section 45 focuses on transparency requirements for AI systems that interact with people or produce content, making sure users are clearly informed. Finally, Section 69 calls for establishing a European Artificial Intelligence Office to oversee supervision and enforcement across Member States. Together, these sections highlight the EU's balanced strategy to encourage innovation while protecting human rights and ethical principles. (European Commission, 2021)

National Frameworks

Pakistan

The Prevention of Electronic Crimes Act (PECA) 2016, along with its 2025 amendment, forms the cornerstone of Pakistan's legislative framework addressing digital offences, with indirect applicability to crimes involving artificial intelligence (AI). Although PECA does not specifically mention AI, its provisions are broad enough to cover AI-generated acts such as unauthorized access, cyber terrorism, identity theft, and the spread of fake or harmful content. In such instances, legal liability may extend to the human developers, users, or controllers of AI systems. The Federal Investigation Agency (FIA), established under the FIA Act 1974 and governed through the FIA Rules of 1975 and 2002, is authorized to investigate and prosecute cybercrimes, including those involving AI. These laws collectively empower Pakistani authorities to hold individuals accountable when AI technologies are misused to commit offences, especially when intent or negligence can be traced back to a human agent. However, this legal framework remains foundational but incomplete for addressing AI and criminal liability, underscoring the need for more explicit legislation in this area (PECA, 2016/2025; Federal Investigation Agency Act, 1974; FIA Rules, 1975 & 2002).

The National AI Policy Consultation Draft V1 of Pakistan is highly relevant to discussions on AI and criminal liability, particularly regarding the creation of a national framework to govern the ethical and legal aspects of artificial intelligence. The policy stresses principles like transparency, accountability, and safety of AI systems, which are essential when assigning liability for AI-driven actions. Its strategic focus on governance and ethics advocates for establishing legal frameworks and accountability mechanisms necessary for handling criminal liability in cases involving autonomous or semi-autonomous AI systems. By promoting a human-centric and responsible approach to AI deployment, the policy lays the foundation for national debates on adapting or extending existing criminal laws to address emerging AI technologies (Ministry of Information Technology and Telecommunication, 2023).

Indonesia

The Indonesian legal system currently lacks explicit provisions for holding artificial intelligence (AI), such as autonomous vehicles, criminally liable. According to Ali et al. (2013), Indonesia's penal code is traditionally anthropocentric, meaning it assumes only human beings can commit crimes. Although the concept of corporate criminal liability has been introduced, extending this principle to non-human actors like AI remains legally unsettled.

The article highlights the difficulty of applying traditional criminal law concepts like *Mens Rea* (guilty mind) and *Actus Reus* (guilty act) to AI systems. Since AI lacks consciousness and moral agency, attributing criminal responsibility is highly problematic. Indonesia might consider alternative legal models such as strict liability or assigning liability to developers, manufacturers, or operators of AI systems. However, these approaches have yet to be codified into law. The authors call for legal reforms to fill this gap, proposing a new legal framework that would recognize AI as a legal subject in limited contexts and potentially hold it criminally liable under certain circumstances (Ali, Putra Mulya, & Permana, 2023).

Hungary

Under Hungarian law, the situation is similar in that current legal norms do not recognize AI, including self-driving vehicles, as entities that can bear criminal liability. The Hungarian Criminal Code requires a human perpetrator with criminal intent (*Mens Rea*) or negligence (*Actus Reus*). As noted in the uploaded paper, criminal acts committed by or through autonomous vehicles are currently dealt with by assigning responsibility to humans—usually the vehicle's user, manufacturer, or software developer—depending on the circumstances.

The Hungarian system adheres to the traditional legal doctrine requiring *Mens Rea*, which AI cannot possess. Therefore, liability is redirected to natural persons or legal entities associated with the AI system. The discussion in the Hungarian context reflects a growing awareness of AI's challenges but highlights that substantive legal reform is necessary to accommodate AI-driven technologies, especially in areas like traffic accidents involving autonomous vehicles.

