Artificial intelligence: Challenges in criminal and civil liability

Jyoti Dharm[1], Anuradha Girme[2], Utpal Gharde[3]
[1] Vice Principal, Assistant Professor, Department of Law, New Law College, Bharati Vidyapeeth (Deemed to Be) University, Pune, Maharashtra, India
[2] Assistant Professor, Department of Law, New Law College, Bharati Vidyapeeth (Deemed to Be) University, Pune, Maharashtra, India
[3] Department of Law, New Law College, Bharati Vidyapeeth (Deemed to Be) University, Pune; LLB, ILS Law College, Pune, IIT Kharagpur, Maharashtra, India

Abstract
As AI continue to become an important part of our daily lives, not only does it bring undeniable advantages but also raises pressing questions: who should be held accountable when AI malfunctions and causes harm? This piece explores the crucial issue, revealing the shortcomings of current legal structures in attributing criminal or civil responsibility for AI-related incidents. The analysis unpacks the hurdles presented by opaque AI systems, the lack of malicious intent, and the dispersed accountability among developers, users, and other involved parties. It emphasizes the immediate need for reform and suggests a way forward by advocating for Explainable AI (XAI) advancements, robust strategies to mitigate algorithmic bias, and the establishment of clear legal frameworks governing AI liability. This timely examination throws light on a pivotal moment in AI's evolution, urging policymakers, researchers, and industry leaders to envision a future where AI thrives alongside ethical standards and safeguards for those affected.

Keywords: AI liability, criminal liability, civil liability, AI crimes, liability attribution, regulatory AI framework, legal framework for AI, victims of AI, ethics, governance

Introduction
Artificial Intelligence (AI) is integrating with various aspects of our lives [1]. From self-driving vehicles navigating city streets to medical algorithms assisting in complex diagnoses, AI's influence is undeniable [2]. However, with this progress comes a critical question: who is accountable when AI malfunctions and causes harm?

This article probes the nascent legal landscape surrounding victims of AI-inflicted harm, with a focus on criminal and civil liability. Traditional legal frameworks in both criminal and civil law often rely on the concept of mens rea – the guilty mind of the perpetrator [3]. However, assigning criminal intent to a complex algorithm devoid of consciousness or moral compass presents a significant challenge [4].

This article seeks to deconstruct this intricate issue, exploring the challenges of assigning liability in the age of AI. We will analyse existing legal frameworks and delve into potential solutions proposed by legal scholars. Should we impose a stricter supervisory duty on those who develop and deploy AI, or is it time to consider granting legal personhood to certain forms of highly autonomous AI?

By exploring these questions, this article strives to contribute to the ongoing discourse surrounding legal frameworks in an AI-driven future. As AI continues to permeate our lives, establishing clear lines of accountability is critical for protecting individuals from AI-inflicted wrongs. This article aims to spark a necessary dialogue, one that will guide us in navigating the legal complexities of our increasingly intelligent world.

Exploring AI-related incidents
AI is rapidly integrating into society, offering significant benefits across various sectors. However, alongside these advantages lies the potential for harm. AI-related incidents refer to situations where AI systems malfunction, produce biased outputs, or cause unintended consequences, leading to negative impacts on individuals and society [5]. These incidents can range from minor accidents and inconveniences to severe safety risks and privacy violations. The potential impact of AI-related incidents can be multifaceted. Individuals may experience

- **Physical harm:** Malfunctioning autonomous vehicles or AI-controlled robots causing accidents [6].

- **Financial loss:** Biased algorithmic decisions in loan approvals or insurance premiums [7].

- **Psychological distress:** Exposure to harmful or offensive content generated by AI systems [8].

On a societal level, AI-related incidents can lead to

- **Erosion of trust:** Repeated incidents can damage public trust in AI technologies.

- **Widening social inequalities:** Biased algorithms can exacerbate existing social and economic disparities.

- **Ethical dilemmas:** The development and deployment of certain AI systems may raise complex ethical questions regarding privacy, autonomy, and control.

