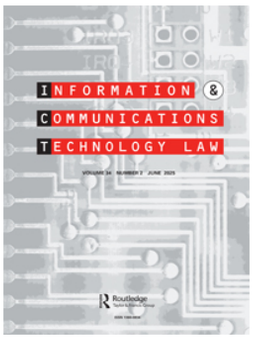

Routledge
Taylor & Francis Group

Information & Communications Technology Law

ISSN: 1360-0834 (Print) 1469-8404 (Online) Journal homepage: www.tandfonline.com/journals/cict20

The rule of law or the rule of robots? Nationally representative survey evidence from Kenya

Brian Flanagan, Guilherme Almeida, Daniel Chen & Angela Gitahi

To cite this article: Brian Flanagan, Guilherme Almeida, Daniel Chen & Angela Gitahi (16 Jul 2025): The rule of law or the rule of robots? Nationally representative survey evidence from Kenya, Information & Communications Technology Law, DOI: 10.1080/13600834.2025.2533042

To link to this article: https://doi.org/10.1080/13600834.2025.2533042

Published online: 16 Jul 2025.

INFORMATION & COMMUNICATIONS TECHNOLOGY LAW
https://doi.org/10.1080/13600834.2025.2533042

Routledge
Taylor & Francis Group

The rule of law or the rule of robots? Nationally representative survey evidence from Kenya

Brian Flanagan[a], Guilherme Almeida[b], Daniel Chen[c] and Angela Gitahi[d]

[a]Maynooth University, Maynooth, Ireland; [b]Insper Institute, São Paulo, Brazil; [c]Toulouse School of Economics, Toulouse, France; [d]McGill University, Montreal, Canada

ABSTRACT

With AI now passing the bar, and with increasing court caseloads worldwide hampering access to justice, there are calls for judges to make use of chatbots to help expedite their work. Such calls pose a normative question: whether our ideal of the rule of law is consistent with judicial reliance on computer generated legal research. In deciding whether artificial intelligence could support the administration of justice in this way, the views of those who stand to gain the most through more readily available dispute resolution will be critical. Collecting nationally representative survey data from Kenya, we report a vignette-based experiment on the acceptability of AI law clerks – assistants whose legal analysis does not decide what the law says but which informs the ultimate decision. We find that an AI's influence on the law's application is seen as no less legitimate than that of a human assistant. This result spurs efforts to systematically investigate whether the integration of AI might make justice systems more efficient, accessible, and trustworthy in practice.

KEYWORDS

Generative AI; legal AI; robot law clerks; legitimacy; vignette survey experiment

> The day should come … when you will be able to feed a set of facts to a machine … and … the machine can then lay out for you … the reasoning process by which you may be able to arrive at a [legal] conclusion. (Reed Lawlor, Law and Electronics Conference, California, 1960)[1]

> No electronic magician could design a computer program that would supply a verdict everyone would accept once the facts of the case along with the text of all past statutes and judicial decisions were put at the computer's disposal. (Ronald Dworkin, Law's Empire, 1986)[2]

Theorists have been elucidating the normative stakes of employing machines to resolve legal disputes since the 1960s.[3] Now, with the explosion in machine learning capability, the question is receiving renewed attention. Until recently, most would have agreed with Dworkin's insistence that human judgment is a prerequisite of competent legal

CONTACT Brian Flanagan brian.flanagan@mu.ie

[1]Reed Lawlor, 'Excerpts from Fact Content of Cases and Precedent – A Modern Theory of Precedent.' [1960] (1972) 12(4) *Jurimetrics Journal* 245.

[2]Ronald Dworkin, *Law's Empire* (Harvard University Press, 1986) 412.

[3]For example, Reed Dickerson, 'Some Jurisprudential Implications of Electronic Data Processing' (1963) 28 *Law & Contemporary Problems* 53; Bruce Buchanan and Thomas Headrick, 'Some Speculation about Artificial Intelligence and Legal Reasoning' (1970) 23(1) *Stanford Law Review* 40.

reasoning. Every situation to which any rule is applied is unique in ways that might – or might not – seem decisive. For all their ingenuity and value, neither expert computer programmes, which adhere to a pre-set decision tree, nor self-learning programmes, which identify salient patterns across legal decisions, can be relied upon to correctly resolve a disputed question of law. With the creation of the Large Language Model (LLM), however, humans' monopoly on the appreciation of legal relevance may have ended. Now, as 'judges, judicial support staff, prosecutors, and lawyers around the globe ... start[] to use chatbots … to draft … judicial decisions, and elaborate arguments',[4] we might have to recognize Lawlor's prescience and acknowledge that the electronic magician has produced her masterpiece.

The paradox of law is that while the complexity of modern society depends on it, law's own complexity limits its accessibility. Knowledge of the law's intricacies is not expected even of those who enact it. For the law to have real purchase in social life, it has been essential that 'there be a competent profession available to offer … advice … [as to] what the law at any given time requires'[5] – and, qua judges, to resolve disagreements over its application. The problem of legal knowledge is posed most directly by litigation. To dispose of cases, judges must master the details of materials which 'must unavoidably swell to a very considerable bulk, and must demand long and laborious study to acquire a competent knowledge of them'.[6] There is always a possibility that judges' research burden will slow the administration of justice to a point where the threat of successful legal action diminishes, and law's promise is undermined. With the recent step-change in information technology, however, a solution to bottlenecks caused by shortages in legal research labour might soon be at hand. The rapid development of generative artificial intelligence has prompted the suggestion that '[j]udge AI has the potential to increase access to justice'[7] and has challenged the assumption that law's complexity and accessibility are inversely related.

Some theorists foresee a 'legal singularity' in which the use of algorithms enables society to establish exponentially more efficient and cheaper litigation services, including computer judges.[8] Conversely, other theorists identify a tension between our commitment to the rule of law and the relinquishment of human responsibility for the individual's legal fate.[9] Drawing on this debate, we take as our starting point the premise that 'technological infrastructures matter, require our attention and must somehow be brought under the Rule of Law'.[10]

[4]Juan David Gutiérrez, 'UNESCO Global Judges' Initiative: Survey on the Use of AI Systems by Judicial Operators' (2024) *UNESCO* 1, 6 <https://unesdoc.unesco.org/ark:/48223/pf0000389786>.

[5]Jeremy Waldron 'The Rule of Law' (2020) *Stanford Encyclopedia of Philosophy* (Summer Edition) <https://plato.stanford.edu/archives/sum2020/entries/rule-of-law/>.

[6]Alexander Hamilton, 'The Federalist No. 78' in Jack Richon Pole (ed), *The Federalist* (Hackett 2005) 418.

[7]Tania Sourdin, 'Robo Justice: Constitutional Issues with Judge AI' (2023) 30(2) *Indiana Journal of Global Legal Studies* 293, 321.

[8]For example, Abdi Aidid and Benjamin Alarie, *The Legal Singularity: How Artificial Intelligence Can Make Law Radically Better* (University of Toronto Press, 2023); Jakub Harasta, Tereza Novotná and Jaromir Savelka, 'It cannot be right if it was written by AI: on lawyers' preferences of documents perceived as authored by an LLM vs a human' (2024) *Artificial Intelligence and Law* <https://doi.org/10.1007/s10506-024-09422-w>; Eugene Volokh, 'Chief Justice Robots' (2019) 68 *Duke Law Journal* 1135; Anthony Casey and Anthony Niblett, 'A Framework for the New Personalization of Law' (2019) 86 *University of Chicago Law Review* 333.

[9]See Timothy Endicott and Karen Yeung, 'The Death of Law? Computationally Personalized Norms and the Rule of Law' (2022) 72(4) *University of Toronto Law Journal* 373.

[10]Mireille Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (2016) 79(1) *Modern Law Review* 1, 2. See also Andrew Coan and Harry Surden, 'Artificial Intelligence and Constitutional Interpretation' (2024) *Arizona Legal Studies Discussion Paper No. 24–30* <https://ssrn.com/abstract=5018779>.

