The Rule of Law or the Rule of Robots?
Nationally Representative Survey Evidence from Kenya

Abstract

With AI now passing the bar, and with court caseloads worldwide hampering access to justice, there are calls for judges to make use of chatbots to help expedite their work of legal interpretation. Such calls pose an empirical question of the practical difference such a reform might be expected to make. They also pose a normative question: whether judicial reliance on computer generated legal research in deciding litigants' rights and obligations is consistent with the rule of law. In this paper, we address the normative question by exploring whether the folk concept of law demands that the law's application be guided by human legal insight only. We report a vignette-based experiment on the acceptability of AI law clerks - assistants whose contribution does not decide what the law says but which may inform the ultimate decision. Collecting nationally representative survey data from Kenya, we find that an AI clerk's influence on legal outcomes is seen as no less legitimate than that of a human clerk. This result spurs efforts to systematically investigate whether the integration of AI might make justice systems more efficient, accessible, and trustworthy in practice.

The Rule of Law or the Rule of Robots?
Nationally Representative Survey Evidence from Kenya

> "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.

> "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.

Until recently, most would have agreed with Dworkin's insistence that human judgment is a prerequisite of competent legal 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 programs, 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. Soon, we might have to recognize Lawlor's prescience and acknowledge that the electronic magician has produced her masterpiece.

Theorists have been elucidating the normative stakes of employing machines to resolve legal disputes since the 1960s (e.g., Dickerson 1963; Buchanan & Headrick 1970). Now, with the explosion in machine learning capability, the question is receiving renewed attention. Some foresee a 'legal singularity' in which the use of algorithms enables society to establish exponentially more efficient and cheaper litigation services, including computer judges (e.g., Aidid and Alarie 2023; Volokh 2019; Casey and Niblett 2019). This possible future could be especially relevant for countries with overburdened judiciaries, for whom AI might offer equitable and timely access to justice. Others identify a tension between our commitment to the rule of law and the relinquishment of human responsibility for the individual's legal fate

(Endicott and Yeung 2022). 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' (Hildebrandt 2016, 2).

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 such measures be considered normatively acceptable? Addressing the latter question, we report a vignette-based study on the legitimacy of robot law clerks. Collecting nationally representative survey data from Kenya, we find that judicial reliance on computer generated legal research is seen as no less legitimate than judicial 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 AI's potential to provide a law clerk for busy judges (Part 2). Part 3 considers how the legitimacy of robot law clerks can be elucidated by empirical inquiry into the folk concept of legality. Then, on presenting our study's method (Part 4), we report its results (Part 5), and discuss their significance and implications for future research (Part 6).

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' (Hart 1961, 270).

These systems faced 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' (MacCormick 1995). Defeasibility refers to the limitation that, no matter how elaborate, no text can specify all intuitively exonerating circumstances: "it is not possible to conclude legal clarity from semantic clarity" (Klatt 2008, 219; similarly, Pasquale 2019, 48). 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 rendered by man or machine: "deduction cannot provide an adequate model of legal reasoning" (Bench-Capon et al 2009, 17). So, while schematized-human-learning systems have been made available commercially to consumers in certain legal domains, notably, in that of tax law (Contos et al. 2011), 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 emphasised. 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 was 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' (Marmor 2005, 115). It proved that initial doubts, e.g., that 'deep structure is to be found in social context and purpose, which are non-computational' (Greinke 1994), were premature. But while the problem of defeasibility is mitigated, it is not yet solved.

A case might include a novel, intuitively exonerating characteristic to which the algorithm is blind for want of any opportunity to learn of its legal salience (Surden 2014, 105-06; Chirimuuta 2023, 14). 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 way (Flanagan 2010, 262-63). Again, with no way of knowing in advance whether a particular case will feature such a characteristic, it is unclear how, in

applying the law, a judge might responsibly rely on the application suggested by an SML program.

The overarching challenge of defeasibility is that both expert and SML AI systems are liable to overlook a situation's legal novelty in ways a human never would: they 'suck[] up the dust and the crickets' alike (Simmons 2018, 1095). 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, reliable legal AI is clearly in prospect.

A large language model (LLM) also relies on machine learning. LLMs learn to accurately predict the next token (which can sometimes 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 specialised 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 stepchange 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 (Jones and Bergen 2024). Similarly, in the legal domain, LLMs have achieved key markers of professional human competence, such as passing the US Uniform Bar Exam (Katz et al 2024; Martinez 2024). Perhaps the most remarkable feature of LLM legal aptitude, however, is its ability to favour the law's spirit over 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 (Almeida et al 2024). The LLMs reproduced humans' propensity to a) sometimes prioritise a law's spirit over its letter, and b) do so especially where the law's purpose was benevolent. Like the human judge, LLMs could recognize a case's inclusion of a novel, intuitively exonerating characteristic on whose legal salience it had not been trained. Other work has shown that LLMs perform very well across a wide array of different legal tasks (Guha et al., 2023).

LLMs still have significant limitations. Chief amongst them is LLMs' well-documented tendency to hallucinate (Maynez et al., 2020; Zhang et al., 2023), i.e., to confidently include false facts in their responses. In the legal domain, a US-based law firm was fined $5,000 after

a lawyer submitted a legal brief which cited non-existent cases to a federal court; the cases were chatGPT hallucinations (Merken 2023). Nevertheless, with the LLM, it seems that legal AI may have overcome the challenge of defeasibility and thereby transcended the traditional barrier to reliability: "Computers 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" (Susskind 1986, 133).