AI-related incidents leading to harm
The potential for harm from AI-related incidents is not merely theoretical. Here are some real-world examples

- **Algorithmic bias:** A facial recognition system used by law enforcement exhibited racial bias, disproportionately misidentifying people of color [9].

- **Self-driving car accidents:** Several accidents involving self-driving cars have raised concerns about their safety and the ability of AI to handle complex driving scenarios [10].

- **AI-generated deepfakes:** These realistic video forgeries can be used to damage reputations and propagate misinformation [11].

These examples highlight the diverse ways AI systems can malfunction or produce unintended consequences, causing harm to individuals and society.

Complexities of attributing liability in AI-related Incidents
Attributing liability in AI-related incidents presents a significant challenge. Traditional legal frameworks rely on the concept of mens rea, that is, the guilty mind of the perpetrator. However, with complex AI systems, determining who is responsible for an incident becomes less clear-cut. Here are some reasons for this complexity

- **Opacity of AI systems:** The inner workings of some complex AI systems are black boxes, making it difficult to establish the cause of error [12].

- **Multiple actors involved:** The development and deployment of AI systems often involve multiple actors, including developers, programmers, and those who integrate the technology into a final product. Assigning blame becomes difficult when responsibility is diffused across different parties [13].

- **Evolving nature of AI:** The rapid pace of AI development means that legal frameworks may struggle to keep up with the complexities of new and emerging AI applications [14].

These complexities necessitate a rational and detailed approach to establishing liability in AI-related incidents. In addition, the legal frameworks will need to adapt to ensure appropriate accountability and protect individuals from harm.

Criminal liability in AI-related incidents
The rapid integration of AI into society raises a crucial question: who can be held criminally liable when AI malfunctions and causes harm? Traditional criminal law struggles to adapt to this new landscape, as it often relies on concepts like mens rea (guilty mind) that are difficult to apply to non-sentient AI systems. This section explores the complexities of criminal liability in AI-related incidents.

Legal principles and frameworks
Criminal law traditionally requires proof of both mens rea (guilty mind or criminal intent) and actus reus (the physical act constituting the crime) to hold someone criminally liable [15]. However, AI systems lack consciousness and intent, making the application of these principles challenging. Here are some potential legal frameworks being considered

- **Criminal liability of programmers/developers:** Developers could be held criminally liable if they deliberately design an AI system to cause harm, like creating a dangerous weapon [16]. However, proving deliberate intent can be difficult.

- **Strict liability for AI companies:** Imposing strict liability, where fault does not need to be proven, could incentivize companies to prioritize safety measures in AI development [17]. This approach may be criticized as unfair in cases of unforeseeable outcomes.

- **Command responsibility:** This principle, used in military law, could be adapted to situations where users give AI explicit instructions to commit a crime [18]. The challenge lies in defining the level of user control over an AI system.

Case studies and legal precedents
Due to the relative novelty of AI, there are few established legal precedents regarding criminal liability in AI-related incidents. However, some relevant cases offer insights

- **Uber self-driving car crash (2018):** An Uber self-driving car struck and killed a pedestrian. The criminal case focused on the human safety driver, raising questions about liability attribution in autonomous systems [19]. This case highlights the difficulty of assigning blame when humans and AI interact.

- **South Korea's "AI ethics guidelines" (2020):** These guidelines advocate for potential criminal liability of AI developers in cases of intentional harm or negligence leading to severe consequences [20]. While not legally binding, this example demonstrates a growing need for accountability in AI development.

The lack of established precedents necessitates the need for a legal framework to address the unique challenges posed by AI-related incidents.

Challenges and limitations in holding AI stakeholders criminally liable
Assigning criminal liability in AI-related incidents presents a significant challenge due to the complexities of AI systems and the current legal framework. The challenges arise due to the following reasons

- **Causation:** Establishing a clear causal link between an AI system and the harm caused can be difficult. In situations involving multiple actors (developers, users, third-party systems) and complex chains of events, pinpointing the primary cause becomes a significant hurdle [21].