AI's new legal capacities present two pressing practical and normative questions: would the integration of legal LLMs make justice systems more efficient and accessible to ordinary people?; and, crucially, would their integration be normatively acceptable? Although robust answers to the first question must await large scale field research, we know that individual judges and their staff are already integrating LLMs into their work. In this context, it is high time to examine whether the judicial use of such technology is considered legitimate by those whom the justice system is intended to serve:

> Unlike in the private sector, where customers can easily cease using a service if dissatisfied, citizens lack alternatives to approaching public administration … Moreover, the imposition of technological solutions upon citizens not only proves ineffective but also contradicts with the principles of democracy, ultimately leading citizens to abandon these solutions.[11]

We know that the quality of procedural justice predicts 'how people evaluate … the court system and the law'[12]; if we are to successfully elucidate citizens' assessment of the legal system in this new era of AI, we must investigate the perceived conformity with procedural justice of judicial reliance on computer-generated legal analysis.[13]

We report a vignette-based study on the legitimacy of robot law clerks. Collecting nationally representative survey data from Kenya, we find that judges' reliance on AI-produced legal research is seen as no less legitimate than judges' reliance on that produced by a human assistant.

Our paper begins with a brief review of the development of artificial intelligence as a means of legal interpretation (Part 1), before discussing its potential to provide a law clerk for busy judges (Part 2). Part 3 presents our method for analysing the legitimacy of robot law clerks. We then report our study's results (Part 4), and discuss their significance, limitations, and implications for future research (Part 5).

Part 1. Artificial legal interpretation

In its first phase, artificial legal interpretive systems attempted to recreate human propositional legal knowledge (knowledge 'that … '.). Known as 'expert' systems, complex decision trees were constructed that reproduced in logical form the content of extant or possible texts, e.g. statutes or legal textbooks/treatises. One might describe such systems as schematized-human-learning: they 'freeze the meaning of the rule so that its general terms must have the same meaning in every case where its application is in question'.[14]

[11]Saja Aljuneidi and others,, 'Why the Fine, AI? The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations' (2024) *Association for Computing Machinery Proceedings of the CHI Conference on Human Factors in Computing Systems* 1, 5. See also Patrick Stewart Hodge, 'The Law and AI: Where are we going?' (Lecture at De Montfort University in Leicester 2023) 22 <https://supremecourt.uk/uploads/speech_231130_336aa18930.pdf>; Margaret Hagan, 'Towards Human-Centred Standards for Legal Help AI' (2024) 382(2270) *Philosophical Transactions of the Royal Society A* 1, 2 <https://doi.org/10.1098/rsta.2023.0157>.

[12]Tom Tyler, 'Court Review: Procedural Justice and the Courts' (2007) 44(1/2) *The Journal of the American Judges Association* 26.

[13]Gizem Yalcin and others, 'Perceptions of Justice By Algorithms' (2023) 31 *Artificial Intelligence and Law* 269, 270 <https://doi.org/10.1007/s10506-022-09312-z>; Dovilė Baryse and Roee Sarel, 'Algorithms in the Court: Does it Matter Which Part of the Judicial Decision-Making is Automated?' (2024) 32 *Artificial Intelligence and Law* 117, 134 <https://doi.org/10.1007/s10506-022-09343-6>.

[14]H.L.A. Hart, *The Concept of Law* (Clarendon Press 1961) 270.

These systems face two significant constraints: that a considerable and expensive human effort is required to explicitly and comprehensively formulate what the law says about most things, and that any such formulation is invariably 'defeasible'.[15] Defeasibility refers to the impediment that, no matter how elaborate, no text can specify all intuitively exonerating circumstances: 'it is not possible to conclude legal clarity from semantic clarity'.[16] Every legal text is liable to have a literal application that diverges from that of the law which it seeks to transcribe; in the right circumstances, every legal text is intuitively 'defeated'.

With no way of knowing in advance whether a particular literal application will appear incorrect, it is unclear how any weight might be placed on a letter-only verdict – whether reached by man or machine: 'deduction cannot provide an adequate model of legal reasoning'.[17] So, while schematized-human-learning systems have been made available commercially to consumers in certain legal domains, notably, in that of tax law,[18] they face important limitations in modelling how humans apply laws. Efforts to advance legal AI soon sought to take advantage of the possibilities of *machine* learning.

Whereas legal AI had initially focused on replicating human propositional knowledge, in its second phase of development, inductive, pattern-spotting reasoning was emphasized. Self-learning algorithms were designed that can be trained on an existing body of discrete rule applications to predict how subsequent cases would be decided in light of what they have learned about the fact combinations characteristic of the training data. Unlike expert systems, specialist machine learning systems (SML) identify for themselves the connections within an existing legal corpus. Accordingly, they have the advantage of potentially avoiding counterintuitive outcomes for which no textually specified exception had been formulated.

SML exemplifies the Wittgensteinian legal theory that 'the meaning of [legal] rules, like those of all symbols, must be determined by the actions themselves, that is, by the way the rules are used'.[19] It proved that some initial doubts, e.g. that 'deep structure is to be found in social context and purpose, which are non-computational',[20] were premature. But while the problem of defeasibility was mitigated, it was not yet solved. A case might always include a novel, intuitively exonerating characteristic to which the algorithm is blind for want of any opportunity to learn of its legal salience.[21] In the application of newly or recently posited provisions, the body of caselaw on which an SML might be trained will be limited, increasing the risk that its induction will be 'defeated' in this

[15]Neil MacCormick, 'Defeasibility in Law and Logic' in Zenon Bankowski, Ian While and Ulrike Hahn (eds), *Informatics and the Foundations of Legal Reasoning* (Kluwer, 1995) 99.
[16]Matthias Klatt, *Making the Law Explicit* (Hart, 2008) 219. See also Frank Pasquale, 'A Rule of Persons, Not Machines: The Limits of Legal Automation' (2019) 87(1) *George Washington Law Review* 1, 48.
[17]Trevor Bench-Capon, Henry Prakken and Giovanni Sartor, 'Argumentation in Legal Reasoning' in Iyad Rahwan and Guillermo Simari (eds), *Argumentation in Artificial Intelligence* (Springer 2009) 17.
[18]George Contos and others, 'Individual Taxpayer Compliance Burden: The Role of Assisted Methods in Taxpayers Response to Increasing Complexity' in Martha Gangi and Alan Plumley (eds), *IRS Research Bulletin: Proceedings of the IRS Research Conference* (Internal Revenue Service, 2010).
[19]Andrei Marmor, *Interpretation and Legal Theory* (Hart, 2005) 115.
[20]Andrew Greinke, 'Legal Expert Systems: A Humanistic Critique of Mechanical Legal Interface' (1994) 1(4) *Murdoch University Electronic Journal of Law*.
[21]See Harry Surden, 'Machine Learning and Law' (2014) 89 *Washington Law Review* 87, 105; Mazviita Chirimuuta, 'Rules, judgment and mechanisation' (2023) 1(3) *Journal of Cross-disciplinary Research in Computational Law* 1, 14 <https://journalcrcl.org/crcl/article/view/22>.

way.[22] Again, with no way of knowing ex ante whether a particular case will feature such a characteristic, it is unclear how a judge might ever responsibly rely on an SML's outcome suggestion.

The overarching challenge of defeasibility is that traditional AI systems – whether expert or SML – are liable to overlook a situation's legal novelty in ways a human never would: they 'suck[] up the dust and the crickets' alike.[23] Neither system can produce analysis that could provide a judge with the same kind of assurance as that of a qualified human clerk. With the advent of large language models, however, the disparity in human and machine legal reasoning has receded.

A large language model also relies on machine learning. LLMs learn to accurately predict the next token (which can be conceptualized as a word) in a series, allowing them to answer questions, e.g. has X a legal right to Y? Rather than being trained on a specialized body of knowledge, such as on a set of legal precedents, state of the art LLMs are trained on a general language corpus. Their capabilities represent a revolution in AI: there is evidence that humans struggle to distinguish conversations with the best known example, GPT4, from conversations with human interlocutors, i.e. that it passes the Turing Test.[24] Similarly, in the legal domain, LLMs have achieved key markers of professional human competence, such as comfortably passing the US Uniform Bar Exam[25] (if not yet topping the class),[26] while evidence of their capacity across an array of discrete legal tasks has steadily accumulated.[27] Perhaps the most remarkable feature of LLM legal aptitude, however, is its ability to favour the law's spirit over the text of the relevant legal materials.

Recent studies have compared how humans and LLMs resolve 'hard' cases, in which the law's text and stated purpose diverge, and for which no precedents are available by which the interpreter might learn to associate certain characteristics with counter-literal outcomes.[28] LLMs reproduced humans' propensity to (a) sometimes prioritize a law's spirit over its letter, and (b) do so especially where the law's purpose is benevolent. Like the human judge, LLMs were found to recognize a case's inclusion of a novel, intuitively exonerating characteristic on whose legal salience it had not already been trained.

[22]Brian Flanagan, 'Revisiting the Contribution of Literal Meaning to Legal Meaning' (2010) 30(2) *Oxford Journal of Legal Studies* 255, 262-63.

