AI & SOCIETY
https://doi.org/10.1007/s00146-025-02292-z

CURMUDGEON CORNER

Justice in the age of algorithms: can AI weigh morality?

Olivia Ruhil[1]

Received: 27 December 2024 / Accepted: 25 February 2025

"We can only see a short distance ahead, but we can see plenty there that needs to be done." — Alan Turing

Artificial intelligence (AI) has transitioned from futuristic speculation to a present-day disruptor in nearly every field, and the legal profession is no exception. From drafting contracts to conducting legal research, AI tools promise to revolutionise how justice is delivered. Yet, when it comes to decision-making—the sacred ground where law and morality intersect—AI encounters a profound chasm it cannot cross.

Legal systems are not merely instruments for enforcing rules; they are arenas where human morality is tested and refined. A judge deciding a criminal sentence weighs not just the facts but the intangible elements of intent, redemption, and societal impact. Courts grappling with constitutional matters must balance individual freedoms against collective welfare. These are deliberations steeped in history, empathy, and a deep understanding of context—qualities that no machine can emulate, no matter how advanced its algorithms.

Proponents of AI in law often highlight its ability to streamline labor-intensive tasks. Tools like GPT-4, Kira Systems, and ROSS Intelligence have demonstrated their prowess in areas such as contract analysis, compliance checks, and legal research. By automating these repetitive functions, AI allows lawyers to focus on higher order intellectual tasks, potentially lowering costs and increasing access to justice.

Estonia offers an intriguing case study. The country introduced AI "judges" to handle small claims disputes under €7,000, a move aimed at reducing case backlogs. These AI systems handle evidence processing and preliminary rulings before final human oversight ensures ethical integrity. The results have been promising: efficiency has improved without sacrificing fairness.

But here is the rub—efficiency is not the same as justice. The ability to process cases faster or cheaper does not address the fundamental question: Can machines truly make decisions that align with the principles of fairness, empathy, and accountability?

This brings us to the concept of the "Moral Turing Test." Building on Alan Turing's original question—"Can machines think?"—the Moral Turing Test asks: Can machines reason ethically? Can they weigh competing interests and arrive at a decision rooted in human values?

The answer, so far, is a resounding no. AI operates on historical data, and history is rife with bias. Take the case of COMPAS (Angwin et al. 2016), an algorithm used in the U.S. to predict recidivism risks. A landmark investigation by ProPublica revealed that Black defendants were disproportionately labeled as high risk compared to white defendants with similar criminal records. This was not just a glitch—it was the logical outcome of training an AI system on biased historical data.

Bias is not the only issue. Accountability is another thorny challenge. Human judges are held to a high standard of scrutiny; their decisions can be appealed, debated, and overturned. But when an AI system like COMPAS (Angwin et al. 2016) makes a flawed recommendation, who takes responsibility? The developers who coded the algorithm? The legal professionals who relied on it? Or no one at all? The opaque, "black-box" nature of many AI systems complicates this further, as even their creators may struggle to explain how a specific decision was reached.

The flaws of AI are most evident in morally complex legal scenarios. Imagine a criminal sentencing case. A judge must consider not only the severity of the crime but also the defendant's intent, remorse, and potential for rehabilitation. These factors require empathy, discretion, and a nuanced understanding of human behavior—qualities that AI inherently lacks.

Similarly, constitutional cases demand a deep appreciation of societal norms and evolving values. When courts weigh matters like freedom of speech versus national security, they engage in a dialogue with history, culture, and

✉ Olivia Ruhil
olivia.ruhil21@nludelhi.ac.in

[1] Indian Institute of Technology Delhi, New Delhi, India

Published online: 12 March 2025

ethics. AI, which relies on pattern recognition and probabilistic models, is ill equipped to navigate this interpretative terrain.

This is where the Human-in-the-Loop (HITL) framework comes into play. HITL ensures that AI systems serve as tools, not decision-makers. In Estonia's small claims courts, for example, human judges review every AI-generated recommendation. This hybrid approach allows AI to handle the heavy lifting of data analysis while humans provide the moral compass.

