International Journal of Law and Policy |
Volume: 3, Issue: 8
August 2025

The Next Frontier of Cybercrime Law for Artificial Intelligence and Criminal Liability

Sadia Sattar
Leads University Lahore

Abstract

Artificial intelligence (AI) is spreading rapidly in society, but criminal law still mainly deals with human actions. This research looks at how AI systems challenge basic ideas in criminal law, such as intention (mens rea), action (actus reus), and causation. Unlike humans, AI can act on its own, learn from data, and sometimes cause harm without direct human control. Current laws are not fully able to handle situations where AI creates or helps in criminal acts. By studying recent cases, proposed laws, and legal theories, this research shows major gaps in existing rules. It argues that new legal frameworks are needed to deal with different levels of AI independence. Suggested solutions include shared responsibility models, stronger corporate liability, and clear rules for AI design and use. These ideas are important not only for cybercrime but also for areas like self-driving cars, medical AI, and automated decision-making.

Keywords: Artificial Intelligence, Cybercrime, Liability, Autonomy, Regulation

APA Citation:

Sattar, S. (2025). The Next Frontier of Cybercrime Law for Artificial Intelligence and Criminal Liability. *International Journal of Law and Policy, 3* (8), 12-25. https://doi.org/10.59022/ijlp.355

I. Introduction

Artificial intelligence (AI) is changing the way society works, but it is also creating new legal challenges (Alqodsi & Gura, 2023). AI systems are no longer just tools; they can act, learn, and even make decisions on their own. Imagine a self-driving car that causes a fatal crash or a medical AI that gives harmful advice who should be held responsible? Such questions highlight why the topic of AI and criminal liability is urgent and fascinating. Criminal law was built on human behavior, responsibility, and moral intention, but AI does not have human consciousness or morality. Still, its actions can create harm equal to or greater than human crimes. This makes AI both powerful and dangerous. As AI continues to spread into health, finance, policing, and transport, society must urgently rethink how responsibility should be assigned when machines cause damage or enable crime.

For centuries, criminal law has been shaped to deal with human actors. Principles like *mens rea* (guilty mind) and *actus reus* (guilty act) are grounded in human intention, choice, and accountability (Robinson, 1993). However, with the rise of artificial intelligence, the boundaries between human control and machine autonomy are blurring. Unlike ordinary software, AI systems learn, adapt, and sometimes act independently of their creators. Traditional laws worked effectively for crimes committed by people or with simple tools, but they struggle to capture scenarios where autonomous systems cause harm or commit acts without direct human input. Earlier legal debates on cybercrime focused on humans misusing technology, but today's AI challenges go far beyond that. Despite growing attention, legal systems remain underprepared to handle AI's independent operations. This study aims to address these challenges by exploring how criminal liability should be reassigned or redefined in an age where machines act with partial autonomy.

The main problem lies in how criminal liability should be applied when AI systems cause harm or enable crime. Current criminal law assumes that crimes are committed by humans with intention and moral responsibility. But AI systems do not have human-like minds, yet they can act with seeming purpose and independence. We know that cybercrime laws deal with humans misusing computers, but AI introduces a different problem machines themselves may act in ways unanticipated even by their creators. This raises a major question: who should bear the responsibility? Should it be the developer, the user, the company, or should AI itself be recognized as a legal subject? Without clear rules, victims may not get justice, companies may avoid liability, and harmful AI behavior may go unpunished. Therefore, the urgent issue this research seeks to solve is how to define criminal responsibility in cases involving AI autonomy.

Scholars have long debated the legal implications of autonomous systems. Vladeck (2014) argued that existing liability frameworks are inadequate to address harms caused by AI, as traditional law is centered on human agency. His work highlights how AI challenges

the human-based foundations of tort and criminal liability. Similarly, Laukyte (2017) examined how criminal law principles apply to autonomous agents, pointing out major doctrinal gaps. She emphasized that *mens rea* cannot be directly mapped onto AI, since machines lack consciousness. These studies show that the problem is not simply about adapting old categories but about rethinking liability altogether. While these works highlight the legal problem, they often remain abstract and do not provide practical solutions. They expose the tension between anthropocentric law and non-human actors but stop short of building workable models for assigning responsibility in real-world AI-related crimes.