[23]Ric Simmons, 'Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System' (2018) 52 *University of California Davis Law Review* 1067, 1095.

[24]Cameron Jones and Benjamin Bergen, 'Does GPT-4 Pass the Turing Test?' <https://arxiv.org/abs/2310.20216>.

[25]See Daniel Martin Katz and others, 'GPT-4 Passes the Bar Exam' (2024) 382(2270) *Philosophical Transactions of the Royal Society A* 1 <https://doi.org/10.1098/rsta.2023.0254>

[26]Eric Martínez, 'Re-evaluating GPT-4's Bar Exam Performance. Artificial Intelligence and the Law' (2024) *Artificial Intelligence and Law* <https://doi.org/10.1007/s10506-024-09396-9>.

[27]Neel Guha and others 'Legalbench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models' (2023) *Osgoode Legal Studies Research Paper No. 4583531* <http://dx.doi.org/10.2139/ssrn.4583531>; Camilla Bignotti and Carolina Camassa, 'Legal Minds, Algorithmic Decisions: How LLMs Apply Constitutional Principles in Complex Scenarios' (2024) <https://doi.org/10.48550/arXiv.2407.19760>; Armin Alimardani, 'Generative Artificial Intelligence vs. Law Students: An Empirical Study on Criminal Law Exam Performance' (2024) 16(2) *Law, Innovation and Technology* 777 <https://doi.org/10.1080/17579961.2024.2392932>; Yonathan Arbel and David Hoffman, 'Generative Interpretation' (2024) 99 *New York University Law Review* 451; Guillaume Zambrano, 'Case Law As Data : Prompt Engineering Strategies for Case Outcome Extraction With Large Language Models in a Zero-Shot Setting' (2024) 6(3) *Law, Technology and Humans* 80; Sascha Schweitzer and Markus Conrads, 'The Digital Transformation of Jurisprudence: An Evaluation of ChatGPT-4's Applicability to Solve Cases in Business Law. (2024) *Artificial Intelligence Law* <https://doi-org.may.idm.oclc.org/10.1007/s10506-024-09406-w>.

[28]Guilherme Almeida and others,'Exploring the Psychology of LLMs' Moral and Legal Reasoning' (2024) 333(104154) *Artificial Intelligence* 1 <https://doi.org/10.1016/j.artint.2024.104145>.

LLM capability is not unlimited, however; rather, it exhibits a 'jagged frontier', where performance can vary sharply across 'tasks that appear [to humans] to be of similar difficulty'.[29] LLMs' chief limitation in the legal context is their well-documented tendency to hallucinate,[30] i.e. to confidently include false facts in their responses. Notoriously, a US-based law firm was fined $5,000 after a lawyer submitted a legal brief which cited non-existent precedents; the cases were ChatGPT hallucinations.[31] Nevertheless, with the LLM, it seems that legal AI may finally have overcome the challenge of defeasibility and thereby transcended a historic barrier to reliability, namely, that '[c]omputers have not yet been programmed … to display the … intuition [and] common-sense … that we, as human beings, expect … of judges acting in their official role'.[32]

Unlike traditional social practices or negotiated orders, law has the essential characteristic of being based on sets of texts.[33] In this sense, law is the prime text-deploying social technology.[34] Conversely, the LLM is an information technology that can create text. With their kindred connection to language, it stands to reason that AI might be positioned to amplify the reach and utility of law. Now, as AI's interpretive response to potentially exonerating case features tracks that of humans, it is no longer true that '[h]uman judges and other persons charged with interpreting legal texts reason in ways that … remain over the horizon of machine capacities'[35] So it is that, in approaching functional parity with human lawyers, LLMs have placed robot generated legal interpretation firmly on the agenda. With global law firms purchasing chatbot systems to support their advisory services,[36] the case for investigating AI's implications for the perception of *adjudicative* systems is clear.

Part 2. Artificial law clerks

A law clerk is a member of a court's staff whose function is to facilitate judges in the performance of their duties, notably, by producing research memos and by assisting them with opinion writing. The employment of law clerks has long been understood as the

[29]Fabrizio Dell'Acqua and others, 'Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality' (2023) *Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24–013* <https://ssrn.com/abstract=4573321>.

[30]See Joshua Maynez and others, 'On Faithfulness and Factuality in Abstractive Summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics' (2020) *Association for Computational Linguistics* 1906; Yue Zhang and others, 'Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models' (2023) <https://doi.org/10.48550/ARXIV.2309.01219>; Matthew Dahl and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' (2024) (16)(1) *Journal of Legal Analysis* 64; Varun Magesh and others, 'Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools' (2024) <https://arxiv.org/abs/2405.20362>

[31]Sara Merken, 'New York lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief' *Reuters* (New York, 26 June 2023) <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/>.

[32]Richard Susskind, 'Review: Detmold's Refutation of Positivism and the Computer Judge' (1986) 49(1) *Modern Law Review*, 125, 133. See also Gerald Postema, Law's Rule (Oxford University Press, 2022) 302.

[33]Fernanda Pirie, 'Beyond Pluralism: A Descriptive Approach to Non-state Law' (2023) 14(1) *Jurisprudence* 1.

[34]Orion Lewis and Sven Steinmo, 'How Institutions Evolve: Evolutionary Theory and Institutional Change' (2012) 44(3) *Polity* 314.

[35]Michael Livermore, 'Rule by rules' in Ryan Whalen (ed), *Computational Legal Studies: The Promise and Challenge of Data-Driven Research* (Edward Elgar, 2020) 239. See also Phillip Sales, 'Algorithms, Artificial Intelligence and the Law, Judicial Review' (2020) 25(1) 46, 50; Richard Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22(2) *Stanford Technology Law Review* 242.

[36]Cristina Criddle, 'Law Firms Embrace the Efficiencies of Artificial Intelligence' *Financial Times* (London 4 May 2024) <https://www.ft.com/content/9b1b1c5d-f382-484f-961a-b45ae0526675>.

judiciary's response to growing caseloads.[37] In line with this assumption, research indicates that increasing the number of clerks available to judges 'help[s] previously underperforming courts in disadvantaged locales to achieve more efficient outcomes'.[38] In principle, AI might supply an abundance of law clerk assistance with comparable potential to 'offer[] valuable insights that assist judges in writing judgments'.[39]

Rather than acting as judge, AI might provide important adjudicative support either by extracting relevant precedents, as happens already in China,[40] or by recommending an outcome for stated legal reasons.[41] Significantly, a study has found that the use of ChatGPT reduces the time spent by humans on drafting advisory legal memos[42] – just as the employment of a human clerk might be expected to do. Judges have noticed the possibilities:

> I asked ChatGPT can you give me a summary of this area of law … and I put it in my judgment. Lord Justice Colin Birss, Court of Appeal of England and Wales, UK 2023.[43]

> [Judges] should consider whether and how AI-powered large language models … might … inform the interpretive analysis. Judge Kevin Newsom, Court of Appeal 11th Circuit, USA, 2024.[44]

> These points formulated by the ChatGPT-4 are very impressive as our law of Civil Procedure, developed over the years has also guided the courts to deliberate on these dimensions while deciding such an application, if circumstances so justify. Judge Amir Munir, Additional District & Sessions Court, Pakistan, 2023.[45]

Credentialed professionals are, of course, a finite, costly resource, which, by comparison, AI is not. Moreover, there are many countries whose judiciaries face significant case backlogs and for whom a low-cost clerk substitute would presumably present considerable time savings in both research and writing.[46] Kenya is one such country.

Facing a 500,000 case backlog, Kenyan judicial authorities have expressed openness to technological innovation.[47] Notably, Kenya has trialled an AI-based case management

[37]For example, Richard Posner, *The Federal Courts: Crisis and Reform* (Harvard University Press, 1985); Jonathan Cohen, 'In the Shadow of the Law Clerk: Assessing the Roles of Law Clerks in the Judicial Process' (1995) 3 *Long Term View* 99; Kermit Lipez, 'Judges and Their Law Clerks: Some Reflections' (2007) 22 *Maine Bar Journal* 112.

[38]Judson Peverall, 'Inside State Courts: Improving the Market for State Trial Court Law Clerks' (2020) 55 *University of Richmond Law Review* 227, 280.

[39]Beenish Chaudhary, Patricia Covarrubia and Gar Yein Ng, 'The Judge, the AI, and the Crown: A Collusive Network' (2024) 33(3) *Information & Communications Technology Law* 330, 349.