But even HITL is not a panacea. Over-reliance on AI can erode critical thinking skills among legal professionals, as they may come to defer to machine-generated recommendations. Ensuring that humans remain active, skeptical participants in the decision-making process is crucial.

Transparency is the cornerstone of any ethical legal system. Traditional courts operate under principles of openness, where decisions are explained and subject to public scrutiny. AI systems, however, often operate as black boxes, their decision-making processes shrouded in mystery.

This lack of transparency is not just a technical issue—it is an ethical one. Without explainability, how can we trust that an AI system is making fair and unbiased decisions? Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) aim to bridge this gap by shedding light on how AI models reach their conclusions. But explainability must go hand in hand with accountability. Regulatory frameworks, such as the European Union's AI Act (European Commission 2021), are a step in the right direction, mandating that AI systems be auditable and that mechanisms for redress be in place.

The risks posed by AI in law extend beyond bias and accountability. They touch on the very nature of justice itself. Legal systems are designed not just to resolve disputes but to reflect societal values. AI, however, is a product of its training data, unable to adapt to changing norms or recognise the intangibles of human experience.

Consider the case of predictive policing (Wexler 2017). Tools like PredPol analyse historical crime data to forecast potential crime hotspots. While these systems can optimize resource allocation, they also risk reinforcing discriminatory practices. Over-policed communities become self-fulfilling prophecies, with more arrests leading to more data that justifies further targeting.

Ethical dilemmas like these underscore the importance of robust safeguards. Regular bias audits, the diversification of training datasets, and the establishment of independent oversight bodies are not just best practices—they are imperatives.

AI is not inherently good or bad—it is a tool, and its value depends on how we wield it. The integration of AI into legal systems offers immense potential, but it also demands vigilance. Policymakers, developers, and legal professionals must collaborate to ensure that AI enhances justice rather than undermining it.

Education is a key component of this effort. Legal professionals must be trained not only in the use of AI tools but also in their limitations. They must understand that AI is a partner, not a replacement, and that the ultimate responsibility for justice lies with humans.

Public engagement is equally important. Building trust in AI-driven legal systems requires transparency, participatory governance, and open dialogue about the ethical implications of this technology.

The promise of AI in law is undeniable, but so are its pitfalls. As we stand on the cusp of a new era, we must decide what kind of justice system we want. One that prioritizes efficiency at the expense of empathy? Or one that harnesses technology to support human judgment without losing sight of the values that make justice meaningful?

Alan Turing's vision of machines that can think was bold. But the challenge before us is even bolder: to ensure that these machines, however intelligent, remain tools of humanity rather than its replacement.

The path forward is clear, if not easy. By embedding ethical safeguards, fostering human-AI collaboration, and prioritising fairness, we can chart a future where technology serves the cause of justice without compromising it. As Turing himself might remind us, there is plenty to be done—but the journey is worth it.

Acknowledgements The author extends heartfelt gratitude to professor Suma Athreye for her unwavering support, encouragement, and inspiration throughout this journey. Her belief in the author's potential has been a constant source of motivation, making everything possible.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Author contribution O.R. conceptualized the study, designed the research framework, and prepared the main manuscript text. O.R. also performed the critical analysis, interpreted the findings, and drafted all sections of the manuscript. All aspects of the manuscript were solely authored and reviewed by O.R.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica.* Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act). Retrieved from https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision-making and a "right to explanation." AI Mag. https://doi.org/10.1609/aimag.v38i3.2741

Guidotti R, Monreale A, Turini F, Pedreschi D, Giannotti F, Ruggieri S (2018) A survey of methods for explaining black-box models. ACM Comput Surv. https://doi.org/10.1145/3236009

Hildebrandt M (2016) Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics. Univ Toronto Law J. https://doi.org/10.3138/utlj.2017-0044

Rudin C (2019) Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nat Machine Intell. https://doi.org/10.1038/s42256-019-0048-x

Wexler, R. (2017). Life, liberty, and trade secrets: Intellectual property in the criminal justice system. *Stanford Law Review, 70*(6), 1343–1429. Retrieved from https://www.stanfordlawreview.org/print/article/life-liberty-and-trade-secrets/