United States

The document is relevant to understanding how the United States approaches the issue of AI and criminal liability through its traditional legal framework. It notes that the U.S. does not yet have specific criminal legislation tailored to AI systems. Instead, liability is assessed based on pre-existing legal doctrines, primarily focusing on the accountability of human actors or corporations that design, control or deploy AI technologies.

The key concept of *Mens Rea* (guilty mind) remains central to establishing criminal responsibility, meaning that AI itself is not considered a legal subject capable of bearing criminal intent. This approach underscores a cautious legal stance that resists assigning direct liability to AI, opting instead to scrutinize the intent and actions of those behind its operation (Abdelaziz, 2025).

China

In the context of China, the document emphasizes a legal and regulatory structure that is primarily state-controlled and precautionary. China's approach does not aim to assign criminal accountability to AI systems, but instead concentrates on strict regulation of individuals, developers, and companies involved in AI technologies. The Chinese government underscores authority over AI's application in socially sensitive fields, supporting adherence and supervision mechanisms that avert harm before it takes place.

Criminal responsibility is attributed to the human operators or organizations that misuse AI, reflecting a collective and practical legal framework. This strategy highlights China's focus on societal order and governmental control over speculative debates regarding AI independence in legal contexts (Abdelaziz, 2025).

Research Gap

Despite increasing academic attention on the intersection of artificial intelligence and criminal law, a notable gap persists in thoroughly examining how current legal principles—particularly those related to criminal liability—can evolve to accommodate AI-driven actions. Most existing legal systems still rely on concepts such as *Mens Rea* and *Actus Reus*, which assume human consciousness and intent, making them challenging to apply to fully autonomous or semi-autonomous systems. While international guidelines like the Tallinn Manual and national regulations such as Pakistan's PECA 2016 offer some direction, they remain vague and inconsistent in addressing whether AI itself, or its developers and users, should bear responsibility. Additionally, the question of whether AI should be granted legal personhood is still largely speculative and unresolved, with few concrete frameworks proposed for differentiating between culpable human actions and independent AI conduct.

This research seeks to fill these gaps by providing a comparative analysis of both international and domestic perspectives, evaluating whether existing legal doctrines are adequate or if new legal frameworks are necessary to hold AI systems criminally liable.

Findings

(i) Incompatibility with Traditional Legal Concepts: Core legal principles such as *Mens Rea* (guilty mind) and *Actus Reus* (guilty act) are not directly applicable to AI systems, as these entities lack consciousness, intent, and moral agency. This creates a legal vacuum in determining culpability when AI causes harm.

(ii) Absence of Legal Personhood for AI: Most jurisdictions, including Pakistan, do not recognize AI as a legal person, making it difficult to assign direct liability to AI entities. The debate on granting AI legal status remains largely theoretical, with significant ethical and practical opposition.

(iii) Fragmented International Approaches: International instruments like the Tallinn Manual and UNESCO's AI ethics guidelines offer some direction, but they lack binding authority and vary

in scope and applicability. There is no unified international legal framework governing AI criminal responsibility.

(iv) National Legal Gaps: Pakistan's PECA 2016 and other national laws provide indirect applicability to AI-related offenses but are silent on autonomous AI actions. This lack of specific legislation hinders the effective attribution of criminal liability.

(v) Comparative Legal Perspectives: Jurisdictions such as the European Union and the United States are experimenting with risk-based and regulatory approaches but still center liability on human agents (developers, users, corporations), not the AI itself.

(vi) Emerging Theoretical Models: Legal scholars propose new models such as strict liability for developers, shared liability between AI and humans, and even independent liability for highly autonomous AI. However, these models are still in the conceptual phase and not reflected in formal legislation.