[40]Jinting Deng, 'Should the Common Law System Be Intelligentized?: A Case Study of China's Same Type Case Reference System' 2019 3(2) *Georgetown Law Technology Review* 223.

[41]Ernest Lim, 'Law by Algorithm' (2023) 43(3) *Oxford Journal of Legal Studies* 650.

[42]Jonathan Choi and Daniel Schwarcz, 'AI Assistance in Legal Analysis: An Empirical Study' (2025) 73(2) *Journal of Legal Education* 384 <https://jle.aals.org/home/vol73/iss2/5/>; Jonathan Choi, Amy Monahan and Daniel Schwarcz, 'Lawyering in the Age of Artificial Intelligence' (2024) 109 *Minnesota Law Review* `147. <http://dx.doi.org/10.2139/ssrn.4626276>.

[43]Colin Birss in Bianca Castro and John Hyde, 'Solicitor condemns judges for staying silent on 'woeful' reforms' *The Law Society Gazette* (London 14 September 2023) <https://www.lawgazette.co.uk/news/solicitor-condemns-judges-for-staying-silent-on-woeful-reforms/5117228.article>.

[44]Kevin Newsom, Snell v. United Speciality Insurance Company (2024) No. 22–12581 11th Circuit US Court of Appeals <https://www.govinfo.gov/app/details/USCOURTS-ca11-22-12581/context>.

[45]Amir Munir in TLTP, ,Additional District Judge uses ChatGPT tool to strengthen court decision' *Pakistan Today* (Pakistan 15 April 2025) <https://www.pakistantoday.com.pk/2023/04/15/additional-district-judge-uses-chatgpt-tool-to-strengthen-court-decision/>.

[46]See Richard Susskind, *Online Courts and the Future of Justice* (Oxford University Press, 2019) 287; Fotios Spyropoulos and Evangelia Androulaki, 'Aspects of Artificial Intelligence on e- Justice and Personal Data Limitations' (2023) 26(3) *Journal of Legal, Ethical and Regulatory Issues* 1, 6; Coan and Surden (n 10) 6.

[47]Nancy Gitonga, 'Justice Denied as Over 520,000 Cases Pend in Courts for Years' *People Daily* (Keyna, 4 March 2024) <https://peopledaily.digital/news/justice-denied-as-over-520000-cases-pend-in-courts-for-years>.

system with which to better distribute existing judicial capacity.[48] In principle, the assistance of an AI-based legal research service might increase the capacity of individual judges. Given the association between slow justice delivery, an adverse business environment, and significant economic and welfare consequences,[49] the case for investigating this possibility is pressing. The starting point must be consideration of the artificial law clerk's normative acceptability, a matter on which existing research has had relatively little to say: '[We have reached] a moment when Assistive Judge AI is proliferating, and not enough consideration has been given to how this impacts normative ideals of how justice should be done'.[50]

There is wide agreement that rule systems should be evaluated by reference to how well they serve the value of legality, i.e. by how faithfully they adhere to the ideal of the rule of law. Theories of the nature of the rule of law articulate various procedural or formal principles, through which law is made and administered. The most prominent effort is that of Lon Fuller, who set out seven, mostly legislative principles, to wit, that statutes must be (i) consistent, (ii) enforced according to their terms, (iii) general in application, (iv) intelligible, (v) directed towards prescribing conduct that is possible, (vi) prospective in application, (vii) stable over time, and (viii) publicly announced.[51] Replicating an initial study by Donelson and Hannikainen (2020), Ivar Hannikainen and colleagues found that, across diverse cultures and linguistic communities, majorities of survey participants commonly recognized Fuller's desiderata as principles to which legal systems ought to adhere.[52] Notably, none of the aforementioned legislative principles would forbid legal AI. Potentially, LLMs could facilitate the application of statutes that are stable, consistent, public, clearly written, and prospective in a manner that is consistent with their terms and which does not demand the impossible. Rather, the issue is whether such a role would respect the rule of law's *adjudicative* dimension, specifically, the principle that litigants should receive normatively acceptable reasons.[53] Must legal reasons reflect purely human insights to be acceptable – or can the law be duly applied with the assistance of AI-generated legal research?

Unlike a human, a robot has nothing at all at stake in adjudicating someone's rights – neither the retention of their job, the advancement of their promotion prospects nor the preservation of their reputation. They cannot be held accountable in any meaningful way for what they decide to be the law's meaning. Arguably, such accountability is crucial to the judicial function: 'Judgments are decisions for which the decision maker has basic responsibility … '.[54] Perhaps for this reason, even advocates concede that the legitimacy

[48]Matthieu Chemin and others, 'Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya' (2024) *TSE Working Papers 22–1391* <https://ideas.repec.org/p/tse/wpaper/127593.html>.

[49]Dani Rodrik, 'Institutions for High-Quality Growth: What They are and How to Acquire Them' (2000) *National Bureau of Economic Research Working Paper 7540* <http://www.nber.org/papers/w7540>; Daron Acemoglu, Simon Johnson and James Robinson, 'The Colonial Origins of Comparative Development: An Empirical Investigation' (2001) 91(5) *American Economic Review* 1369 <https://www.aeaweb.org/articles?id=10.1257/aer.91.5.1369>.

[50]Brian Barry, 'AI for Assisting Judicial Decision-Making: implications for the Future of Open Justice' (2024) *Australian Law Journal* 98 656.

[51]Lon Fuller, *The Morality of Law* (Yale University Press, 1964) 96–97.

[52]Ivar Hannikainen and others, 'Are There Cross-Cultural Legal Principles? Modal Reasoning Uncovers Procedural Constraints on Law' (2021) Cognitive Science 45(8) e13024.

[53]See Lawrence Solum, 'Procedural Justice' (2004) 78 *Southern California Law Review* 181; Antonino Rotolo and Giovanni Sartor, 'Argumentation and Explanation in the Law' (2023) 6 *Frontiers in Artificial Intelligence* <https://doi.org/10.3389/frai.2023.1130559>.

[54]Endicott and Yeung (n 9) 398. See also W. Bradley Wendel, 'The Promise and Limitations of Artificial Intelligence in the Practice of Law' (2019) 72(1) *Oklahoma Law Review* 21, 42; J Tasioulas, 'The Rule of Algorithm and the Rule of Law' in Christoph Bezemek, Michael Potacs and Alexander Somek (eds), *Vienna Lectures on Philosophy* (Bloomsbury, 2023) 17, 35.

of AI judges is 'likely to be … counterintuitive'.[55] Law clerks might be a different matter, however.

Consider the normative question on which the deployment of artificial law clerks would seem ultimately to depend: whether, in discharging their duty to provide a reasoned decision, it is any less legitimate for judges to rely on the legal analysis of an inexpensive robot as it is for them to rely on the analysis of a qualified but costly human? For some theorists, robot clerks would be unproblematic: 'AI judicial staff attorneys that draft proposed opinions for judges to review – would be … legitimate'.[56] There remain sceptics, however, for whom *any* interpretive reliance on machines would risk a 'legitimacy deficit' that would pose 'a real threat to social wellbeing'.[57]

Would an artificial law clerk be considered consistent with the value of legality or would it, instead, be thought to bespeak a rule of robots? It is true that, on issues of resource allocation, such as hiring or university admission,[58] there is increasing evidence that people prefer decisions to be made by a human rather than an algorithm. Equally, on the question of AI's resolution of the factual aspects of legal disputes, that is, of disagreement as to the relevant circumstances, there are indications that machine judges are considered less legitimate than human ones.[59] One might imagine that the core adjudicative activity of legal interpretation – of deciding what the law says – would engage the value of legality even more directly. Indeed, in his 2023 annual report, US Chief Justice John Roberts described 'a persistent public perception of a "human-AI fairness gap", reflecting the view that human adjudications, for all of their flaws, are fairer than whatever the machine spits out'.[60] The evidence is mixed. Initial research on US public opinion appears to bear out Roberts' description,[61] while research in Germany, Lithuania and the Netherlands reports a more positive perception of AI's possible judicial role.[62] Crucially, were AI to merely assume the role of a law clerk rather than judge, ultimate interpretive responsibility would remain in human hands. Perhaps, in this context, AI's acceptability would simply depend on its perceived legal expertise: 'It is … thanks to … the quality of their reasoning – that [courts'] rulings enjoy authority and