- **Attribution of intent (Mens Rea):** Traditional criminal law relies on mens rea (criminal intent) of the perpetrator. However, AI systems lack consciousness and intent, making it difficult to hold them criminally liable [22].

- **Opacity of AI systems:** The inner workings of some AI systems, particularly deep learning models, can be opaque and complex. This opacity makes it challenging to understand how an error occurred or who is responsible for fixing it [23].

- **Evolving nature of AI:** The rapid pace of AI development poses a challenge for legal frameworks. The law struggles to keep up with the complexities of new and emerging AI applications, leaving gaps in how liability is assigned [24].

- **Shared responsibility:** The development and deployment of AI systems often involve multiple stakeholders, including developers, programmers, companies integrating the technology, and potentially even users who provide training data. Diffused responsibility across these actors makes it difficult to pinpoint who should be held criminally liable [25].

- **Potential for chilling effect:** Overly strict or ambiguous legal frameworks could stifle innovation in the field of AI. Developers may become hesitant to take risks or explore new avenues for fear of criminal prosecution [26].

Civil Liability in AI-related incidents

As AI becomes an inseparable part of society, the potential for harm increases. Civil liability, which focuses on compensating victims for damages, becomes crucial. This section explores the legal landscape surrounding civil liability in AI-related incidents.

Legal concepts and standards

Civil liability frameworks aim to provide compensation to those who are harmed by the actions or negligence of others. Here are some key concepts governing civil liability in AI-related incidents

- **Duty of care:** A party has a duty of care to avoid causing harm to others. In the context of AI, this duty could extend to developers, manufacturers, and those who deploy AI systems [27].

- **Standard of care:** This refers to the level of care expected from a party fulfilling a duty of care. The standard could depend on the specific AI application [28].

- **Causation:** The plaintiff (injured party) must prove that the AI system caused the harm. This can be challenging in complex AI systems with multiple actors involved [29].

Legal theories in civil litigation

Several legal theories may be applicable in civil lawsuits involving AI-related harms

- **Negligence:** Plaintiffs can argue that the developer, manufacturer, or user of the AI system acted negligently in its design, development, deployment, or use, leading to harm [30].

- **Product liability:** This theory may apply if the AI system is considered a product. Plaintiffs can claim the AI system was defective or unreasonably dangerous, causing harm [31].

- **Vicarious liability:** This theory holds a party liable for the actions of another under their control. For example, a company could be held vicariously liable for the actions of an AI system it deploys [32].

Challenges in seeking compensation

Despite these legal theories, seeking compensation for AI-related harms through civil litigation faces several challenges

- **Causation:** As mentioned earlier, proving that the AI system was the primary cause of harm can be difficult, especially in situations with multiple actors.

- **Opacity of AI systems:** It is difficult to how AI programs arrived at a particular decision or why they malfunctioned due to opacity of their inner working. This diminishes the ability to pinpoint fault.

- **Novelty of AI technology:** The legal system is still grappling with AI, and established legal principles may not fully apply to these new technologies.

- **Cost of litigation:** Civil lawsuits are inherently expensive and time-consuming. Victims of AI-related harms may lack the financial resources to hire legal representation and technical experts needed to navigate the complexities of these cases, especially against well-funded corporations developing and using AI systems [33].

Ethical considerations and policy implications of AI-related incidents

The increasing presence of AI in society raises critical ethical questions and necessitates policy considerations.

Ethical Implications of AI-Driven Outcomes

AI systems can deliver biased or discriminatory outcomes, impacting individuals and communities. Here's a breakdown of some key ethical concerns

- **Algorithmic bias:** AI systems trained on biased data can perpetuate those biases in their decisions. This can lead to unfair hiring practices, loan denials, or discriminatory policing practices [34].

- **Lack of transparency:** The inner workings of complex AI systems can be opaque; thus it is difficult to understand how they arrive at decisions. When AI's working is not clear, it makes it hard to trust and know who is responsible [35].

- **Privacy concerns:** AI systems often rely on vast amounts of personal data to function which can potentially violate data privacy [36].

