

Criminal Liability for Autonomous AI-Generated Crimes: Accountability Challenges in Pakistan's Criminal Justice System

Osama Muhammad¹,

Abstract

The proliferation of autonomous artificial intelligence (AI) systems ranging from self-driving vehicles and algorithmic trading bots to autonomous weapons and generative AI models has introduced a category of harmful conduct that does not fit comfortably within the doctrinal architecture of criminal law. Criminal liability has traditionally rested on two pillars: a wrongful act (*actus reus*) and a culpable mental state (*mens rea*), both of which presuppose a human or corporate agent capable of intention, knowledge, or recklessness. Autonomous AI systems complicate this picture because their outputs may not be foreseeable even to their designers, and the causal chain linking human conduct to the eventual harm may be attenuated by layers of machine learning, self-modification, and emergent behaviour. This article examines how Pakistan's criminal justice system anchored in the Pakistan Penal Code 1860, the Prevention of Electronic Crimes Act 2016, and general principles of criminal responsibility inherited from the common law is positioned to respond to AI-generated crimes. It surveys competing theories of AI liability developed in comparative scholarship (*respondeat superior* analogies, natural-probable-consequence liability, direct liability, and proposals for a novel legal status for AI), and applies them to Pakistan's statutory and institutional context. The article argues that Pakistan's existing framework is structurally unable to attribute *mens rea* to autonomous systems, that corporate and vicarious liability doctrines offer only partial solutions, and that investigative and forensic capacity within the police and judiciary lags behind the technical complexity of AI systems. It concludes with a set of legislative, institutional, and procedural reforms aimed at closing the accountability gap without stifling innovation.

Article History

Received 15 April 2026
Revised 25 June 2026
Accepted 24 June 2026
Published 30 June 2026

Corresponding Author*
OPEN ACCESS

Keywords

artificial intelligence, criminal liability, autonomous systems, mens rea, Pakistan Penal Code, Prevention of Electronic Crimes Act, algorithmic accountability, cybercrime

1. Introduction

Artificial intelligence systems are no longer confined to narrow, human-supervised tasks. Machine learning models now drive vehicles, execute financial trades, moderate online content, generate synthetic media, and operate industrial machinery with minimal real-time human oversight. As the autonomy of these systems increases, so does the possibility that their operation will produce outcomes that, had they been caused by a human being, would readily be classified as criminal offences: causing death through a defective autonomous vehicle's decision-making, generating child sexual abuse material through an unmoderated image-generation model, executing fraudulent financial transactions through an autonomous trading algorithm, or producing defamatory or incendiary deepfake content that incites violence (Mangi et al., 2025; Gul et al., 2025).

Pakistan is not insulated from this trend. The country has seen rapid uptake of AI-enabled financial technology, a growing deepfake-driven disinformation problem around elections and public figures,

¹ LL.M. osama56640@gmail.com

increasing reliance on automated content moderation and surveillance tools by both state and private actors, and early-stage deployment of autonomous and semi-autonomous systems in logistics, security, and agriculture. Pakistan's criminal justice system, however, was built for a world in which a criminal act could always be traced to a identifiable human mind possessing intention, knowledge, or recklessness. This article asks a deceptively simple question with no simple answer: when an autonomous AI system causes a harm that would be a crime if committed by a human, who if anyone is criminally liable under Pakistani law, and is the existing framework adequate to answer that question with legal certainty? The article proceeds in six parts. Part 2 reviews the conceptual and comparative literature on AI criminal liability. Part 3 maps the doctrinal foundations of Pakistani criminal law relevant to this inquiry. Part 4 analyses the specific accountability gaps that arise when autonomous AI is the proximate cause of harm. Part 5 draws on comparative approaches to assess reform options suited to Pakistan's legal and institutional context. Part 6 offers recommendations and concludes (Pandey et al., 2026; Gul et al., 2025).