[55]Volokh (n 8) 1142.
[56]ibid 1141.
[57]Adrian Zuckerman, 'Artificial Intelligence in the Administration of Justice' in Andrew Higgins (ed), *The Civil Procedure Rules at 20* (Oxford University Press 2020) 304. See also Xavier Rodriguez, 'Artificial Intelligence (AI) and the Practice of Law' 2023 24(2) *Sedona Conference Journal* 783; Juan David Gutiérrez, 'Critical Appraisal of Large Language Models in Judicial Decision-Making' in Regine Paul, Emma Carmel and Jennifer Cobbe (eds), *Handbook on Public Policy and Artificial Intelligence* (Elgar, 2024); Henrique Marcos, 'Can Large Language Models Apply the Law?' (2024) *AI & Society* <https://doi.org/10.1007/s00146-024-02105-9>; Inyoung Cheong and others, '(A)I Am Not a Lawyer, But … : Engaging Legal Experts towards Responsible LLM Policies for Legal Advice' (2024) *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* <https://doi.org/10.1145/3630106.3659048>; Amin Ebrahimi Afrouzi, 'John Robots, Thurgood Martian, and the Syntax Monster: A New Argument Against AI Judges' (2024) 37(2) *Canadian Journal of Law & Jurisprudence* 369.
[58]David Newman, Nathanael Fast and Derek Harmon, 'When Eliminating Bias Isn't Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions' (2020) 160 Organizational Behavior and Human Decision Processes 149; Marius Claudy, Karl Aquino and Maja Graso, 'Artificial Intelligence Can't Be Charmed: The Effects of Impartiality on Laypeople's Algorithmic Preferences' (2022) 13 *Frontiers in Psychology* 1.
[59]Benjamin Minhao Chen, Alexander Stremitzer and Kevin Tobia, 'Having your Day in Robot Court' (2022) 36(1) *Harvard Journal of Law & Technology* 127.
[60]John Roberts, '2023 Year-End Report on the Federal Judiciary' (2023) 6 <https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf.>.
[61]Yalcin (n 13).
[62]Aljuneidi (n 11); Dovilė Baryse, 'People's Attitudes towards Technologies in Courts' (2022) 11(5) *Laws* 71 <https://doi.org/10.3390/laws11050071>; Theo Araujo and others, 'In AI We Trust? Perceptions About Automated Decision-Making by Artificial Intelligence' (2020) 25 *AI & Society* 611 <https://doi.org/10.1007/s00146-019-00931-w>.

legitimacy'.[63] The salience of the distinction between AI as judge and as clerk is evinced by research on Portuguese judges, a representative sample of whom were reported to express scepticism about the former combined enthusiasm for the latter.[64] More generally, there is evidence that, when it comes to *advice* rather than *decision*, people may actually prefer to rely on that provided by an algorithm over that provided by a human.[65] Accordingly, we hypothesized that judges would be perceived as being no less legitimately guided by computer-generated legal research as by research produced by a human clerk.

Part 3. Method

We investigate the legitimacy of chatbot law clerks by means of a vignette experiment. Conducting surveys of perceptions of the legitimacy of real-world institutions, whether those of citizens,[66] of litigants,[67] or of judges,[68] is limited by the fact that, to our knowledge, outside China, no country has so far adopted this innovation. Moreover, any correlation between opinions of an institution's legitimacy and an institutional reform might be influenced by third common causes. On the other hand, theorists warn against the dangers of abstract questions that require ordinary people to synthesize their conceptual knowledge,[69] such as asking whether the rule of law demands exclusive reliance on human legal analysis. With a vignette study, we avoid these limitations. Testing participants' responses to scenarios, such a study documents participants' *practical* application of their concept of legitimacy. Equally, experimental manipulation across vignettes offers the advantage of a robust basis from which to infer whether the variable of interest causes a change in participants' assessments.[70]

To achieve a fine-grained understanding of participants' attitudes, the experimental condition must be carefully selected. A study might alternatively refer to a general decisional practice[71] or to a specific decisional consideration.[72] In this context, the latter approach promises more readily interpretable data. A study which manipulated whether the court had a practice of making use of AI assistance might prompt opinions about the virtue of innovation and judicial reform in general or perhaps about the

[63]Koen Lenaerts, 'New Horizons for the Rule of Law Within the EU' (2020) 21 *German Law Review* 29.
[64]Andreia Martinho, 'Surveying Judges About Artificial Intelligence: Profession, Judicial Adjudication, and Legal Principles' (2025) 40 *AI & Society* 569 <https://doi.org/10.1007/s00146-024-01869-4>.
[65]See Jennifer Logg, Julia Minson and Don Moore, 'Algorithm Appreciation: People Prefer Algorithmic to Human Judgment' (2019) 151 Organizational Behavior and Human Decision Processes 90; Yochanan Bigman and Kurt Gray, 'People are Averse to Machines Making Moral Decisions' (2018) 181 *Cognition* 21.
[66]For example, Noam Gur and Jonathan Jackson, 'Procedure–content Interaction in Attitudes to Law and in the Value of the Rule of Law' in Denise Meyerson, Catriona Mackenzie and Therese MacDermott (eds), *Procedural Justice and Relational Theory* (Routledge, 2020) 111; James Gibson, Gregory Caldeira and Lester Spence, 'Measuring Attitudes toward the United States Supreme Court' (2003) 47(2) *American Journal of Political Science* 354.
[67]Avital Mentovich, J.J. Prescott and Orna Rabinovich-Einy, 'Legitimacy and Online Proceedings: Procedural Justice, Access to Justice, and the Role of Income' (2023) *Law and Society Review* 57(2) 189.
[68]Frans van Dijk, 'Legitimacy as Expressed versus Legitimacy as Experienced: Methodologies to Assess an Elusive Concept' (2023) 19(2) *Utrecht Law Review* 105.
[69]Edouard Machery, *Philosophy within its Proper Bounds* (Oxford University Press, 2017).
[70]See Kevin Tobia, 'Methodology and Innovation in Jurisprudence' (2023) 123(8) *Columbia Law Review* 2483; Fiery Cushman and Joshua Greene, 'Finding Faults: How Moral Dilemmas Illuminate Cognitive Structure' (2012) 7(3) *Social Neuroscience* 269.
[71]For example, Norman Poythress and others, 'Procedural Justice Judgments of Alternative Procedures for Resolving Medical Malpractice Claims' (1993) 23(20) *Journal of Applied Social Psychology* 1639.
[72]For example, Stephen Garcia, Patricia Chen and Matthew Gordon, 'The Letter Versus the Spirit of the Law: A Lay Perspective on Culpability' (2014) 9(5) *Judgment and Decision Making* 479.

importance of backlog reduction rather than about the legitimacy of such assistance per se. Accordingly, we focused participants on the logic of using AI legal research in the resolution of a discrete legal question and analysed their evaluations of the legitimacy of contrasting human – and AI – guided legal interpretations.

Whereas vignette studies of attitudes on philosophically salient topics often make use of convenience samples, we conducted the survey on a nationally representative sample. The risk with convenience sampling is that data may be collected from participants who are outliers in the general community, and, in consequence, that the study's conclusions offer a misleading account of the importance of the relevant factors.[73] A representative sample addresses this challenge by providing a stronger basis on which to generalize to the broader population – a national population which already forms a standard unit of social scientific analysis. We chose to survey Kenya because its views on legal AI have particular significance in light of the Kenyan judiciary's interest in the potential of e-justice measures to reduce existing case backlogs, as demonstrated by its willingness to conduct a nationwide randomized control trial on information technology's impact on case management practices.[74] Our choice also responds to criticism that most attitudinal research has been focused on W.E.I.R.D. (White Educated Industrialized Rich and Democratic) populations, which have been found to deviate systematically from global trends along several metrics.[75]

2,246 (1,198 male, 1,045 female, 987 between 18 and 29 years old, 603 between 30 and 39 years old, 349 between 40 and 49 years old, 231 between 50 and 59 years old, and 76 who were 60 years old or older) participants completed our survey in either English or Swahili. Data was collected by TGM Research from a national panel with a sampling strategy that aimed at producing a representative sample of Kenya's population by age, gender, and region (for the distribution of participants per region, see supplementary materials). The study was approved by Maynooth University Research Ethics Committee.[76]

The study compared the responses of four nationally representative cohorts (totaling 2,246[77]) to a suite of four test cases, each of which featured the same fact situation but which varied according to (a) whether the verdict aligned with either the law's text or its purpose, and (b) whether the verdict relied on the legal analysis of either a human or an artificial law clerk. To allay the risk that participant responses might be a function of some peculiar feature of one particular vignette, we investigated the overall response to a suite of four test cases, each involving an everyday situation. In designing them, we

[73]Kenneth Himma, 'Replacement Naturalism and the Limits of Experimental Jurisprudence' (2023) 14(3) *Jurisprudence* 348, 369.