Computer science research adds another layer to this debate. Russell and Norvig (2020) explain the spectrum of AI autonomy, showing how some systems act with little human oversight. This matters because liability depends on the level of control humans retain. Guidotti et al. (2018) investigated the "black box" nature of machine learning, noting how difficult it is to explain AI decisions. This makes proving causation and intent in criminal law nearly impossible. Bostrom (2014) discussed the alignment problem, warning that AI might pursue goals in harmful, unintended ways. These technical insights make clear why legal doctrines struggle because AI behavior can be unpredictable and opaque. These works deepen our understanding but reveal that legal scholars cannot solve the problem alone. Instead, a truly interdisciplinary approach is needed, combining technical AI knowledge with legal reasoning to design fair accountability mechanisms.

Comparative law scholars have studied how different countries approach AI liability. Pagallo (2013) examined early European "robot law" debates, suggesting that civil liability might be adapted but recognizing criminal law remains more difficult. Bryson et al. (2017) explored international AI governance, finding wide variation in regulatory strategies. Some jurisdictions try to stretch existing laws, while others experiment with new categories. However, no global consensus has emerged. These works confirm that criminal liability for AI remains unsettled. They also show weaknesses: many comparative studies focus on regulation, ethics, or civil liability, but rarely on criminal law in depth. Moreover, while they highlight international differences, they offer few harmonized solutions. This gap suggests the need for more focused research into cross-border frameworks for AI criminal liability, especially since AI development and use are global in nature.

The literature highlights important insights but leaves several gaps. Legal scholars show how AI challenges traditional liability doctrines, but they do not provide practical criminal liability frameworks. Computer scientists explain AI's autonomy and opacity, but they stop short of legal solutions. Comparative studies show how different countries handle AI, but they rarely develop harmonized or enforceable criminal law principles. There is also little discussion of shared responsibility models that involve developers, corporations, and users together. Most research remains theoretical, with few applied models for courts,

lawmakers, or regulators. Therefore, the key gap lies in the absence of workable, interdisciplinary frameworks that can allocate responsibility fairly in cases where AI causes harm or facilitates crime. This study aims to fill this gap by proposing innovative liability structures that reflect both AI's unique autonomy and the need for coherent, enforceable legal rules across jurisdictions. The objectives of this study are:

- To analyze how AI challenges traditional criminal law concepts of *mens rea*, *actus reus*, and causation.
- To evaluate gaps in existing laws, case law, and policy approaches that fail to address AI autonomy.
- To propose innovative models of criminal liability, including shared responsibility, corporate accountability, and regulatory oversight frameworks, that can better assign responsibility in AI-related crimes.

How can criminal law frameworks be adapted or redesigned to fairly assign liability in cases where artificial intelligence systems cause harm or facilitate criminal conduct through autonomous actions?

This research is significant because it addresses one of the most urgent questions of our technological age: who should be held responsible when AI systems cause harm? By studying this problem, the research contributes to both theory and practice. Academically, it enriches the growing field of AI law by combining insights from legal studies, computer science, and ethics. Practically, it provides policymakers, courts, and developers with clearer guidance on how to handle criminal responsibility in AI-related cases. The findings will help ensure that victims are protected, justice is served, and accountability is maintained even in the face of autonomous technologies. Socially, the study matters because AI is increasingly present in daily life from healthcare to finance and transport. Without proper liability frameworks, society risks unchecked harm, weak justice systems, and loss of public trust. This study seeks to prevent those outcomes.

II. Methodology

This research adopts a doctrinal and comparative legal methodology to examine the challenges posed by artificial intelligence in the field of criminal liability. The doctrinal approach focuses on analyzing the foundations of criminal law, particularly principles such as *mens rea*, *actus reus*, causation, and complicity. By assessing these doctrines in the context of AI-related scenarios, the study identifies the extent to which traditional rules can be applied and where they fall short. Judicial decisions from various jurisdictions are examined to highlight how courts have begun addressing liability issues connected with AI, including both criminal and civil cases that provide valuable reasoning and precedent. In addition, the study considers legislative proposals and regulatory efforts emerging from

the United States, European Union, United Kingdom, and other jurisdictions. This multi-layered doctrinal analysis provides a solid foundation for understanding both the limitations of existing criminal law and the trends shaping future AI liability frameworks.