- **Job displacement:** AI pose a threat to certain jobs, potentially leading to unemployment and economic hardship [37].

These ethical concerns can have a significant impact on victims of AI-related incidents. For example, someone denied a loan due to algorithmic bias may face financial difficulties.

Role of regulatory authorities and policymakers

Regulatory authorities and policymakers have a crucial role to play in addressing these ethical concerns and shaping legal frameworks for AI development and deployment. Some key areas of focus

- **Developing ethical guidelines:** Policymakers can create guidelines for ethical AI development and deployment [38].

- **Data privacy regulations:** Strong data privacy regulations can ensure that the personal data is not misused [39].

- **Algorithmic bias mitigation:** Regulatory frameworks can encourage developers to implement techniques for mitigating bias in AI systems [40].

- **Standards for explainability:** Standards for explainable AI can help ensure that AI decisions are understandable and auditable [41].

Recommendations for ethical AI development
The following recommendations aim to enhance accountability, protect victims' rights, and promote ethical AI development

Enhancing accountability
- **Standardized impact assessments:** Mandate developers to conduct comprehensive social and ethical impact assessments before deploying AI systems, considering potential biases and risks [42].
- **Clearly demarcated responsibility:** Establish areas of responsibility for different stakeholders involved in the AI lifecycle, from development to deployment [43]. This clarifies who can be held accountable in case of incidents.

- **Explainable AI (XAI) standards:** Implement industry-wide standards for explainable AI, ensuring decisions made by AI systems are understandable and auditable [44].

Protecting victims' rights
- **Right to explanation:** Grant individuals the right to receive an explanation for decisions made by AI systems that impact them, allowing them to challenge unfair outcomes [45].

- **Accessible legal mechanisms:** Establish clear and accessible legal mechanisms for victims of AI-related harms to seek compensation and hold perpetrators accountable [46].

- **Algorithmic due process:** Develop legal frameworks that ensure algorithmic due process, protecting individuals from discriminatory or unfair treatment by AI systems [47].

Promoting ethical AI development
- **Diversity in AI development teams:** Encourage diverse and inclusive teams in AI development to mitigate bias and foster a wider range of perspectives [48].

- **Public-Private partnerships:** Foster public-private partnerships to promote ethical AI development and research, ensuring collaboration between stakeholders [49].

- **Ethical AI education:** Integrate AI ethics education into computer science and engineering curricula, equipping future developers with a strong foundation in ethical considerations [50].

Conclusion
The rapid integration of Artificial Intelligence (AI) into society presents a complex challenge: who is accountable when AI malfunctions and causes harm? This analysis has explored the difficulties in assigning criminal and civil liability in AI-related incidents.

Traditional legal frameworks struggle to grapple with the unique characteristics of AI technology. The lack of intent inherent in AI systems and the opacity of some AI decision-making processes make it challenging to establish causation and pinpoint responsibility. These challenges leave victims of AI-related harms facing an uncertain path to compensation and accountability.

However, this analysis is not a call for limiting the use of AI. The potential benefits of AI are vast, and responsible development can mitigate the risks. Moving forward, several areas demand focus

- **Standardizing explainable AI (XAI):** XAI advancements can make AI decision-making more transparent, fostering trust and accountability.

- **Algorithmic bias mitigation:** Research on mitigating algorithmic bias is crucial to ensure fair and non-discriminatory AI outcomes.

- **Legal frameworks for AI liability:** Developing clear legal frameworks for criminal and civil liability in AI-related incidents will establish a foundation for holding stakeholders accountable and ensuring victims have recourse.

Addressing these areas necessitates collaboration between policymakers, researchers, industry leaders, and ethicists. By working together, we can resolve the complexities of AI governance and make way for a future where AI serves humanity without compromising ethical principles or victim's rights.