[74]See Chemin (n 48).

[75]See Joseph Henrich, Steven Heine and Ara Norenzayan, 'Most people are not WEIRD' (2010) 466(29) *Nature* 29; H. Clark Barrett, 'Deciding What to Observe: Thoughts for a Post-WEIRD Generation' (2022) 41(5) *Evolution and Human* Behavior 445.

[76]SRESC-2024-38110. Our preregistered analysis plan is available here: <https://aspredicted.org/CJ9_3LX>. We pre-registered a sample of 2,000 participants. However, in order to ensure that the data accurately represented all age-groups, further data was needed. Restricting the analysis to the first 2,000 responses produces essentially the same significance patterns. Code and data for this alternative analysis are available in the online supplementary materials: <https://osf.io/4386v/?view_only=b00326e9816049908314a00a7a1172e6>.

[77]The study followed a 2 (assistance type: human vs. AI) between-subjects × 2 (case type: text consistent verdict vs. purpose consistent verdict) between-subjects × 4 (scenario: No Travel; No Bodabodas; No Sleeping; No Swimming Attire) within-subjects design. Participants received one case from each scenario (totalling 4 cases) in a random order, two of which with text consistent verdicts and two of which with purpose consistent verdicts. Notably, by independently manipulating whether the computer program indicated that text or purpose should be prioritised, our design disentangled the question of the source of assistance from that of the legal outcome.

incorporated cultural norms, and, in the survey's English language version, local idiom. For example, our materials referred to a 'bodaboda', a motorcycle taxi that is common in Kenya. Likewise, they referred to a 'legal researcher' rather than a 'law clerk', as the latter would denote an official whose functions were purely administrative.

To test our prediction that Kenyans would deem the decisions of AI-advised judges to be no less legitimate than those of human-advised judges, we proposed fitting a mixed-effects model of legitimacy judgments with fixed effects for case type, assistance type, and the case type*assistance type interaction, while allowing random intercepts for scenario and participant. Specifically, we predicted that an ANOVA based on this mixed-effects model would reveal significant main effects of case type ($p < .05$), but no effects of assistance type or of the case type*assistance type interaction (ps > .05).

For instance, the 'No Bodabodas in the mall' vignette was presented as follows:

> The government has issued a rule: "It shall be an offence to ride a bodaboda in a shopping mall".
>
> This rule is intended to prevent injuries to shoppers.

Then, we described a situation in which an agent had acted contrary to the law's text but consistently with its purpose:

> Witnessing a violent attack inside a mall, Martin rides his bodaboda into the mall to stop it.
>
> Martin is later charged with the offence of riding a bodaboda in a shopping mall.

Finally, we described a legal proceeding that varied both according to its outcome and according to the source of the legal research on which the court relied:

> The court, guided by legal research performed by a legal researcher/special computer program, decides that Martin violated/did not violate the rule.[78]

We chose the phrase 'special computer programme' because a general Kenyan audience would be more familiar with it than with the more technical term, 'artificial intelligence'.[79] Of course, both computer programmes and AIs can help judges with the sorts of tasks facing any office worker, from email services to database searches. To isolate the assistance of interest, we therefore specified the relevant task as 'performing legal research'. This was designed to exclude computer facilities such as those traditionally offered by Westlaw or LexisNexis that help find and sort legal materials. Any research conducted with the assistance of such tools is naturally understood to be conducted by the person using them rather than by the facilities themselves: we would not consider a student's reliance on a library catalogue as presenting an issue of academic credit. On the other

[78]Complete stimuli available in the online supplementary materials: https://osf.io/4386v/?view_only=b00326e9816049908314a00a7a1172e6.

[79]Comments on earlier iterations of this draft suggested that 'special computer program' could be interpreted in ways that are significantly different from 'artificial intelligence program', such that we should not use the former to assess the legitimacy of the latter. To check whether this criticism was valid, we ran a pre-registered experiment (https://aspredicted.org/9w42-cc9n.pdf) with 200 participants (mean of age = 38.53, 90 male, 108 female, 2 non-binary) on Prolific.co following a 2 (condition: special computer program, AI) between subjects × 4 (scenario: same as main study) within subjects design. The results revealed no significant difference (BF01 = 12.02) in legitimacy ratings between participants assigned to the special computer program or to the AI program. Full data and analysis are available in the online supplementary materials. Given these results, we are confident that the study reported below warrants inferences about the perception of AI legitimacy.

hand, any student who relied on a computer programme to draft their research essay would be expected (at a minimum) to acknowledge their use of such assistance.

Participants were then asked to indicate their agreement with the sentence, 'The court's decision is legitimate', on a 5-point Likert scale. To ensure internal validity by preventing participants from guessing our hypothesis and responding accordingly, participants only received cases that featured *either* a human *or* an AI law clerk. Our choice of dependent variable reflected our interest in identifying whether participants considered legal reasoning an exclusively human sphere of activity or one that might accommodate judicial reliance on AI. Unlike the experimental literature on the nature of legal meaning, therefore, we did not ask participants to apply the rule themselves (or whether the court's decision was correct) (e.g. Struchiner et al 2020).

Part 4. Results

To analyse the data, we fitted the preregistered mixed-effects model of legitimacy judgments with fixed effects for case type, assistance type, and the case type*assistance type interaction, while allowing random intercepts for scenario and participant. As predicted, an ANOVA based on this model revealed significant main effects of case type ($F_{(1,\ 2233)} = 24.40$, $p < .001$), but no effects of assistance type ($F_{(1,\ 2233)} = 0.34$, $p = .563$; $BF_{01} = 24.82$) or of the interaction between case type and assistance type ($F_{(1,\ 2233)} = 2.78$, $p = .096$; $BF_{01} = 5.31$). Participants tended to view the court's ruling as more legitimate when it interpreted the rule in accordance with its purpose ($M = 3.59$ [3.47, 3.72]) than when it interpreted the rule in accordance with its text ($M = 3.40$ [3.28, 3.53]). Crucially, however, participants rated court decisions which relied on AI assistance ($M = 3.49$ [3.36, 3.61]) to be just as legitimate as those which relied on human assistance ($M = 3.51$ [3.39, 3.63]). Similarly, on average, participants saw AI clerks as more legitimate than not (see Figure 1).[80]

An exploratory model including fixed effects for scenario revealed a significant two-way interaction between scenario and case type ($F_{(3,\ 6735)} = 197.46$, $p < .001$). Inspecting the marginal means for each case type and each scenario, we observed that participants significantly preferred decisions which interpreted the rule according to its purpose in the 'No sleeping in the train station' ($b = 0.81$, $t = 14.02$, $p_{Tukey} < .001$) and the 'No entry to government buildings in swimming attire' ($b = 0.78$, $t = 13.48$, $p_{Tukey} < .001$) scenarios, but that that preference was reversed for the 'No bodabodas in the shopping mall' ($b = -0.29$, $t = -4.99$, $p_{Tukey} < .001$) and 'No driving without a license' ($b = -0.55$, $t = -9.51$, $p_{Tukey} < .001$) scenarios. The same model also revealed a small, but significant interaction between assistance-type and scenario ($F_{(3,\ 6735)} = 3.10$, $p = .025$). The effects of assistance-type were non-significant for all ($|b|s < .08$, $|t|s < 1.35$, $ps_{Tukey} > .17$) but the 'No sleeping in the train station' scenario ($b = -0.12$, $t = -2.07$, $p = .039$).[81] These trends are represented in Figure 2.

[80]The latter result was replicated in a pre-registered analysis performed in the convenience sample study described in footnote 79. (Given that norms concerning the appropriate role of *human* law clerks vary between different jurisdictions, it is notable that our participants also, on average, considered a legal finding guided by legal research performed by a human assistant to be legitimate.)

[81]This result is driven by the larger disparity in the rating of the 'No sleeping' scenario's text-based and purpose-based interpretations. Notice that three of the four scenarios had a marginally higher rating for humans for textual applications (the fourth features an equal rating). Perhaps greater doubts about convicting the dozing commuter based on an AI suggestion might derive from stereotypes of computer decision-making as overly rigid.