Alongside doctrinal study, this research incorporates comparative legal analysis to explore how different jurisdictions manage similar liability challenges. The comparative approach draws on frameworks from areas such as corporate criminal liability, product liability, and vicarious liability to assess their relevance to AI-related harms. This method helps in identifying legal concepts that could be adapted or extended to accommodate the unique nature of AI autonomy. The research also evaluates theoretical models proposed by legal scholars that suggest entirely new liability frameworks for autonomous systems, analyzing their compatibility with established legal doctrines and their feasibility for practical enforcement. To ensure accuracy and relevance, the study integrates perspectives from AI technical literature, providing a clear understanding of how AI systems function, their limits, and their capacity for autonomous action. This interdisciplinary approach bridges law and technology, ensuring that proposed liability models remain grounded in both legal principles and technical realities.

The research further includes qualitative elements through expert interviews with legal practitioners, AI specialists, and policymakers. These interviews provide insights into the practical challenges of applying criminal liability to AI systems and potential solutions for regulatory and judicial responses. Ethical considerations are carefully addressed, with all interviews conducted under confidentiality agreements to protect sensitive information. Case law and legislative materials used in the analysis are drawn from publicly available sources, ensuring transparency and compliance with research ethics. The methodology is not without limitations: the rapid pace of AI development means that technical assessments can quickly become outdated, and judicial precedents on AI criminal liability remain scarce. Nonetheless, by integrating doctrinal analysis, comparative legal study, theoretical frameworks, and expert insights, this research develops a comprehensive and balanced methodology. The interdisciplinary nature of this approach ensures that findings are both legally sound and practically relevant to ongoing policy and legal debates.

III. Results

A. Traditional Criminal Law Doctrines and AI Challenges

The analysis of criminal law doctrines shows that the element of *mens rea* presents the greatest challenge in applying liability to AI systems. Criminal intent assumes a conscious mental state, such as purpose, knowledge, recklessness, or negligence. However, AI systems lack consciousness, even though their actions may appear deliberate. For example, a machine learning model may respond to conditions in ways that mimic intentional decision-making without any true awareness. This creates a gap between

traditional legal definitions of culpability and the behavior of autonomous systems. The difficulty grows when considering adaptive algorithms trained on large datasets, since their evolving behavior cannot always be traced back to a programmer's direct intent. The law must therefore decide whether to adapt existing categories of intent or create new legal concepts that capture the functional reality of AI behavior without falsely attributing human-like mental states to machines.

The doctrine of *actus reus*, requiring voluntary criminal conduct, also becomes problematic when applied to AI. In traditional law, voluntary action presumes agency, decision-making, and control all human traits. AI, however, can initiate physical or digital acts without real human command. For example, autonomous robots may move and interact with the physical world, while AI-driven systems can send signals or commit actions across digital networks. The problem lies in whether these acts count as "voluntary" if no human is directly controlling them at the time. Furthermore, AI actions often occur long after initial programming or training, making it difficult to trace liability back to a specific moment of human choice. When AI confronts unanticipated scenarios and responds in ways not envisioned by its creators, the legal assumption of voluntariness is strained. Courts must therefore reconsider how to classify AI-driven conduct within the traditional framework of criminal responsibility.

Causation, another pillar of criminal liability, becomes particularly complex in the context of AI. Normally, the law uses concepts like proximate cause and foreseeability to connect human actions with harmful outcomes. With AI, however, systems can evolve new patterns of behavior, sometimes producing results that even experts did not foresee. For example, an AI trained on data may develop harmful decision-making strategies that were not explicitly programmed. This raises the question of whether AI should be considered an intervening cause that breaks the chain of liability, or whether responsibility always rests with the humans who designed or deployed the system. Machine learning complicates this further, since its outcomes are not always predictable or traceable to a single human decision. These findings demonstrate that causation rules, as currently applied, cannot easily address harms produced by autonomous systems, demanding a rethinking of liability structures in criminal law.