Discussion

The results of this study indicate that current legal frameworks—both internationally and domestically—are insufficient to address the complexities of assigning criminal liability to artificial intelligence systems, the analysis reveals that while concepts like strict liability or vicarious liability offer partial solutions, they often fall short in situations where AI acts independently or unpredictably, particularly in cases involving black-box technologies such as deep neural networks. This underscores the difficulty of applying traditional legal principles like *Mens Rea* and *Actus Reus* to non-human entities that lack consciousness, intent, or emotional awareness, furthermore, a comparative analysis with jurisdictions like the EU, China, and the United States shows that, while there are efforts to regulate AI's impact, no unified legal framework for criminal responsibility exists, for example, the EU's AI Act focuses on risk management and transparency but refrains from granting AI legal personhood, similarly, Pakistan's PECA 2016 and National AI Policy remain silent on autonomous liability, concentrating instead on human misuse.

These discrepancies highlight the urgent need for legal reform, whether by extending existing laws or creating new categories of liability that consider the autonomous decision-making capabilities of advanced AI systems, therefore, the findings stress the necessity for both theoretical and legislative developments to ensure accountability in a future where AI takes on increasingly autonomous roles in society.

Recommendations

- i. Develop specific legal definitions and classifications for different types of AI (e.g., weak AI, AGI, ASI) to guide the application of liability frameworks.
- ii. Introduce new legislation that explicitly addresses AI-related offenses, rather than relying solely on existing cyber or criminal laws.
- iii. Establish criteria under which AI systems could be granted limited legal personhood for accountability purposes in specific contexts.
- iv. Create standardized international agreements that harmonize AI criminal liability principles across borders.
- v. Implement strict liability schemes for developers and manufacturers of AI systems in high-risk sectors such as healthcare, finance, and autonomous vehicles.

- vi. Mandate transparency in AI design to help determine causation and responsibility in legal cases.
- vii. Require rigorous ethical and legal testing of AI systems before deployment, especially for systems involved in public decision-making or autonomous actions.
- viii. Encourage collaboration between legal experts, technologists, and ethicists to draft adaptable laws that reflect AI's evolving capabilities.
- ix. Insert such types of features which can help in determine the Reason of harmful event occurred by Artificial intelligence.
- x. Educate law enforcement, legal professionals, and policymakers on AI technologies to better understand and enforce AI liability frameworks.

Conclusion

This research underscores the increasing urgency to revise criminal liability frameworks in response to the growing autonomy and societal integration of artificial intelligence (AI), while traditional legal principles, such as *Mens Rea* and *Actus Reus*, are effective in human contexts, they fall short when applied to entities that lack consciousness, intent, or moral agency. The study shows that although international tools like the Tallinn Manual and the EU's AI Act attempt to regulate AI's actions, they do not provide clear mechanisms for holding AI itself criminally responsible, on a national level, countries like Pakistan, along with others analyzed in this study, continue to adhere to anthropocentric liability models, typically assigning accountability to developers or users, but leaving a significant gap when it comes to AI's independent actions.

Moreover, the lack of consensus on whether AI should be granted legal personhood continues to impede significant legal reform, and the findings suggest that without a shift in both legal philosophy and legislative policy, the justice system will struggle to address crimes committed by or through AI. A cohesive, future-proof legal approach—potentially combining human accountability with modified ideas of electronic personhood or regulatory responsibility—is crucial, and until such reforms are introduced, legal systems worldwide will face challenges in ensuring justice and accountability in the era of autonomous technologies.