2. Research Methodology

Research methodology refers to the systematic methods and procedures adopted to investigate the research problem and achieve the objectives of the study. It explains how relevant information is identified, selected, collected, organized, processed, and analysed, as well as the rationale for choosing particular research methods and techniques. The methodology ensures that the research is conducted in a structured, reliable, and academically sound manner. Depending on the nature of the study, it may involve doctrinal, qualitative, quantitative, comparative, or empirical methods, along with the use of primary and secondary sources to obtain and critically evaluate relevant information.

3. The Doctrinal Foundations of Pakistani Criminal Law

3.1 The Pakistan Penal Code, 1860

Pakistan's substantive criminal law remains codified principally in the Pakistan Penal Code, 1860 (PPC), a colonial-era statute drafted on classical common-law principles of individual culpability. The PPC's offence definitions are built around a human 'person' who acts with a specified mental state intention, knowledge, or in some provisions, negligence or rashness. Sections dealing with common intention and common object (analogous to joint enterprise doctrines) allow liability to be distributed among multiple human participants in a crime, but they do not contemplate a non-human participant. The PPC's general exceptions—covering infancy, unsoundness of mind, and accident were designed to excuse human actors whose capacity for culpable choice was impaired; they offer no ready analogue for an AI system that 'acts' without any capacity for choice in the legal sense at all. Because the PPC has no concept of a non-human offender, any prosecution arising from AI-caused harm must be framed against a human or corporate defendant: a programmer, a deploying company, an operator, or a supervising official. This is workable where the harm can be traced to an identifiable act of negligence, recklessness, or intentional misuse by such a person for example, a company that deliberately trains a generative model to produce defamatory deepfakes, or an operator who disables safety interlocks on autonomous machinery. It is far less workable where the harm arises from the AI's autonomous, statistically emergent behaviour that no individual foresaw or could reasonably have been expected to foresee (Abbas et al., 2024; Ahmed et al., 2025).

3.2 The Prevention of Electronic Crimes Act, 2016

The Prevention of Electronic Crimes Act, 2016 (PECA) is Pakistan's principal cyber-specific statute and the closest existing analogue to an 'AI law.' PECA criminalises unauthorised access to information systems, electronic fraud, cyberstalking, the dissemination of offensive or defamatory content, and related conduct, and it grants investigative powers to the Federal Investigation Agency (FIA) through its Cybercrime Wing (the National Response Centre for Cyber Crime). PECA was amended in 2025 to expand its scope over disinformation and to create the Social Media Protection and Regulatory Authority, reflecting growing legislative attention to online harms, including those amplified by algorithmic systems. However, PECA, like the PPC, is drafted around identifiable human actors—'whoever' commits an act with intent or knowledge and does not address the distinct evidentiary and attributional problems posed by autonomous systems whose outputs are not directly authored, in the traditional sense, by any single human being (Osman et al., 2025 Ahmed et al., 2025).

3.3 Corporate Criminal Liability in Pakistan

Pakistani law recognises corporate criminal liability in a limited and unevenly developed form. Regulatory statutes (in sectors such as banking, securities, and anti-money laundering) impose liability on companies and, frequently, on their 'directors, officers, or persons in charge,' but general criminal law has been slower to develop a robust doctrine of corporate mens rea comparable to the identification doctrine or the broader 'failure to prevent' offences found in some other common-law jurisdictions. Where an autonomous AI system is developed and deployed by a corporate entity, prosecutors are often left to prove fault at the level of an identifiable natural person within the company (a director, a lead engineer, a compliance officer) rather than attributing an aggregate organisational failure such as inadequate testing protocols, insufficient human oversight mechanisms, or a corporate culture that deprioritised safety directly to the company as an offence in itself (Sani, 2025; Gul et al., 2025).