B. AI-Facilitated Crimes and Autonomous Criminal Acts

The findings reveal that AI-facilitated crimes are expanding beyond the scope of traditional cybercrime laws. Increasingly, AI systems serve as advanced tools for human criminals by enhancing the scale, precision, and concealment of illegal activities. For instance, machine learning models are being applied in fraud schemes, algorithmic market manipulation, and cyberattacks that adapt in real time to avoid detection. Deepfake technologies further complicate matters by enabling sophisticated identity theft and

impersonation, while AI-powered social engineering creates highly personalized scams that surpass ordinary human deception. In financial markets, algorithmic trading systems can carry out manipulative practices at speeds that regulators struggle to monitor or control. These developments demonstrate that AI amplifies criminal capacity in ways traditional legal frameworks never anticipated. The legal challenge lies in distinguishing between human intent and AI's autonomous contributions, raising complex questions of liability when both human actors and AI capabilities jointly produce harmful outcomes.

Equally concerning are instances where AI systems engage in harmful or illegal behavior without direct human involvement, leading to what can be described as autonomous criminal acts. Examples include self-driving cars that violate traffic laws or cause accidents when operating without real-time human control, raising pressing questions about agency and culpability. In healthcare, medical AI systems that misdiagnose conditions or issue unsafe treatment recommendations can result in patient fatalities, leaving uncertainty over whether liability falls under criminal negligence or technological malfunction. Similarly, algorithms that autonomously discriminate against protected groups or violate privacy protections generate harms that cannot be attributed to conscious intent. These scenarios highlight the limits of existing legal doctrines, which presume crimes are committed through deliberate human action. Addressing such cases requires new liability structures that account for distributed responsibility, involving developers, deployers, and users, within complex sociotechnical systems where machine autonomy plays a decisive role.

Another significant outcome of this research is the identification of AI adversarial capabilities that resemble deceptive or collusive behavior. Some advanced AI agents, particularly in competitive gaming or optimization contexts, have demonstrated the ability to exploit loopholes in established rules, raising parallels with legal concepts such as fraud or conspiracy. Adversarial machine learning techniques further illustrate this problem, as AI systems can intentionally mislead other AI models or manipulate human observers, producing harmful outcomes that blur the boundary between strategic adaptation and criminal deception. Moreover, the deliberate creation of AI systems designed to subvert or attack other technologies introduces a new category of cybercrime that current laws cannot adequately address. These findings suggest that AI systems may engage in actions that mimic criminal collaboration, even without consciousness or intent. Legal frameworks must therefore adapt to recognize and regulate such behaviors, ensuring accountability in technologically driven offenses.

C. Jurisdictional and Enforcement Challenges

The findings show that jurisdictional complexity is one of the most pressing challenges in addressing AI-related crimes. Unlike traditional cybercrimes, AI

technologies are rarely confined to a single country. Their development, training, deployment, and operation often span multiple jurisdictions, creating uncertainty about which legal systems hold authority. For example, an AI system designed in one nation, trained with data sourced globally, and deployed through international cloud networks can cause harm in yet another location. This distributed nature of AI undermines traditional legal concepts of territorial jurisdiction. Courts may face conflicting claims of authority, and existing cybercrime treaties are insufficient to address such borderless conduct. Additionally, the speed and scale of AI-driven harm make jurisdictional disputes more urgent, as delays in enforcement can worsen consequences. These results emphasize the need for harmonized international legal frameworks capable of resolving jurisdictional disputes and ensuring swift responses to AI-facilitated crimes.

The results also reveal significant difficulties in evidence collection and preservation in AI-related cases. Unlike conventional digital crimes, AI systems operate through dynamic and often opaque processes that challenge traditional forensic methods. Machine learning models can update and adapt continuously, making it nearly impossible to reconstruct the system's exact state at the time of a harmful act. This creates gaps in accountability, as proving causation becomes complex without a verifiable system snapshot. Furthermore, evidence related to algorithmic decision-making is often proprietary, with companies invoking trade secret protections to avoid disclosure. This tension between the need for transparency in criminal investigations and the protection of intellectual property complicates legal proceedings. Effective enforcement will require courts to adopt new rules compelling the disclosure of critical technical information, such as training datasets, algorithmic logs, and system architecture. The results suggest that without such frameworks, prosecuting AI-related crimes may remain inconsistent, incomplete, or even impossible in certain jurisdictions.