References
1. Sabharwal A, Selman B. S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Third Edition. Artif Intell,2011:175:935-937. DOI: 10.1016/j.artint.2011.01.005.
2. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.
3. Duff RA. Intention, Agency and Criminal Liability: Philosophy of Action and the Criminal Law. Blackwell, 1990. Available from: SSRN: https://ssrn.com/abstract=2637418
4. Le M, Jung E-C. Analysis of intent-design relationship for artificial intelligence design agent model based on product purchasing process. Proceedings of the Design Society: Design Conference,2020:1:285-294. DOI: 10.1017/dsd.2020.146
5. OECD. The OECD AI Policy Observatory: Expert Group on AI Incidents, 2019. Available from: https://oecd.ai/en/network-of-experts/working-group/10836

6. Wang J, Zhang L, Huang Y, Zhao J. Safety of Autonomous Vehicles. Journal of Advanced Transportation,2020:2020:8867757. DOI: 10.1155/2020/8867757

7. Garcia ACB, Garcia MGP, Rigobon R. Algorithmic discrimination in the credit domain: what do we know about it?. AI & Soc,2023. DOI: 10.1007/s00146-023-01676-3

8. Doki S, S S, Hori D, *et al*. Comparison of predicted psychological distress among workers between artificial intelligence and psychiatrists: a cross-sectional study in Tsukuba Science City, Japan. BMJ Open,2021:11(6):e046265. DOI: 10.1136/bmjopen-2020-046265

9. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. March, New York: Association for Computer Machinery – ACM, 2021.

10. National Highway Traffic Safety Administration (NHTSA). Self-Driving Car Safety,2023. Available from: https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety

11. Shao J, Zheng Y, Qin J, Bao L, Li Z. Detecting Deepfakes in the Wild: A Survey, 2020. Available from: https://arxiv.org/abs/2004.11138

12. Lipton ZC. The Mythos of Model Interpretability, 2018.

13. Santoni de Sio F, Mecacci G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. Philos Technol,2021:34:1057–1084. DOI: 10.1007/s13347-021-00450-x

14. Mishra PK. AI and the legal landscape: Embracing innovation, addressing challenges. Available from: https://www.livelaw.in/lawschool/articles/law-and-ai-ai-powered-tools-general-data-protection-regulation-250673?infinitescroll=1

15. Duff RA. Intention, Agency and Criminal Liability: Philosophy of Action and the Criminal Law. Blackwell, 1990. Available from: SSRN: https://ssrn.com/abstract=2637418

16. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.

17. Whitby B. Design for a Knowing and Responsible Artificial Agent. Metaphilosophy,2000:31(1-2):5-36. Available from: https://www.cambridge.org/core/journals/proceedings-of-the-design-society-design-conference/article/analysis-of-intentdesign-relationship-for-artificial-intelligence-design-agent-model-based-on-product-purchasing-process/13224DD1363F7329104B677C0973879F

18. Allen C, Wallach W. Moral Machines: Ethical Reasoning in Artificial Intelligence. Oxford University Press, 2009.

19. National Highway Traffic Safety Administration (NHTSA). Self-Driving Car Safety, 2023. Available from: https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety

20. Kim SH. Artificial Intelligence and Criminal Law in South Korea. Journal of International Criminal Justice,2020:18(2):381-398

21. Selbst AD, *et al*. Regulating Artificial Intelligence: A Multi-Stakeholder Approach. Center for Security and Emerging Technology, Georgetown University, 2020. Available from: https://cset.georgetown.edu/article/what-were-reading-on-ai-regulation/

22. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.

23. Lipton ZC. The Mythos of Model Interpretability. [Internet],2018 [cited 2024 Mar 23]. Available from: https://arxiv.org/abs/1606.03490

24. Matthias A. The Risks of Artificial Intelligence. Cambridge University Press, 2020.

25. Allen C, Wallach W. Moral Machines: Ethical Reasoning in Artificial Intelligence. Oxford University Press, 2009.

26. Tartaro A, Leon Smith A, Shaw P. Assessing the impact of regulations and standards on innovation in the field of AI. Papers,2023. arXiv:2302.04110.