3.4 Constitutional and Procedural Constraints

Article 4 and Article 9 of the Constitution of Pakistan, 1973 guarantee due process and protection of life and liberty, while Article 25 guarantees equality before law. Criminal procedure under the Code of Criminal Procedure, 1898 (as amended) and the Qanun-e-Shahadat Order, 1984 (Pakistan's law of evidence) requires that guilt be established through admissible evidence linking a specific accused to a specific criminal act and mental state. These procedural safeguards, while essential to due process, create acute difficulties in AI-related prosecutions: forensic reconstruction of an AI system's decision pathway may require expertise in machine learning that is scarce within Pakistan's police, prosecution, and judicial cadres, and the evidentiary chain connecting a human defendant's conduct to an AI's autonomous output may not meet the standard of proof beyond reasonable doubt without specialised technical testimony that the system presently struggles to produce or evaluate (Hyder et al., 2025; Malik et al., 2025).

4. Accountability Challenges in AI-Generated Crime

4.1 The Mens Rea Gap

The central doctrinal obstacle is the *mens rea* gap: Pakistani offences generally require proof of intention, knowledge, or at minimum rashness or negligence on the part of a human defendant. Where an autonomous AI system produces an unforeseen harmful output—for instance, a large language model generating instructions that facilitate a crime, or an autonomous vehicle's perception system misclassifying a pedestrian in a manner its designers did not anticipate and could not have tested for given the near-infinite variability of real-world conditions—it becomes difficult to locate the requisite mental state in any single human actor. Prosecutors are left with two unsatisfying options: stretch the concept of recklessness or negligence to cover the general risk of deploying imperfect AI (which risks over-criminalising socially beneficial innovation and chilling investment), or decline to prosecute at all, leaving victims without recourse in the criminal justice system though civil remedies in tort or contract may remain available (Shaikh et al., 2023; Malik et al., 2024).

4.2 Causation and the Multiple-Hands Problem

AI systems are typically the product of many hands: data engineers who curate training data, model developers, third-party vendors who supply foundation models, integrators who fine-tune or deploy the system for a specific use case, and end users who operate it. Harm may result from an interaction among several of these contributions—flawed training data supplied by one party, an inadequately tested fine-tuning process by another, and careless deployment by a third—none of which, viewed in isolation, would satisfy the causation and fault requirements for a criminal conviction. Pakistani criminal procedure, like most common-law systems, is oriented toward identifying a discrete causal chain from a specific defendant's act to a specific harm; the distributed, networked nature of AI development strains this model considerably. A particularly acute manifestation of these challenges in Pakistan concerns generative AI misuse: AI-generated deepfake pornography and non-consensual intimate imagery, synthetic audio used in financial fraud (voice-cloning scams), and AI-generated disinformation deployed around elections. PECA's provisions on cyberstalking, defamation, and the dissemination of offensive material can, in principle, be applied where a human actor knowingly generates and disseminates such content. The harder cases arise where an AI platform's own content-generation defaults or content-moderation failures rather than a specific user's deliberate misuse—are the proximate cause of harm, raising questions about platform and developer liability that current Pakistani law does not clearly resolve (Ibrahim et al., 2026; Gul et al., 2025).

5. Comparative Perspectives and Lessons for Pakistan

Jurisdictions grappling with these questions have adopted varied strategies, several of which offer instructive models for Pakistan, while others illustrate pitfalls to avoid.

5.1 The European Union

The European Union has pursued a risk-based regulatory approach through the AI Act, which classifies AI systems by risk tier and imposes graduated obligations on providers and deployers, including mandatory risk management, human oversight, and record-keeping for high-risk systems. While the AI Act is primarily a regulatory rather than a criminal statute, its emphasis on documented human oversight and traceability creates an evidentiary infrastructure—logs, risk assessments, conformity documentation—that would materially assist criminal investigators in later