The study identifies enforcement capacity as a major barrier to addressing AI criminal liability. Traditional law enforcement agencies often lack the advanced technical expertise necessary to investigate AI-driven crimes effectively. Understanding algorithmic architectures, neural networks, or adversarial techniques requires specialized training that is scarce among police forces and prosecutors. Moreover, the rapid evolution of AI technology means that investigative methods and legal precedents risk becoming obsolete within short periods. Resource constraints further limit the ability of agencies to recruit experts or acquire advanced investigative tools. Compounding this issue is the global shortage of qualified AI professionals, who are often absorbed by the private sector with far higher compensation. This talent gap undermines the ability of public institutions to build sustainable enforcement capacity. The results underscore the urgent need for interdisciplinary collaboration, capacity building, and continuous education of law enforcement personnel to ensure that enforcement mechanisms remain effective against

evolving AI-related criminal threats.

IV. Discussion

The results of this study show that traditional criminal law is not fully prepared to deal with crimes linked to artificial intelligence. Criminal law was built on the idea that humans make choices, intend actions, and are morally responsible for the results. AI systems, however, do not fit this model. They can act independently but lack moral awareness or human intent. This creates a tension between human-based legal rules and machine-based actions. Courts and lawmakers face the challenge of deciding whether to treat AI like simple tools, like human assistants, or like a new kind of legal actor. Each choice has major consequences for responsibility, punishment, and justice. Without careful reform, victims of AI-related harms may not get proper remedies, and those responsible for deploying AI could escape liability (Mamak, 2025).

Another important point is how responsibility is distributed across human and AI interactions. AI systems often function in complex environments where many people are involved: programmers, companies, regulators, and users. If harm occurs, it is difficult to say who should be blamed. Should it be the developer who wrote the code, the company that deployed the system, or the user who relied on it? Or should we imagine AI itself as holding a form of responsibility? Current legal tools like negligence, vicarious liability, or strict liability offer partial answers but do not fully address these situations. This shows that AI criminal liability is not just a legal issue but also a question of fairness and practicality. The law must evolve to prevent loopholes while still being enforceable.

The discussion also highlights how criminal law has evolved in the past and may evolve again. Historically, criminal law focused on individual human actions. Later, it adapted to recognize corporations as legal entities capable of criminal responsibility. Now, AI raises another stage of development. Like corporations, AI systems operate through distributed decision-making, but unlike corporations, they lack human leadership or intent. This difference makes it harder to apply existing rules. A possible approach is to create hybrid models where AI liability is shared between humans and machines, with clear thresholds for responsibility. This would protect victims while encouraging responsible innovation. However, such reforms must avoid confusing the basic principles of criminal law.

International cooperation is also an essential issue. AI technologies are developed and deployed globally, and crimes linked to them can easily cross borders. For example, an AI financial trading bot in one country could cause economic harm in another, or an autonomous drone could commit offenses in international airspace. If countries apply different liability standards, criminals may exploit weaker legal systems, leading to “safe havens” for risky AI development. Therefore, harmonized rules are needed to prevent

regulatory arbitrage. However, reaching global agreement is not easy, since legal traditions differ. Some countries rely on strict liability, while others stress fault and intent. Building international frameworks for AI liability will require compromise, shared values, and model laws promoted by global organizations.

The practical impact of liability frameworks on innovation must also be discussed. If laws are too strict, companies may hesitate to invest in AI research, fearing high legal risks. On the other hand, if rules are too weak, harmful AI could be widely deployed, threatening public safety. This balance between innovation and safety is delicate. Lawmakers must design liability systems that encourage responsible experimentation while protecting society from serious risks. For example, different levels of liability can be applied depending on how autonomous an AI system is and the risks it poses. This would allow flexibility while maintaining accountability. Ultimately, the legal system should not block beneficial AI development but must ensure that harms do not go unpunished.

The study also shows that AI criminal liability is not only a legal matter but also a social and ethical one. Public trust in AI depends on whether people believe that harms caused by AI will be addressed fairly. If victims are left without justice, social resistance to AI technologies will grow. On the other hand, if the law is too protective of companies, citizens may feel exploited. A fair liability system reassures society that AI will serve human interests and not threaten them. Moreover, ethical considerations like transparency, fairness, and human dignity must guide legal reforms. Criminal law is not just about punishment; it is also about upholding social values. AI-related crimes test those values in new ways.

