Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
July-December 2023 4(2), 5848–5852

LEGAL ACCOUNTABILITY OF AUTONOMOUS AI SYSTEMS IN CRIMINAL JUSTICE

Gurprem Monga [1]

[1] Bachelor of Science (B. Sc non-med), Bachelor of Laws (LL.B.), Master of laws (LL.M.)- Criminology, India

DOI
10.29121/shodhkosh.v4.i2.2023.6172

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2023 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License.

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

OPEN ACCESS

ABSTRACT

The speed of change caused by autonomous AI systems poses a basic problem for principles of criminal law. When an AI system takes an action that is normally a crime if taken by a human being, it becomes far less clear as to how to impose liability under law. Current legal frameworks, which are based on concepts of human agency, intent, and responsibility, lack the legal suitability to apply these authorities to an act that has autonomously been executed, with particular attention to concepts of AI blackness and autonomy. Further muddling the issue is the way in which many modern AI systems are "black boxes," especially deep learning systems that learn and evolve in unintelligible ways even to their inventors. Given this black box characteristic, it becomes nearly impossible to track a harmful action to a particular programming mistake, dataset flaw, or decision. Without this ability to show that a human actor (developer, agency owner, or consumer) caused the harmful output of AI, responsibility will be diffused along the chain of actors as it will be exceedingly hard to match the harmful output to a single actor responsible for that harmful output. The arrival of autonomous artificial intelligence systems necessitates a reconsideration of basic legal concepts. Actus reus traditionally relies on a morally culpable human actor - and it falls short of addressing AI harm. While there are serious legal complications with blaming the programmer or user, corporate criminal responsibility for manufacturers is a more feasible option. As AI continues to evolve, our legal systems must adjust their approach to ensure accountability, protect the public from harm, and fairly allocate responsibility when the distinction between human and machine actions erodes.

Keywords: Legal, Accountability, Autonomous, AI Systems, Criminal, Justice, Law

1. INTRODUCTION

While autonomous AI systems can lead to a more efficient and objective criminal justice system, they also pose risks—most notably, bias, lack of transparency, and removal of human judgement. If these AI systems are developed without careful consideration, they could lead to unjust outcomes. The future will depend on the ability of policymakers, technologists, and legal experts to work together and develop a model that will allow our society to benefit from AI systems, while maintaining the essential elements of justice, equity and human rights. (Gerding, 2022)

Actus reus, which is a basic element of criminal law, is the activity or inaction that constitutes the offence. Mens rea is the guilty mind. In order for someone to be liable for a crime, there must be both the guilty act and the guilty mind.

Criminal law is typically founded on two essential elements:

- **Actus Reus:** The guilty act or the physical commission of a crime.
- **Mens Rea:** The guilty mind or the criminal intent behind the act.

An autonomous AI may be able to commit an actus reus (like an autonomous vehicle causing a fatality), but it cannot satisfy mens rea. An AI system cannot consciously think or morally reason or have the ability to intend,reason,or act negligently like a human. Whatever the AI system does is algorithmic, and it is nothing more than data processing; it has no "guilty mind." This is a legal issue. Since the AI system cannot fulfill the fundamental aspects of a crime, traditional legal theory will argue that the AI system is not being criminally liable. (Fatima, 2022)

Legal scholars and policymakers are exploring several approaches to address this accountability gap:

Liability of Human Actors: This is the most common and practical approach today. The liability for an AI's actions is assigned to a human or corporation.

Manufacturer/Developer Liability: Holding developers liable for deliberate or negligent programming that leads to a crime.

User/Operator Liability: Placing responsibility on the end-user for improper use, a failure to supervise the AI, or providing it with criminal instructions.

Corporate Liability: Treating the AI's actions as an extension of the company that created or deployed it. This is similar to how corporations are held liable for the actions of their employees.

Strict Liability: Another radical idea we could consider would be to place strict liability on all companies creating and deploying high-risk AI. In other words, if the company develops a high-risk AI that causes harm, it is responsible for that harm, regardless of whether the company had anything to do with the eventual deployment of the AI. The hope is this would get companies to prioritize safe deployment and encourage rigorous testing of the AI prior to its use. (Chidiogo, 2024)

AI as a Legal Person: Some legal scholars have proposed throwing "legal person" rights onto AI systems as a limited version of a corporation. In this situation, AI systems would be having "legal person" status, the AI itself is the entity that provides the legal action, punishment and penalties could be as extreme as the deletion of the AI or a forced reprogramming of the program or algorithm. This is a very contentious issue because it eliminates the mens rea question entirely, and raises other questions regarding the legal personhood and personhood capability of AI systems.

Many autonomous AI systems operate as "black boxes," meaning their decision-making processes are opaque and difficult to understand. This lack of transparency makes it nearly impossible for individuals to challenge an AI's output or for judges to scrutinize the rationale behind a recommendation. This opacity erodes due process and accountability, as it becomes unclear who is responsible when an AI system makes a harmful or incorrect decision. Was it the developer, the cop, or the judge who depended on the tool? (Singh, 2021)

If society overly relies on artificial intelligence, then human judgement and compassion may decline, which is necessary to make judgements presented in a nuanced fashion. Legal decision-making, whether focused on sentencing, parole, or bail requires a choice made with respect to personal context. When a purely quantitative and algorithmic principle is adopted, these essential human considerations are lost.

The use of AI for surveillance and predictive policing poses deeper issues about invasion of privacy because it includes using big data to draw from enormous collections of personal data in order to analyze it, typically without explicit consent. This typically creates a chilling effect on civil liberties so that citizens engage in constant self-surveillance.

AI has the capacity to analyze vast amounts of data including criminal reports, video data from surveillance cameras, and social media data, at speeds unattainable for humans. The researchers presenting this work offered new ways of leveraging AI through predictive policing models to see and potentially predict crime. (Jennifer, 2021)

2. LITERATURE REVIEW

Fahim et al. (2024): Predictive policing models allow law enforcement to leverage advanced, predictive algorithms to identify the areas that have high amounts of crime or crime areas in order to consistently better allocate resources. A department can potentially use historical crime data and analyze for any latent patterns to predict the high-risk areas where police are deployable for intervention before the crime actually occurs.

Sayyed et al. (2024): AI based tools can examine work such as forensic analysis to assist in evidence processing of DNA and/or other evidence, or examine digital evidence sometimes more adept and consistently than one single human expert can do.

Dennis et al. (2020): When utilizing automation with AI, human error and changes in subjective bias, which may include bias against a suspect may be hindered. The judiciary can also use AI based tools that can search hyperlaps across legal research databases, review implications of upcoming and prior case precedents against proposed assignments, and provide case management including preparation and determinations through speed and consistency.

Osmani et al. (2020): In most legal systems around the world there is a significant chronic backlog. The existence of that backlog is also compounded by the amount of time it takes to accomplish these tasks on upcoming cases to have them completed already.

Aggarwal et al. (2020): Automated clerks, using AI, can help clear the backlog by coding amongst other tasks assignments including transcribing transcripts, digitizing court documents, externally storing files, and automating manual processes. This will free up human actors to carry on with their human or technological related processes while at the same time dealing with the backlog crises.

Chatterjee et al. (2021): AI models are only as good as the data to train them. If historical crime data reflects existing societal biases—for example, if certain neighborhoods have been over-policed or if particular demographics have been disproportionately arrested—the AI will learn and amplify these biases. This can lead to a vicious cycle where AI-powered systems unfairly target marginalized communities, reinforcing systemic inequalities and creating discriminatory outcomes in sentencing, bail, and parole decisions.

3. RESULTS AND FINDINGS

AI systems are trained on large datasets. If this data represents current social biases, the AI will learn and perpetuate our biases. It can learn these biases and may enhance existing ones. For example, if a risk assessment tool is trained on historical arrest data, it may also take biased decisions on certain groups regardless of individual's situations. If someone is impacted by a decision that is biased by an AI and that AI was trained from data that points to previous bias in our system, this is contrary to equal protection under the law, and has the potential to yield discriminatory results in bail, sentence, and parole decisions.

Present laws were not written with autonomous AI in mind and lack clear guidance on items like, data privacy, liability, and evidence admissibility. We have a mishmash of regulations across multiple jurisdictions, with numerous conflicting rules and regulations. So this lack of clarity creates jurisdictional challenges to establishing a lawful approach to autonomous AI in the criminal justice system and facilitates the development and use of AI for purposes that are contrary to the laws that underpin our society.

Predictive policing, which is perhaps the most prominent example of autonomous AI in the criminal justice system, is based on systems and tools that can investigate historical crime data in order to determine "hot spots" where crimes will likely happen in the future. While predictive policing was generally intended to make policing more efficient, it introduces problematic legal and ethical questions.

AI is only as good and ethical as the data it learns from. If the historical data contains any biases, structural in nature, there could be an enormous problem. For instance, if a police department over-policed a neighborhood historically, that neighborhood will likely have disproportionately higher arrest data, unduly informing the AI about crime because the AI will not learn that the data is also affected by the history of over-policing there. This arguably creates a vicious cycle in which police are disproportionately deployed to that neighborhood, leading to more arrests and further reinforcing the underlying data bias.

The ProPublica investigation of Compass (Correctional Offender Management Profiling for Alternative Sanctions) in the United States is arguably the best-known case study of this possibility. In their investigation, they reported that Compass was more likely to wrongfully label black defendants as future criminals than white defendants.

Who is responsible for the biased outcome? The police department that is deploying the use of Compass? The company that built Compass and therefore owned the algorithm? The people who collected the data that was flawed?

The legal world must also deal with negligence, product liability, and due process violations. Unclear is whether a defendant may be successful in challenging a risk score in light of the algorithm's "black box" phenomenon, that is, the inability to explain how a particular decision was made. AI-enabled surveillance, including facial recognition surveillance and drones, signals important accountability issues. The extreme case, albeit in largely theoretical terms with respect to use in the domestic context, is lethal autonomous weapons systems (LAWS).

An autonomous drone might be patrolling a city when it identifies a person as a potential threat based on a series of gait analyses and facial recognitions in conjunction with behavioral patterns. Consequently, the drone takes action, such as to deploy a non-lethal deterrent, leading to injury or death. However, there is not a human "trigger-puller." Rather, the drone has swept through layers of evidence to cognize the need for action.

The task of distinguishing criminal responsibility is immensely challenging. Does the AI have legal personhood, such that it may be held accountable for its own actions? Most legal academics would say no, as the AI lacks mens rea (criminal intent). If there is no algorithmic guilt, ultimate accountability would rest with the human creators and operators.

- **Manufacturer/Developer:** Could they be held liable for a flaw in the design or programming? This would fall under product liability law, but proving negligence in a complex, self-learning system is a significant challenge.
- **Operator/Commander:** The person who deployed the drone might be held responsible, but they didn't directly cause the harm. Their liability would likely depend on whether they had reasonable oversight and whether the system was used for its intended purpose.
- **The Problem of Distancing:** As AI systems become more complex and autonomous, the chain of causation between a human decision and a harmful outcome becomes more diffuse, creating an accountability gap.

Another important sector is using AI to assist judges in making decisions on bail, sentencing, and parole. These systems are built to analyze the information about a defendant (criminal record, age, employment, etc.) to predict their likelihood of recidivism.

In much the same way as predictive policing, these systems can incorporate and exacerbate bias prevalent in society. For example, if a system assigns a defendant from a low-income community a higher risk score, it is only doing so because the demographic has a higher historical rate of arrest according to the training data, resulting in a harsher sentence or denial of bail. This violates the tenets of individualized justice and undermines the principles of due process. While a defendant has the right to confront the reasons for their conviction, how does one cross-examine a company proprietary algorithm?

If a judge applies a biased AI recommendation resulting in an excessive sentence, is the judge liable? Or is the company that sold the software? Current legal thought may suggest the ultimate accountability rests with the human judge, but this provides a false sense of security, as the judge may end up relying solely on the AI recommendation and unnecessarily disclaim their judicial discretion.

There needs to be strict rules set out by the legal system regarding how and when AI can be used, such that human discretion remains the final say. This, I believe, is part of the rationale behind international legal frameworks like the EU AI Act focusing on high risk applications and requiring increased transparency and human oversight.

Governments need to set strict legal and ethical frameworks for the development and implementation of AI in criminal justice. This includes regular audits of AI systems for bias, responsibilities for algorithms to be transparent, and unambiguous lines of responsibility when errors are made.

The focus also shouldn't be on fully autonomous systems rather on "human in-the-loop" models; so that AI becomes a powerful tool that augments human judgement rather than replacing it. This ensures that humans retain ultimate authority over the decision, and can use AI's information to prescriptions to assist decision making whilst applying reasoning, care and discretion.

In terms of addressing algorithmic bias, AI models need to be trained on diverse and representative training datasets, de-biasing historical datasets, and be careful that our method of generating new data does not replicate inequalities.

The fundamental quandary in attributing actus reus to an AI system is the absence of a human agent capable of executing the harmful act. One could argue that the programmer is accountable because he or she wrote the code that caused the harmful act. This is problematic.

The developer may have made an inadvertent creation, and based on the understandable intention of creating a system with no intention of causing harm, the autonomous and self-learning capability of the AI may reasonably induce unforeseeable results that no person could have predicted were possible. Of course, holding the programmer to be legally liable for a result he or she could NOT foresee as possible, would likely taint legal precedent.

The user of an autonomous system may also be liable, at law, similar to that of a car owner being responsible for a car's actions, but if the user had no active and direct involvement into the AI system's decision at the time of the act, then the user's involvement could be passive, and at law, to assign liability for actus reus becomes a serious legal question; when the AI or automated design process acted alone or autonomously.

A second possibility is that the company or corporation that produced the AI system might be held responsible under corporate criminal liability theory, and maybe this is the most plausible outcome. Corporate criminal liability does see

the corporation as a legal person and can hold them accountable for their products' actions, in particular if the corporation has acted irresponsibly in their design, tests and/or safety protocols. This will hold the corporation accountable rather than the individual human responsible for the technology they released into the world.

An often highly debated and path breaking idea is essentially giving a narrow form of legal personhood to highly autonomous AIs, this would enable the AI itself to be made a legal entity subject to actions, but then there would be very difficult questions about punishment and a robot's "rights". Moreover it is not much more than a philosophical question than a viable solution for the near future.

4. CONCLUSION

The legal frameworks that are currently being developed , such as the EU's AI Act exist largely as legal regulations based off of a risk-based approach. This leads to AI systems that are deemed "high-risk" (e.g., law enforcement AI or AI used in critical infrastructure) to be more heavily regulated than "low-risk" AI systems. The purpose is to find a happy medium between encouraging innovation while ensuring that society is protected from harm and exploitation. As AI becomes ingrained within the criminal justice landscape and into our daily lives, new laws and international treaties will be necessary to make the lines of responsibility clear, and reinforce ideals of fairness, accountability, and openness.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Sadaf Fahim, Ethico-Legal Aspect of AI-driven Driverless Cars: Comparing Autonomous Vehicle Regulations in Germany, California, and India 186 (Oxford University Press, Delhi, 1st edn., 2024)

Hifajatali Sayyed, "Artificial Intelligence and Criminal Liability in India: Exploring Legal Implications and Challenges", 10 Cogent Social Sciences 15-34 (2024).

Dennis J. Baker & Paul H. Robinson, Artificial Intelligence and the Law: Cybercrime and Criminal Liability, 2020

Nora Osmani, "The Complexity of Criminal Liability of AI Systems", 14 Masaryk University Journal of Law and Technology 53 (2020).

Thomas C. King, Nikita Aggarwal, et. el., "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions", 26 Science and Engineering Ethics 89-120 (2020).

Sheshadri Chatterjee and Sreenivasulu N.S., "Artificial Intelligence and Human Rights: A Comprehensive Study from Indian Legal and Policy Perspective", 10 International Journal of Law and Management 94 (2021)

Jennifer Cobbe & Jatinder Singh, "Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges", 42 Computer Law & Security Review 579 (2021).

Chidiogo Uzoamaka Akpuokwe, et. et., "Legal Science & IT Research Journal Challenges of Artificial Intelligence and Robotics: A Comprehensive Review", 5 Computer 546 (2024).

Fatima Dakalbab, Manar Abu Talib, et el., "Artificial Intelligence & Crime Prediction: A Systematic Literature Review", 6 Social Sciences & Humanities Open 142 (2022).

Vahid Yazdanpanah, Enrico H. Gerding, et. el., "Reasoning About Responsibility in Autonomous Systems: Challenges and Opportunities", 38(4) AI & Society (2022).