reconstructing fault if harm occurs. Pakistan currently has no equivalent regulatory baseline requiring AI developers or deployers to maintain such records. Both jurisdictions have generally resisted creating AI-specific criminal offences, instead relying on existing fault-based doctrines applied flexibly to new fact patterns—for example, treating a company's failure to implement adequate safety testing before deploying an autonomous vehicle as evidence of corporate negligence or, in egregious cases, gross negligence manslaughter. The United States has also seen sector-specific regulatory action (for example, by transportation and aviation safety regulators) operating alongside, rather than replacing, general criminal law. This approach avoids the definitional difficulties of drafting bespoke 'AI crime' offences, but depends on prosecutors and courts having sufficient technical fluency to apply general doctrines to novel facts—a dependency that mirrors Pakistan's capacity constraints (Warraich & Butt, 2024;).

5.2 China and Algorithmic Accountability

China has moved toward direct regulatory obligations on generative AI providers, including content-labelling, security assessment, and provider registration requirements, backed by administrative and, in serious cases, criminal sanctions for providers who fail to prevent the generation of illegal content. This model demonstrates that criminal accountability can be anchored in provider-level compliance failures rather than solely in the autonomous act itself, offering a template Pakistan could adapt without requiring a wholesale reconceptualisation of mens rea doctrine (Kuang et al., 2025).

5.3 Lessons for Pakistan

Three lessons emerge. First, ex ante regulatory obligations (mandatory testing, logging, human-oversight requirements) generate the evidentiary record that ex post criminal accountability depends upon; Pakistan's absence of any AI-specific regulatory statute leaves investigators with little to work from when harm occurs. Second, jurisdictions that have avoided drafting free-standing 'AI offences' have instead strengthened doctrines of corporate and management-level fault, a path more compatible with incremental legislative reform than with the more radical proposals (such as AI legal personhood) that remain unimplemented anywhere in the world. Third, technical capacity-building within investigative and judicial institutions is a precondition for any of these models to function in practice, regardless of which liability theory is adopted (Wassan et al., 2023).

6. Recommendations for Reform

6.1 Legislative Reforms

- Enact a dedicated AI governance statute, or amend PECA, to impose mandatory risk assessment, testing, logging, and human-oversight obligations on developers and deployers of high-risk autonomous AI systems operating in or targeting Pakistan, creating the evidentiary trail necessary for later criminal or civil accountability.
- Introduce a statutory offence of reckless or grossly negligent deployment of an autonomous AI system causing death, serious injury, or significant financial harm, calibrated to capture culpable failures of testing, oversight, or safety design without criminalising ordinary technological risk-taking or good-faith innovation.

- Strengthen corporate criminal liability doctrine, moving toward a 'failure to prevent' model under which a company can be held criminally liable where it fails to implement reasonable procedures to prevent AI-facilitated offences committed in connection with its business, shifting the evidentiary burden away from proving individual director-level fault alone.
- Amend PECA's extraterritoriality and mutual legal assistance provisions to facilitate cooperation with foreign jurisdictions and technology companies whose AI systems cause harm within Pakistan, and pursue bilateral and multilateral data-sharing arrangements with major AI-developing jurisdictions.

6.2 Institutional and Procedural Reforms

- Establish a specialised AI and digital forensics unit within the FIA Cybercrime Wing, with dedicated training in machine-learning model auditing, log analysis, and adversarial data forensics.
- Create a panel of court-appointed technical assessors or a specialised cyber/AI bench within relevant High Courts to assist judges in evaluating complex algorithmic evidence, consistent with practices in specialised commercial and banking courts already present in Pakistan's judicial architecture.
- Develop prosecutorial guidelines, issued jointly by the FIA and provincial prosecution services, clarifying charging standards for AI-related offences and the evidentiary thresholds required to establish corporate or individual fault.