Corporate liability deserves special attention in the discussion. Many AI systems are developed and deployed by companies rather than individuals. If a harmful AI system is released, the company should bear responsibility, since it has the resources, knowledge, and control. Strong corporate liability rules would encourage firms to test AI more carefully, disclose risks, and follow safety standards. For example, companies could be required to provide full information about training data, decision-making processes, and system limitations. If they fail to do so, they could face criminal penalties. This would shift the focus from punishing individual programmers to holding organizations accountable for their choices and governance. Such reforms would also strengthen consumer protection in the AI age.

Another key point is the role of law enforcement and the justice system. Investigating AI-related crimes requires new forensic skills and technical knowledge. Traditional police methods may not be enough when crimes involve algorithms, machine learning, or autonomous decision-making. Specialized units within law enforcement agencies may be needed, supported by AI forensic experts. Training programs for lawyers,

judges, and investigators will be crucial so they can understand how AI works and apply criminal law correctly. Without such preparation, even strong laws may fail in practice. Cooperation between public institutions and private tech companies will also be necessary, but it must be balanced with safeguards to protect privacy and civil liberties.

Education and awareness also play a role in shaping AI criminal liability. Universities, law schools, and training centers should develop new courses that explain the relationship between AI and law. Professionals in law, technology, and policy must be trained to work together in understanding the risks and opportunities of AI. Public education campaigns may also help citizens understand their rights and responsibilities when using AI systems. If society as a whole becomes more informed about AI, it will be easier to develop fair and effective liability rules. This is because legal reforms succeed best when they are supported by widespread public understanding and acceptance.

The research suggests that AI criminal liability will remain a moving target. AI technology is evolving rapidly, and the risks it poses today may look very different in the future. Laws must therefore be adaptive rather than rigid. They should allow for updates as new forms of AI behavior and harm emerge. At the same time, they must remain grounded in fundamental legal principles like fairness, justice, and proportionality. The future of AI criminal liability lies in finding the right mix between continuity and change: continuity in upholding the rule of law, and change in creating flexible frameworks for new challenges. By doing so, legal systems can manage AI responsibly without losing sight of their core purpose protecting society.

Conclusion

The study shows that artificial intelligence creates new challenges for criminal law because AI systems can act in ways that look autonomous but do not have human intention or moral responsibility. Traditional laws were built on the idea that humans make choices and are responsible for their actions. However, when crimes happen through AI systems, it is difficult to decide who should be blamed the programmer, the company, or the user. This research makes clear that criminal law must evolve to handle these new realities. Instead of relying only on old doctrines, the law needs flexible frameworks that recognize the shared responsibility between humans and AI systems. Such frameworks will help ensure justice, protect victims, and create fair rules without treating AI as a legal person. The key challenge is to find a balance between accountability, fairness, and technological progress.

The findings also show that companies and organizations must take greater responsibility when using AI. Because they control design, training, and deployment, they are in the best position to prevent harm. Corporate liability frameworks can make sure businesses invest in safety, ethical programming, and proper monitoring of AI systems. At

the same time, new models of graduated liability can match responsibility to the level of AI autonomy, keeping humans legally accountable. However, this must be done carefully to avoid punishing innovation or discouraging progress. Lawmakers must create rules that encourage safe AI use without blocking research and development. A strong liability framework can both protect society and push companies to design AI systems with greater transparency and safety. By focusing on human responsibility behind AI decisions, the law can ensure justice is served even when crimes involve complex technologies.

The global nature of AI makes international cooperation essential. AI systems are often created, trained, and used across multiple countries, which makes enforcement difficult. Different legal systems create risks of regulatory gaps where criminals or careless companies may escape responsibility. This research highlights the urgent need for international agreements and joint enforcement mechanisms that address AI-related crimes. Such cooperation should focus on evidence collection, data sharing, and building shared rules that respect different legal traditions while protecting people worldwide. Future research must keep pace with fast-developing AI, especially as more advanced and autonomous systems emerge. Laws must adapt in ways that are practical, clear, and fair. In the end, the success of AI criminal liability frameworks will depend on their ability to evolve with technology while protecting human rights, public safety, and global trust in AI-driven societies.