References

1. Abdelaziz, D. K. A. (2025). A comparative analysis of legislation and international conventions regarding criminal liability for AI misuse and crimes. *Journal of Infrastructure, Policy and Development*, 9(1), 10722.
2. Arora, T., & Thakur, S. (2024). Criminal liability of artificial intelligence: A comprehensive analysis of legal issues and emerging challenges. *International Journal of Research Publication and Reviews*, 5(11), 1886–1891.
3. Atif, A. (2021). Challenges in defining legal aspects of criminal liability for counter-productive actions of machine learning and artificial intelligence. *International Journal of Instructional Technology and Educational Studies (IJITES)*, 2(3). ISSN (Print): 2682-3918, ISSN (Online): 2682-3926.
4. Bayraktar, K. (1977). Suçişlemegetahrikcürmü.FormülMatbaası.
5. Chalmers, R., Liddy, J. & Mason, T. (2023). The legal challenges of autonomous intelligence systems, *Oxford Journal of Legal Studies*, 44(3), 673–699.
6. Chaudhary, G. (2020). The liability paradox. *ILI Law Review*, Summer Issue.
7. Dharm, J., Girme, A., & Gharde, U. (2024). Challenges in criminal and civil liability. *International Journal of Law*, 10(2), 52–57.

8. Giannini, A. (2023). Criminal behavior and accountability of artificial intelligence systems [Doctoral thesis, Maastricht University & University of Florence].Eleven Publishers.
9. Hallevy, G. (2016). The criminal liability of artificial intelligence entities – From science fiction to legal social control. *Akron Law Journals*, March 2016.
10. Hodula, M. (2023).Criminal liability of self-driving vehicles under Hungarian criminal law. *State and Law: Legal and Political Sciences Quarterly*, 66(2).
11. Jani, C. S., &Rathor, S. P. (2024).A legal framework for determining the criminal liability and punishment for artificial intelligence.Tuijin Jishu / Journal of Propulsion Technology.
12. Jones, P., & Walker, T. (2023).Ethics and regulation of autonomous systems.Springer.
13. Kan, C. H. (2024). An evaluation in the context of the general theory of crime and fundamental principles. *International Journal of Eurasia Social Sciences (IJOESS)*.
14. Kumar, P. (n.d.). Determination of civil and criminal liability of artificial intelligence.National University of Study and Research in Law, Ranchi, India.
15. Lagioia, F., & Sartor, G. (2020). A legal analysis and a regulatory perspective. *Philosophy & Technology*, 33, 433–465.
16. Ministry of Information Technology and Telecommunication. (2023). National Artificial Intelligence Policy – Consultation Draft V1. Government of Pakistan.
17. Na, J., &Xin, T. (2025).The possibility and limitation of legal subject status of artificial intelligence from the perspective of criminal law.*International Journal of Forensic Sciences*, 10(1).
18. Nirma University Law Journal, 8(2), July 2019.
19. Pagallo, U. (2013). The laws of robots (Vol. 10).Springer.Retrieved January 30, 2024.
20. Padhy, A. K., &Padhy, A. K. (2019).Criminal liability of the artificial intelligence entities.
21. Prevention of Electronic Crimes Act, 2016 (PECA), Pakistan, amended (2025).
22. Rubinstein, D. (2022). Criminal liability for AI-caused harm: Mens Rea and beyond. *British Journal of Criminology*, 62(1), 257–273.
23. Sartor, G. (2009). Cognitive automata and the law: Electronic contracting and the intentionality of software agents. *Artificial Intelligence and Law*, 17, 253.
24. Schmitt, M. N. (Ed.). (2017). Tallinn Manual 2.0 on the international law applicable to cyber operations.Cambridge University Press.
25. Shiwani, &Datta, N. (2023).Research paper on artificial intelligence and criminal justice system.*International Journal of CReative Research Thoughts (IJCRT)*, 11(1). ISSN: 2320-2882.
26. Smith, H. (2022). The economic impact of over-regulation on AI startups in Europe. *European Innovation Review*, 45(1), 78–92.
27. Solum, L. B. (2021).Yapayzekâlarınhukukikişiliği (M. Fincan, Trans.).SeçkinYayıncılık.
28. United Nations.(1988). Rome Statute of the International Criminal Court. The Hague, Netherlands.
29. Wolfgang, F., &Ansay, T. (2022).Tüzelkişiliknazariyelerivetatbikat.