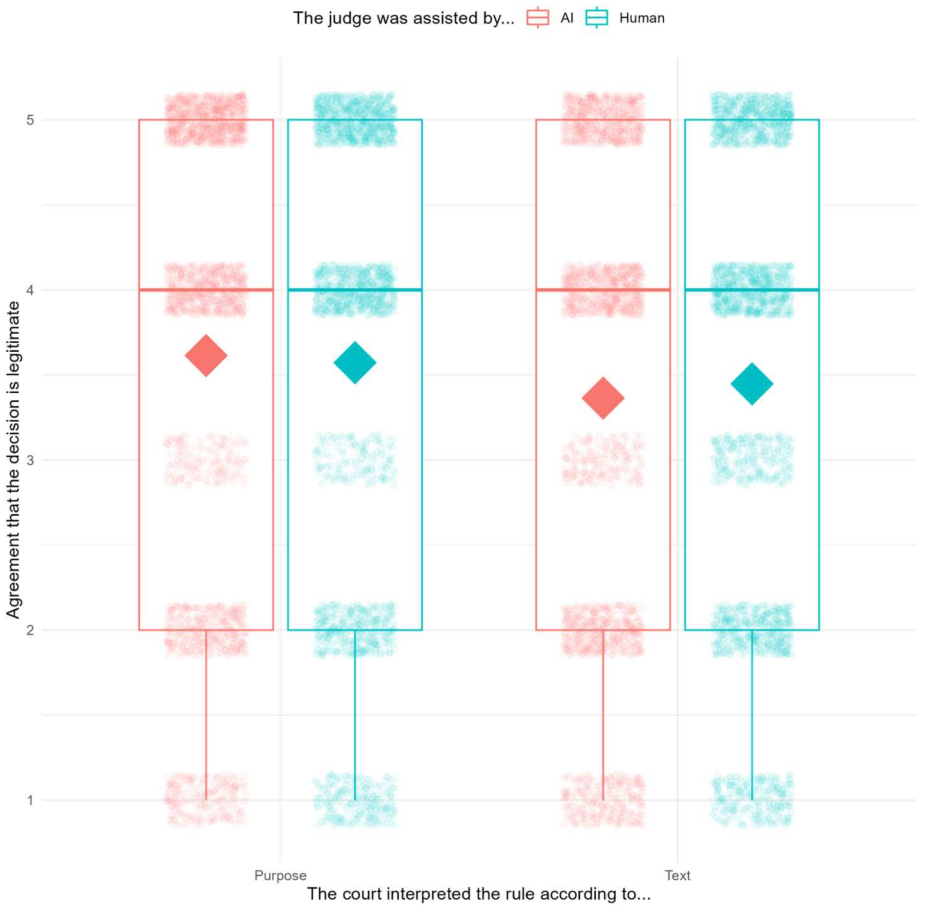

Figure 1. Agreement with the statement that the court's decision is legitimate by case-type and assistance-type, collapsed across scenarios. Diamonds represent the mean.

Part 5. General discussion

We tested perceptions of the legitimacy of the administration of justice through judicial reliance on AI-generated legal analysis. Confirming our preregistered hypothesis, the study revealed no overall difference in the perceived legitimacy of AI – and human-assisted legal interpretations. Participants considered legal decisions that relied on AI-generated legal research to be just as legitimate as decisions that relied on human-authored research. On the other hand, whether the court's decision prioritized the law's text or its purpose – a factor known to impact rule application – did significantly influence the decision's perceived legitimacy.

Justice systems have been shown to facilitate economic development by promoting competitive credit markets and firm productivity,[82] and by spurring investment in the

[82]For example, Reshad Ahsan, 'Input Tariffs, Speed of Contract Enforcement, and the Productivity of Firms in India' (2013) 90(1) *Journal of International Economics* 181; <https://ideas.repec.org/a/eee/inecon/v90y2013i1p181-192.html>; Sandra Sequeira, 'Corruption, Trade Costs, and Gains from Tariff Liberalization: Evidence from Southern Africa' (2016) 106(10) *American Economic Review* 3029.

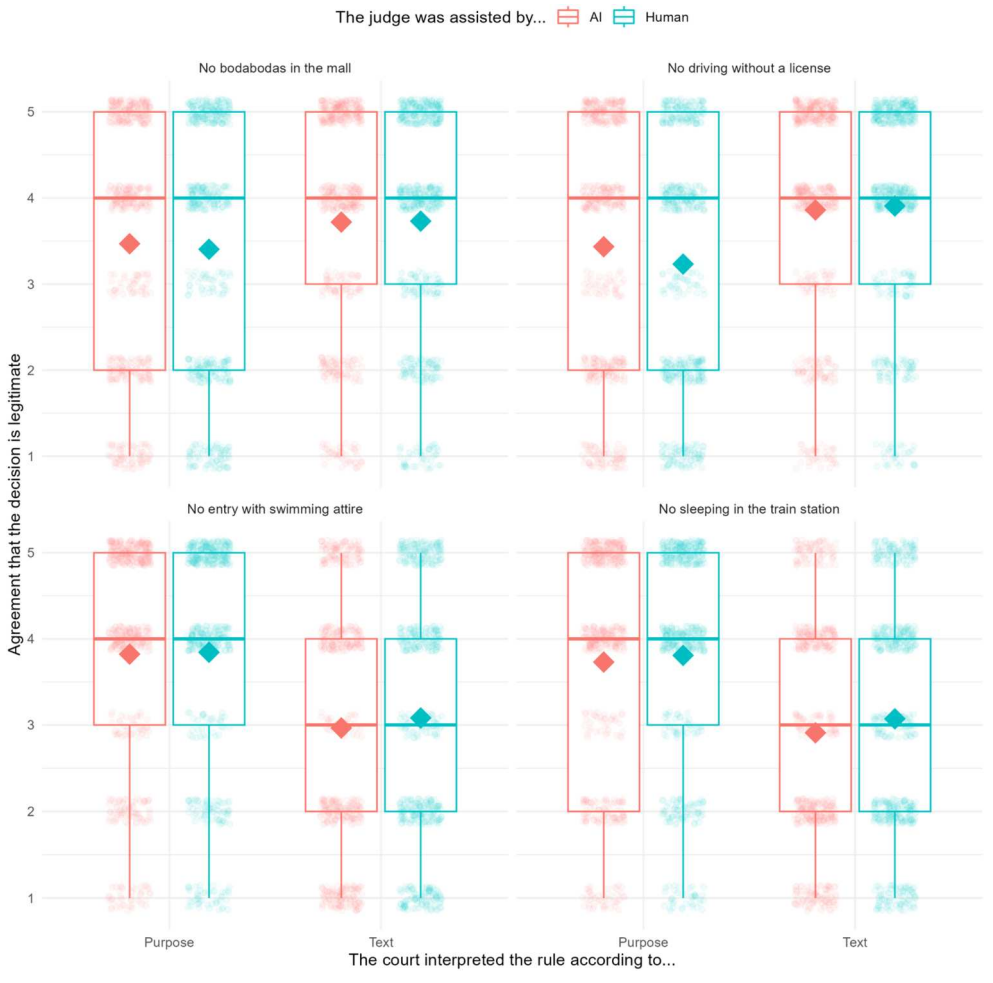

Figure 2. Agreement with the statement that the court's decision is legitimate by case-type and assistance-type, faceted by scenarios. Diamonds represent the mean.

business environment.[83] But an estimated 1.5 billion individuals globally struggle to access the law in respect of administrative, criminal, or civil challenges.[84] By analogy with the positive contribution of human law clerks,[85] the use of LLMs to assist judicial legal research promises to significantly enhance the efficiency of formal adjudication.[86] Notably, judges who have pioneered the use of LLMs in their own research have

[83]Guilherme Lichand and Rodrigo Soares, 'Access to Justice and Entrepreneurship: Evidence from Brazil's Special Civil Tribunals' (2014) 57(2) *Journal of Law and Economics* 459; Matthieu Chemin, 'Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India' (2012) 28(3) *Journal of Law, Economics, & Organization* 460.

[84]World Justice Project, 'Measuring the Justice Gap: A People Entered Assessment of Unmet Justice Needs around the World' Technical report (Washington, D.C. 2019) URL: <https://worldjusticeproject.org/sites/default/files/documents/WJPMeasuring%20the%20Justice%>.

[85]Peverall (n 38).

[86]Miriam Stankovich and others, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (2023) *UNESCO* 58 <https://unesdoc.unesco.org/ark:/48223/pf0000387331>

emphasized this potential dividend: 'it provides judges … with an inexpensive research tool'.[87] Justice delayed is justice denied: whether LLMs can increase access to legal adjudication without compromising rule of law is a key question of our age.