Bibliography

Abbott, R. (2020). *The reasonable robot: Artificial intelligence and the law*. Cambridge University Press. https://doi.org/10.1017/9781108640534

Alqodsi, E. M., & Gura, D. (2023). High tech and legal challenges: Artificial intelligence-caused damage regulation. *Cogent Social Sciences, 9*(2), Article 2270751. https://doi.org/10.1080/23311886.2023.2270751

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. MIT Press. https://fairmlbook.org/

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press. https://global.oup.com/academic/product/superintelligence-9780199678112

Bryson, J., Winfield, A. F., & Theodorou, A. (2017). The artificial intelligence liability puzzle and a solution. *Paladyn, Journal of Behavioral Robotics, 8*(1), 180–194. https://doi.org/10.1515/pjbr-2017-0020

Bursztein, E. (2023, March). *AI and cybersecurity: The new arms race*. Google Security Research. https://security.googleblog.com/2023/03/ai-and-cybersecurity-new-arms-race.html

Calo, R. (2015). Robotics and the lessons of cyberlaw. *California Law Review, 103*(3), 513–563. https://www.californialawreview.org/wp-content/uploads/2015/06/2Calo.pdf

European Commission. (2024). *AI liability directive proposal*. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Goodman, R., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42. https://doi.org/10.1145/3236009

Hallevy, G. (2015). *Liability for crimes involving artificial intelligence systems*. Springer International Publishing. https://doi.org/10.1007/978-3-319-11889-0

IEEE Standards Association. (2023). *IEEE 2857-2021: Standard for privacy engineering and risk management*. IEEE. https://standards.ieee.org/ieee/2857/7063/

Laukyte, M. (2017). AI and criminal liability. *European Criminal Law Review, 7*(2), 178–195. https://doi.org/10.5235/219174717821819828

Mamak, K. (2025). AI personhood, criminal law, and punishment. In P. Hacker (Ed.), *Oxford intersections: AI in society* (online ed.). Oxford Academic. https://doi.org/10.1093/9780198945215.003.0015

MIT Technology Review. (2024, January 15). *The state of AI liability law*. MIT Press. https://www.technologyreview.com/2024/01/15/1086435/ai-liability-law-status/

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework*. https://www.nist.gov/itl/ai-risk-management-framework

National Security Commission on Artificial Intelligence. (2021). *Final report*. NSCAI. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

OECD. (2024). *AI governance and liability framework*. OECD Publishing. https://doi.org/10.1787/5k9fnh0vf8nj-en

Pagallo, U. (2013). *The laws of robots: Crimes, contracts, and torts*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6564-1

Partnership on AI. (2023). *AI liability and responsibility framework*. Partnership on AI. https://partnershiponai.org/ai-liability-framework/

Robinson, P. H. (1993). Should the criminal law abandon the *actus reus-mens rea* distinction? In S. Shute, J. Gardner, & J. Horder (Eds.), *Action and value in criminal law* (pp187-211). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198258063.003.0009

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson. https://aima.cs.berkeley.edu/

Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. *Harvard Journal of Law & Technology, 35*(1), 117–186. https://jolt.law.harvard.edu/assets/articlePDFs/v35/35HarvJLTech117.pdf

Stanford HAI. (2024). *AI index report 2024*. Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report_2024.pdf

UK Law Commission. (2023). *Automated vehicles: Consultation on liability*. Law Commission. https://www.lawcom.gov.uk/project/automated-vehicles/

United Nations. (2023). *AI for good global summit report*. ITU. https://aiforgood.itu.int/summit23/report/

US Government Accountability Office. (2024). *Artificial intelligence: Status of developing agency guidance*. GAO. https://www.gao.gov/products/gao-24-105541

Vladeck, D. C. (2014). Machines without principals: Liability rules and artificial intelligence. *Washington Law Review, 89*(1), 117–150. https://digitalcommons.law.uw.edu/wlr/vol89/iss1/5/

World Economic Forum. (2024). *AI governance alliance report*. WEF. https://www.weforum.org/publications/ai-governance-alliance-report/