6.3 Preventive and Regulatory Measures

- Mandate pre-deployment risk assessments and post-deployment incident reporting for high-risk AI systems (autonomous vehicles, medical AI, financial algorithms, content-generation platforms with significant Pakistani user bases), administered by a designated regulatory authority.
- Require clear disclosure and labelling of AI-generated content, particularly synthetic media capable of being mistaken for authentic human-generated content, to support both preventive deterrence and evidentiary attribution in prosecutions involving deepfakes and disinformation.
- Encourage mandatory liability insurance for operators of high-risk autonomous systems, ensuring victims retain a route to compensation even where criminal prosecution proves unsuccessful or where liability is diffused among multiple actors.

7. Conclusion

Autonomous AI systems expose a structural fault line in criminal law's foundational assumption that culpability can always be traced to an identifiable human mind. Pakistan's criminal justice system, built on the Pakistan Penal Code 1860 and supplemented by the Prevention of Electronic Crimes Act 2016, retains the doctrinal tools to address AI-facilitated crime only where a human actor's fault can be clearly established; it offers little guidance where harm emerges from the genuinely autonomous, statistically emergent behaviour of a machine-learning system. This accountability gap is not unique to Pakistan, but Pakistan's comparatively limited regulatory infrastructure, nascent corporate liability doctrine, and constrained investigative and judicial technical capacity make the gap more acute than in jurisdictions that have begun to legislate AI-specific obligations. Closing this gap does not require inventing legal personhood for machines; it

requires the more modest, achievable steps of building an evidentiary infrastructure through ex ante regulatory obligations, sharpening doctrines of corporate and management fault, and investing in the technical capacity of investigators, prosecutors, and courts. Absent such reform, Pakistan risks a criminal justice system that is either unable to hold anyone accountable for genuinely novel AI-caused harms, or one that responds to that inability by attributing fault unpredictably and unevenly to whichever human actor happens to be within reach—an outcome that serves neither victims, nor the rule of law, nor the legitimate development of AI innovation in Pakistan.

References

- Abbas, T. N. A., Hameed, R., Kadhim, A. A., & Qasim, N. H. (2024). Artificial intelligence and criminal liability: exploring the legal implications of ai-enabled crimes. *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, (22), 140-159.
- Ahmed, F. A., Gul, S., & Shahzad, S. (2025). Ensuring accountability and transparency in AI-driven corporate governance. *International Journal of Social Sciences Bulletin*, 3(5), 330-341.
- Ahmed, F. A., Zafar, S., & Gul, S. (2025). Analyzing PECA amendments: Press freedom, democratic values, and digital regulation in Pakistan. *Traditional Journal of Law and Social Sciences*, 4(01), 41-51.
- Asaro, P., 'A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics', in *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2012.
- Chesterman, S., *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*, Cambridge University Press, 2021.
- Code of Criminal Procedure, 1898.
- Constitution of the Islamic Republic of Pakistan, 1973.
- European Parliament, Report with Recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL), 2017.
- Federal Investigation Agency, Pakistan, National Response Centre for Cyber Crime, Annual Reports (various years).
- Gul, S., & Ahmad, R. (2025). Consent and Coercion: Examining Contractual Autonomy in Islamic Jurisprudence and Anglo-American Law. *Pakistan Journal Of Law, Analysis And Wisdom*, 4(5), 01-11.
- Gul, S., Ahmad, R., & Rahman, S. U. (2025). Constitutional Dualities: Reconciling Islamic Normativity with Common Law Principles in Hybrid Legal Systems. *Indus Journal of Social Sciences*, 3(2), 674-693.
- Gul, S., Ahmad, R., & Rahman, S. U. (2025). The myth of neutrality: Judicial review, ideology, and constitutional interpretation in Pakistan and the United Kingdom. *The Critical Review of Social Sciences Studies*, 3(2), 1742-1754.

- Gul, S., Malik, W., & Qureshi, G. M. (2025). Cybersecurity And Sovereignty: The Role Of International Law In Governing State Behaviour In Cyberspace. *Policy Journal of Social Science Review*, 3(5), 121-135.
- Gul, S., Saman, A., & Ahmad, F. (2025). The Convergence of Human Rights and Humanitarian Law: Recalibrating Legal Boundaries in Contemporary Conflict. *The Critical Review of Social Sciences Studies*, 3(2), 1391-1403.
- Hallevy, G., 'The Criminal Liability of Artificial Intelligence Entities', *Akron Intellectual Property Journal*, Vol. 4, 2010.
- Hallevy, G., *When Robots Kill: Artificial Intelligence Under Criminal Law*, Northeastern University Press, 2013.
- Hyder, G., Rehman, T. U., & Saeed, A. (2025). Predictive Policing and the Rule of Law: Evaluating AI's Role in Criminal Justice through Global and Pakistani Legal Frameworks. *The Critical Review of Social Sciences Studies*, 3(4), 1119-1127.
- Ibrahim, A., Gourari, F. M., AbdelAziz, G. M., & Hashish, A. (2026). Artificial intelligence in criminal justice governance: Opportunities and legal challenges (A comparative study). *Corporate Law & Governance Review*, 8(2), 77.
- Kuang, Y., Zheng, J., Sun, M., & Xu, X. (2025). The role of algorithmic fairness, accountability, and transparency in shaping user satisfaction on digital platforms: A quantitative study in China. *International Journal of Human-Computer Interaction*, 1-17.
- Malik, W. Gul, S. (2024). Cyber conflict and international security: Legal challenges and strategic solutions in cyberspace.
- Malik, W., Gul, S., & Qureshi, G. M. (2025). Regulating artificial intelligence: Challenges for data protection and privacy in developing nations. *Journal of Social Signs Review*, 3(05), 95-108.
- Mangi, D. B., Butro, I., & Memon, T. A. (2025). AI and criminal liability: Theoretical dilemmas in applying criminal law to artificial intelligence. *The Critical Review of Social Sciences Studies*, 3(2), 2174-2186.
- Osman, Y. H. M., John, D. J., Ahmed, D. A., Mohamed, R. M. A., & Ali, R. M. S. A. (2025). Criminal Responsibility in the Age of AI: Rethinking Legal Accountability. *African Journal of Law and Justice System*, 4(3), 179.
- Pakistan Penal Code, 1860 (Act XLV of 1860).
- Pandey, S., Chauhan, A. S., & Baghel, B. S. (2026). Artificial Intelligence, Intellectual Property and Criminal Liability: Exploring the role of AI in technology-driven crimes and the legal challenges of accountability. *International Journal of Special Education*, 41(11s), 1456-1469.
- Prevention of Electronic Crimes Act, 2016 (Act No. XL of 2016), as amended.
- Qanun-e-Shahadat Order, 1984 (P.O. No. 10 of 1984).
- Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

Sani, A. J. (2025). Artificial Intelligence and the Criminal Justice System: An Analysis of Liabilities and Consequences. *Legal Studies in Digital Age*, 1-14.

Selbst, A. D. and Barocas, S., 'The Intuitive Appeal of Explainable Machines', *Fordham Law Review*, Vol. 87, 2018.

Shaikh, M. A., Rahman, S. U., Tahir, M., Khuharo, K. H., & Shahid, M. (2023). FIXING CRIMINAL LIABILITY AS PER ELEMENTS OF A CRIME: A REVIEW IN MODERN ERA OF AI AND ROBOTICS. *Russian Law Journal*, 11(4), 1183-1207.

Warraich, K. M., & Butt, J. S. (2024). Revealing the economic consequences of prosecution failures within Pakistan's criminal justice systems: A focus on ethical considerations, with comparative insights from Norway. *Minhaj International Journal of Economics and Organization Sciences*, 4(1), 1-23.

Wassan, R., Bhatti, Z. A., Ahmed, S., Oad, S. S., & Detho, S. H. (2023). Implementing community policing model for crime reduction: lessons for Pakistan. *Journal of Social Sciences Review*, 3(1), 178-189.