As it is, we know that '[j]udges everywhere face crowded dockets and enormous time pressures'.[88] Equally, we know that, in response to this pressure, some judges may choose to expedite legal research and judgment writing at a cost to our normative ideals. For instance, busy judges' chambers have been shown to rely on the legal analysis of user-generated internet content, namely, that of anonymous Wikipedia editors.[89] In principle, it would seem preferable for adjudicators to make use of the legal analysis of an entity that has passed the bar than of that of an unknown internet user who might have no legal training whatsoever. Judicial use of an LLM might then offer a less problematic alternative; more positively, such a practice might no more compromise our ideal of the rule of law than the employment of a human law clerk. On this question – on whether, all other things being equal, reliance on a machine's analysis is as acceptable as reliance on that of a qualified human – the results of our study suggest that, intuitively, the answer may be affirmative.

In theory, the capacity to draft legal judgments explaining the law's application to particular disputes can serve the functions of judicial support and decision alike. Accordingly, any discussion of the legitimacy of LLMs as law clerks can be analysed for its implications for the prospect of LLMs as judges, and vice versa. In the case of judges, one prominent objection is that the process by which LLMs operate is simply too opaque to litigants to count as 'an instance of adjudication'.[90] This objection's logic might be extended to the human judge's reliance on an artificial law clerk, to whom the latter's operation will be equally opaque. Our study did not test the issue of opacity directly, and it would be valuable to know more about its salience in public perceptions.

In designing such further research, it may be helpful to consider the 'companions in guilt' argument that points to the opacity of the human mind itself. Defending Wisconsin's use of opaque criminal sentencing algorithms, for instance, an official noted that: 'We don't know what's going on in a judge's head; it's a black box, too'.[91] Indeed, the absence of a formula for systematically mapping some entity's ostensible beliefs onto their causal determinants has been taken by some theorists to qualify that entity as an agent in its own right, whether the entity in question is an individual human or a

[87]Newsom (n 44). See also Juan Manuel Padilla García in Alejandro León, 'Sentencia la tomé yo, ChatGPT respaldó argumentación: juez de Cartagena usó inteligencia artificial' *Blu Radio* (Columbia, 2 February 2023) <https://www.bluradio.com/judicial/sentencia-la-tome-yo-chatgpt-respaldo-argumentacion-juez-de-cartagena-uso-inteligencia-artificial-pr30>.

[88]Jeffrey Rachlinski and Andrew Wistrich, 'Judging the Judiciary by the Numbers: Empirical Research on Judges' (2017) 13 *Annual Review of Law and Society* 203, 223. See also Holger Spamann and Lars Klohn, 'Justice is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges' (2016) 45(2) *Journal of Legal Studies* 255, 274.

[89]Neil Thompson and others, 'User-Generated Content Shapes Judicial Reasoning: Evidence from a Randomized Control Trial on Wikipedia' (2024) 35(4) *Information Systems Research* <https://doi.org/10.1287/isre.2023.0034>.

[90]William Lucy, 'Algorithms and adjudication' (2024) 15(3) *Jurisprudence* 251, 269. See also Kalliopi Terzidou, 'The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial' (2022) 31 *Journal of Judicial Administration* 154, 162; Eden Sarid and Omri Ben-Zvi, 'Machine Learning and the Re-Enchantment of the Administrative State' (2024) 87(2) *Modern Law Review* 371.

[91]Christine Remington in Jason Tashea 'Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions' *American Bar Association Journal* (Chicago, 1 March 2017) <http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole>.

group of humans,[92] including a collective legislature.[93] If, in producing a statute or a judgment, a legislature or a multi-member court can be said to act collectively in virtue of the opacity of the connection to members' respective individual inputs, then it would seem to follow that the opacity of an LLM's inner workings would similarly serve to qualify its production of a text as an act, whether of adjudication or simply of advice. Certainly, the connection between arguments from opacity that would enlarge the scope for collective moral responsibility and those that would limit artificial judicial capacity merits further consideration.

Although representative, our data speaks to the attitudes of just a single country. Cross-national investigation will be necessary to allow robust inference as to the existence of any univocal folk understanding, or alternatively, to identify the dimensions of cross-cultural variation. Likewise, a vignette experiment such as ours suppresses the complexities of ordinary decision-making. Because our study was conducted in (virtual) laboratory conditions, rather than conducted in-situ in a real-world setting, our results may lack some degree of 'ecological' validity. This represents a tradeoff between the inclusion of controls that allow a means of direct causal inference and the creation of an artificial environment which may not reproduce the ordinary circumstances under which people might make judgments in practice. Notwithstanding both limitations, however, the study represents a 'first step'[94] in ascertaining the intuitive acceptability of judicial reliance on artificially generated legal reasoning. In doing so, moreover, it contributes to a nascent movement that seeks to apply contemporary social scientific methods to longstanding topics in legal theory, 'experimental' jurisprudence.

Although HLA Hart famously characterized his landmark work of analytic legal philosophy, 'The Concept of Law', as exercise in 'descriptive sociology',[95] for many years, systematic evidence of lay legal intuitions remained scant. Traditionally, legal scholars have offered theories of the nature of law that have taken the form of an analysis of the relevant legal concept. Defending such theories, scholars have routinely invoked their consistency with *folk* legal concepts, i.e. with our linguistic intuitions.[96] This practice is equally evident in the identification of particular values as characteristic of the rule of law.[97] A notable advantage of an intuitive theory of the rule of law is that it would remove any need for a supplementary error theory, by which to account for people's misconception of a familiar institution.[98] A challenge for jurisprudence, however, has been to establish the content of folk concepts from the philosopher's armchair without falling victim to problems of 'groupthink and information cascades'.[99] Seeking to answer this challenge, a new socio-legal research agenda now looks to uncover relevant folk legal concepts

[92]Christian List and Philip Pettit, *Group Agency: The Possibility, Design, and Status of Corporate Agents* (Oxford University Press, 2011).
[93]Richard Ekins, *The Nature of Legislative Intent* (Oxford University Press, 2012).
[94]van Dijk (n 68) 106.
[95]H.L.A. Hart, *The Concept of Law* (Clarendon Press, 1961) vi.
[96]David Plunkett and Daniel Wodak, 'Legal Positivism and the Real Definition of Law' (2022) 13(3) *Jurisprudence* 317.
[97]For example, Lon Fuller, The Morality of Law (Yale University Press, 1964); Kristen Rundle, *Forms Liberate: Reclaiming the Jurisprudence of Lon L Fuller* (Bloomsbury, 2012); Jonathan Crowe, 'Between Morality and Efficacy: Reclaiming the Natural Law Theory of Lon Fuller' (2014) 5(1) *Jurisprudence* 109.
[98]Emad Atiq, 'Legal Positivism and the Moral Origins of Legal Systems' (2023) 36(1) *Canadian Journal of Law & Jurisprudence* 37. See further Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford University Press, 1998).
[99]Kevin Tobia, 'Experimental Jurisprudence' (2022) 89(3) *University of Chicago Law Review* 735.

through more systematic methods.[100] As evidence of the popular acceptability of AI law clerks, the data we have collected can also be viewed as an empirical contribution to the traditional jurisprudential project of elucidating common intuitions about the rule of law's adjudicative dimension. Specifically, our findings would suggest that any theory of law's intrinsic nature that excluded non-human judicial insight may have to explain away people's possession of a broader, more accommodating conception of legality.

Conclusion

Formal dispute resolution is an achievement of human civilization that allows justice to be dispensed in a deliberate, predictable fashion and that facilitates the development of complex modes of organization that contribute to individual prosperity and social equality alike. The recent step-change in artificial intelligence offers a potential resource with which to expand such systems' reach. Optimistically, the creation of the LLM might be to the administration of justice what the invention of the carbon microphone was to speech and music. Of course, key to any assessment will be the question of what LLMs might actually help administer: justice, or mere state coercion. In reaching an answer, the views of those who stand to gain the most through more readily available dispute resolution will be critical. Reporting a nationally representative survey experiment, we found that Kenyans consider judicial reliance on the legal opinion of an AI programme to be just as legitimate as reliance on that of a human legal professional. Clearly, much work remains to be done on the question of legal AI's consistency with our ideals of legality. But our findings give impetus, also, to systematic investigation of whether the integration of legal LLMs might make justice systems more efficient, accessible, and trustworthy in practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

We would like to acknowledge the financial support of Maynooth University's Faculty of Social Science and Social Sciences Institute.

[100]Karolina Prochownik, 'The Experimental Philosophy of Law: New Ways, Old Questions, and how not to get Lost' (2021) 16(12) *Philosophy Compass e12791*.