27. Allen C, Wallach W. Moral Machines: Ethical Reasoning in Artificial Intelligence. Oxford University Press, 2009.

28. Matthias A. The Risks of Artificial Intelligence. Cambridge University Press, 2020.

29. Ponzio R, Yusuf N, Mallinson F, Shahrukh M. Future of International Cooperation Report 2023. The Stimson Center, Doha Forum, and Global Institute for Strategic Research, Washington, D.C., USA and Doha, Qatar, 2023.

30. Akindele R, Jesuloluwa Adewuyi S. Navigating the Ethical and Legal Terrains of AI Tool Deployment: A Comparative Legal Analysis. Communications of the IIMA, 2023, 21(1). Article 8. DOI: 10.58729/1941-6687.1449. Available from: https://scholarworks.lib.csusb.edu/ciima/vol21/iss1/8

31. Calo R. Artificial Intelligence Policy: A Primer and Roadmap [Internet], 2017 Aug 8 [cited 2024 Mar 23]. Available from: https://ssrn.com/abstract=3015350 or http://dx.doi.org/10.2139/ssrn.3015350

32. Whitby B. Design for a Knowing and Responsible Artificial Agent. Metaphilosophy,2000:31(1-2):5-36. Available from: https://www.cambridge.org/core/journals/proceedings-of-the-design-society-design-conference/article/analysis-of-intentdesign-relationship-for-artificial-intelligence-design-agent-model-based-on-product-purchasing-process/13224DD1363F7329104B677C0973879F

33. Buiten M, de Streel A, Peitz M. The law and economics of AI liability. Computer Law & Security Review,2023:48:105794. DOI: 10.1016/j.clsr.2023.105794.

34. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.

35. Lipton ZC. The Mythos of Model Interpretability. [Internet],2018 [cited 2024 Mar 23]. Available from: https://arxiv.org/abs/1606.03490

36. Laux J, Wachter S, Mittelstadt B. Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance,2024:18:3-32. DOI: 10.1111/rego.12512

37. Ford M. Rise of the Robots: Technology and the Threat of a Jobless Future. Basic Books, 2015.

38. Organization for Economic Co-operation and Development (OECD). Recommendation of the

Council on Artificial Intelligence. [Internet],2019 [cited 2024 Mar 23]. Available from: https://legalinstruments.oecd.org/api/print?ids=648&lang=en

39. European Union General Data Protection Regulation (GDPR),2016. [Internet]. Available from: https://gdpr-info.eu/

40. S, Dr. Varsha P. How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights,2023:3(1):100165. DOI: 10.1016/j.jjimei.2023.100165.

41. DARPA (Defense Advanced Research Projects Agency). Explainable AI (XAI) Program,2023. [Internet]. Available from: https://www.darpa.mil/program/explainable-artificial-intelligence

42. P. S, Dr. Varsha. How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights,2023:3(1):100165. DOI: 10.1016/j.jjimei.2023.100165.

43. Harasimiuk D, Braun T. Regulating Artificial Intelligence: Binary Ethics and the Law,2021. DOI: 10.4324/9781003134725.

44. DARPA (Defense Advanced Research Projects Agency). Explainable Artificial Intelligence (XAI) Program,2023. [Internet]. Available from: https://www.darpa.mil/program/explainable-artificial-intelligence

45. European Commission. Ethics Guidelines for Trustworthy AI,2020. [Internet]. Available from: https://digital strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

46. Mantelero A. Regulating AI. In: Beyond Data. Information Technology and Law Series, vol 36. The Hague: T.M.C. Asser Press, 2022. DOI: 10.1007/978-94-6265-531-7_4.

47. Akindele R, Jesuloluwa Adewuyi S. Navigating the Ethical and Legal Terrains of AI Tool Deployment: A Comparative Legal Analysis. Communications of the IIMA,2023:21(1):8. DOI: 10.58729/1941-6687.1449. Available from: https://scholarworks.lib.csusb.edu/ciima/vol21/iss1/8

48. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.

49. OECD. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

50. European Commission. Ethics Guidelines for Trustworthy AI, 2020. [Internet]. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai