The Criminal Code of Machines: A Deep Dive into the Risks of Delegating Moral Judgment to Artificial Intelligence

Shruti Sinha[1], Dr. Debashree Chakraborty[2]

[1]Assistant Professor, Techno India Law College
[2]Associate Professor and Head of Department, Techno India Law College

Abstract

As artificial intelligence systems become increasingly involved in normative decision-making, particularly in the criminal justice system, their moral and legal implications must be rigorously examined. This article explores the dangers of assigning moral judgment to AI, focusing on the risks of bias, lack of accountability, ethical opacity, and the erosion of human-centric jurisprudence. It argues for strict limits on AI autonomy in legal contexts and proposes frameworks for responsible deployment.

Introduction

In the digital age, legal systems across the globe are increasingly integrating artificial intelligence (AI) tools into their decision-making frameworks. These technologies are being used in various domains, including predictive policing, risk assessments, and even sentencing recommendations. Algorithms, often operating behind the scenes, are gradually reshaping how justice is administered by offering greater speed, consistency, and efficiency in handling vast amounts of data.[1] However, as these tools become more embedded in the criminal justice process, they raise critical ethical and philosophical questions. Can machines—devoid of human consciousness, empathy, and moral reasoning—truly make decisions that impact people's lives and freedoms? And more importantly, should they be entrusted with such responsibilities? The use of AI in law challenges our traditional understanding of justice, accountability, and fairness. As we move forward, it becomes imperative to scrutinize whether delegating moral judgment to machines undermines the core principles that legal systems are built upon.

AI and the Myth of Objectivity

Artificial Intelligence is frequently praised for its impartiality, consistency, and logical decision-making, seemingly free from the emotional biases that often cloud human judgment. This perception creates a myth of objectivity—that AI is inherently neutral. However, this belief does not hold up under closer examination. AI systems are not created in a vacuum; they are trained on historical data that often reflects deeply rooted social biases, including discriminatory policing, unequal access to justice, and systemic racism.[2] Rather than removing bias, these systems frequently encode and amplify it, making prejudiced decisions appear scientifically justified.[3]

A well-known illustration of this is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm, used in several U.S. jurisdictions to predict the likelihood of a defendant reoffending. Investigations revealed that the system disproportionately labelled Black defendants as high-

risk compared to white defendants, even when their actual reoffending rates were lower.[4] This example underscores a dangerous reality: AI tools can perpetuate and legitimize existing social inequalities under the guise of objectivity. As these systems become more integrated into legal decision-making, it is vital to interrogate their assumptions, demand transparency, and ensure that technology serves justice rather than distorting it.

Delegation of Moral Judgment

Criminal justice is not merely about the mechanical application of laws or strict adherence to legal codes; it is fundamentally a human enterprise, deeply rooted in judgment, empathy, and moral reasoning. While statutes and precedents provide a framework, every legal decision—particularly in sentencing—demands careful consideration of individual context. Judges must weigh multiple human factors: the intent behind a crime, the degree of remorse demonstrated by the offender, the socioeconomic background of the individual, and the broader societal consequences of the punishment. These dimensions are not easily quantifiable and cannot be fully captured by algorithms or statistical models, no matter how sophisticated.

Artificial Intelligence (AI) systems function based on data, probability, and pattern recognition. They lack consciousness, self-awareness, and the ability to comprehend abstract moral concepts such as mercy, justice, or fairness. Unlike humans, AI cannot grasp the emotional and ethical subtleties that influence legal decisions. For instance, deciding whether a person deserves rehabilitation over incarceration often depends on a judge's sense of compassion, intuition, and understanding of human behavior—traits machines do not possess.

As we increasingly integrate AI into the legal system, there is a growing risk that we might delegate critical moral responsibilities to machines. While AI can assist with tasks such as legal research, risk assessment, or identifying patterns in large datasets, it should never replace human judgment in matters involving ethical deliberation. Doing so would not only diminish the humane core of our justice system but also lead to decisions that may appear procedurally correct while being morally and socially unjust.

Moreover, the use of AI in criminal justice raises questions of legitimacy and public trust. People are more likely to accept and respect legal decisions when they believe those decisions are made by empathetic, accountable human beings who understand their lived experiences. Replacing that human element with machine-generated outcomes risks alienating the public and undermining faith in the system.

In short, while AI can play a supportive role, criminal justice must remain a domain guided by human conscience. It must continue to reflect our collective sense of morality, fairness, and compassion. Efficiency is important—but justice, at its core, is a human responsibility. It is this human touch that gives the law its legitimacy and ensures its continued relevance in a just society.

Accountability Vacuum

One of the most pressing legal challenges arising from the integration of autonomous artificial intelligence (AI) into the justice system is the complex issue of accountability. In conventional legal frameworks, when a human judge, lawyer, or official makes an error, established mechanisms such as appeals, judicial reviews, and disciplinary proceedings allow for the correction of those mistakes. Responsibility can be directly attributed to an individual or institution, ensuring transparency and fairness. However, when an AI system makes a flawed or harmful decision—such as incorrectly assessing a defendant's risk, recommending an unjust sentence, or misinterpreting evidence—the situation becomes far more complicated.

The question arises: who is to be held accountable for the error? Is it the software developer who designed the algorithm, the technology company that marketed and sold the tool, the institution that implemented it, or the individual legal professional who used the AI's output to guide their decision? This diffusion of responsibility creates a legal grey area that current laws are ill-prepared to navigate. Unlike human actors, AI systems cannot be held morally or legally liable, nor can they be subjected to the same ethical obligations.

As AI becomes increasingly embedded in decision-making processes within the legal system, this accountability vacuum could erode public trust and compromise the integrity of justice itself. Without clear statutes and regulatory frameworks to assign liability and ensure oversight, individuals affected by AI-driven errors may be left without adequate legal recourse. Addressing this accountability gap is not merely a technical or procedural issue—it is a fundamental requirement for safeguarding the principles of justice, transparency, and the rule of law. As legal technology continues to evolve, so too must our legal systems, to ensure that responsibility remains clearly defined and that human rights remain fully protected.

Legal and Ethical Implications

One of the most critical and complex legal issues in the deployment of autonomous artificial intelligence (AI) within the justice system is the question of accountability. In traditional judicial settings, when a human judge makes an error, the judgment can be challenged through appeals, and responsibility can be clearly assigned. Human actors—judges, lawyers, and institutions—are held to ethical and legal standards. However, when an AI system makes a flawed or harmful decision, accountability becomes far less clear. Who should be held liable in such cases—the coder who wrote the algorithm, the company that developed or sold the software, the government agency or institution that implemented it, or the end-user who relied on the outcome? This ambiguity presents a serious challenge to legal norms.

Current legal systems are ill-equipped to handle such scenarios, as most frameworks were designed around human agency and intention. Automated systems lack consciousness and cannot be held morally or legally accountable in the same way humans can. Without clear regulatory guidelines or liability structures, there is a risk that no one will be held responsible when errors occur. As AI tools become more deeply integrated into legal processes, resolving this accountability gap is essential to preserving public trust, legal integrity, and access to justice.

A Framework for Responsible Use

While artificial intelligence (AI) should never be granted moral authority or the power to make final legal decisions, it can still play a valuable role within the justice system when confined to strictly assistive functions. To ensure that AI supports rather than undermines human judgment, its use must be governed by robust ethical safeguards and legal frameworks. First and foremost, **transparency** is essential—algorithms must be explainable, auditable, and understandable not only to developers but also to legal professionals and the public. Black-box models that make decisions without clear reasoning have no place in systems where justice and accountability are paramount. Second, **human oversight** must be non-negotiable. AI tools should provide recommendations or support, but final decisions must always rest with trained human professionals. Third, **bias auditing** must be conducted regularly to detect and correct discriminatory patterns in algorithmic outputs. Without this, AI risks perpetuating existing inequalities under the illusion of objectivity. Lastly, meaningful **legal reforms** are urgently needed to define responsibility and assign liability when harm results from AI deployment. These reforms must clarify who

is accountable—developers, vendors, or implementing institutions. If implemented responsibly, AI can be a powerful tool for efficiency and accuracy, but it must remain firmly under human control.

Conclusion

Delegating moral judgment to machines may promise operational efficiency, consistency, and speed, but it comes at a significant cost—the erosion of foundational legal and ethical principles. Justice is not merely a mechanical application of rules; it is a deeply human endeavour grounded in empathy, moral reasoning, and social understanding. While artificial intelligence can analyse data, identify patterns, and offer predictions, it lacks the conscience, compassion, and ethical intuition that guide fair decision-making. Relying too heavily on AI risks reducing individuals to data points and stripping the legal process of its human sensitivity. The law must serve people, not statistics. As we integrate emerging technologies into our legal systems, we must remember that progress is not measured solely by efficiency but also by humanity. The future of law must embrace innovation without compromising on justice. It must remain not only intelligent and data-driven—but fundamentally humane and anchored in human values.

References

1. Andrew Guthrie Ferguson, *The Rise of Big Data Policing* (New York University Press, 2017) at 91.
2. Kate Crawford & Ryan Calo, "There is a Blind Spot in AI Research" (2016) 538 *Nature* 311.
3. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).
4. Julia Angwin et al., "Machine Bias" *ProPublica* (May 23, 2016), online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
5. Jack Balkin, "The Three Laws of Robotics in the Age of Big Data" (2015) 78:4 *Ohio St LJ* 1217.
6. Bert-Jaap Koops et al., "Bringing the Robot Within the Law" in Roger Brownsword et al. (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press, 2017) 207.
7. Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (Edward Elgar Publishing, 2015) at 82.
8. M.C. Esposito, "Ethics in the Age of Algorithms" (2020) 45 *AI & Society* 319.