With AI's interpretive response to novel but potentially exonerating features tracking 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' (Livermore 2020, 239). So it is that, in approaching functional parity with human lawyers, LLMs have placed the role of robot generated legal interpretation firmly on the agenda. Global law firms have been quick to embrace the opportunities (Criddle 2023).

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 judiciary's response to growing caseloads (e.g., Posner 1985; Cohen 1995; Lipez 2007). 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' (Peverall 2020, 280). Perhaps AI could help.

Rather than acting as judge, AI might provide important adjudicative supports either by extracting relevant precedents, as happens already in China (Deng 2019), or by recommending an outcome for stated legal reasons (as envisaged in, e.g., Lim 2023; Choi and Schwarcz forthcoming). Significantly, a study has found that the use of ChatGPT reduces the time spent by humans on drafting advisory legal memos (Choi et al forthcoming) - just as the employment of a human clerk might be expected to do. Judges have noticed:

> '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, 2023.

> ‘[Judges] should consider whether and how AI-powered large language models… might… inform the interpretive analysis.’ Judge Kevin Newsom, US Court of Appeal, 2024.

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 (Susskind 2019, 287). Kenya is one such country.

Facing a 500,000 case backlog, Kenyan judicial authorities have expressed openness to technological innovation (Gitonga 2024). Notably, Kenya has trialled an AI-based case management system with which to better distribute judicial capacity (Chemin et al 2023). 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 (e.g., Rodrik 2000; Acemoglu et al. 2001), the case for systematically investigating this possibility is pressing. The starting point must be consideration of the artificial law clerk’s normative acceptability.

Part 3 The Folk Concept of the Rule of Law

There is wide agreement that rule systems can be evaluated by reference to how well they serve the value of legality, i.e., by how faithfully they adhere to the rule of law. Theories of the rule of law articulate procedural or formal principles, through which law is made and administered. The most prominent effort is that of Lon Fuller who set out seven conditions, 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 (1964, 96–97). Notably, none of the aforementioned principles proscribes AI. Potentially, AI could assist in 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. Might people also endorse an additional principle requiring that the law be applied with reliance only on human legal insight?

In their identification of particular values as characteristic of the rule of law, theorists have typically appealed to our folk concept (e.g., Fuller 1964, 62; Rundle 2012, 80; Crowe 2014, 113). For a long time, systematic evidence of lay legal intuitions was scant. Now, a new research agenda, 'experimental jurisprudence', seeks to uncover such evidence through methods of psychology scholarship (Prochownik 2021). Replicating an initial study by Donelson and Hannikainen (2020), Ivar Hannkainen and colleagues found that, across a diverse set of cultures and linguistic communities, majorities of survey participants commonly recognized Fuller's desiderata as principles to which legal systems ought to adhere (Hannikainen et al 2021; see also, Gur and Jackson 2020). To shed light on whether an artificial law clerk would be consistent with the rule of law or, instead, would bespeak a rule of robots, we follow experimental jurisprudents in investigating the folk concept of legality.

Consider the normative question on which the deployment of artificial law clerks would seem ultimately to depend: is it 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, robot clerks would be unproblematic: "AI judicial staff attorneys that draft proposed opinions for judges to review—would be… legitimate" Volokh 2019, 1141. There are sceptics, however, for whom interpretive reliance on machines would involve a 'legitimacy deficit' that would pose 'a real threat to social wellbeing' (Zuckerberg 2020, 304).

It is true that, on issues of resource allocation, such as hiring or university admission (e.g., Newman et al 2020; Claudy et al 2022), 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 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 (Chen et al 2022). 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. Conversely, were AI to merely assume the role of law clerk, ultimate interpretive responsibility would remain in human hands. Moreover, when it comes to *advice* rather than *decision*, there is evidence that people prefer to rely on that provided by an algorithm over that provided by a human (Logg et al 2019). Accordingly, we hypothesized that judges would be perceived as being no less legitimately guided by computer generated legal research as by that produced by a human.

Part 4 Method

We explore the legitimacy of chatbot law clerks by conducting a nationally representative survey experiment of Kenya, a society whose views on such matters have particular salience in light of the Kenyan judiciary's willingness to test the effects of e-justice measures. Our choice of population also responds to criticism that experimental jurisprudence has so far been focused on W.E.I.R.D. (White Educated Industrialized Rich and Democratic) populations (Tobia 2024), which have been found to deviate systematically from global trends along several metrics (Henrich et al 2010; Barrett 2020).

2,246 (1,198 male, 1,045 female) participants completed our survey in either English or Swahili. The sample was recruited by TGM Research and is representative of Kenya's population by age, gender, and region. The study was approved by the Research Ethics Committee at a major university.

Experimental philosophers have long warned against the dangers of asking lay participants to synthesize their own conceptual knowledge (Machery, 2017); simply asking people whether the rule of law demands exclusive reliance on human legal opinion would be a flawed strategy. To gain insight into the structure of ordinary normative understandings, we therefore use a vignette-based experimental design to analyse participant evaluations of the legitimacy of concrete situations involving contrasting human- and AI- guided legal interpretations (see Cushman & Greene 2012).

The study compared the responses of four nationally representative cohorts (totalling 2,246[1]) 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.

[1] The study followed a 2 (assistance type: human vs. AI) between-subjects x 2 (case type: text consistent verdict vs. purpose consistent verdict) between-subjects x 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.

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 propose to fit 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 ($ps > .05$). (Prediction and analysis preregistered.)[2]

For instance, the "No Bodabodas[3] 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:

[2] 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, which led to this deviation from the original pre-registered plan. Restricting the analysis to the first 2,000 responses renders 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

[3] Bodabodas are bicycle or motorcycle taxis that are common in Kenya.

> The court, guided by legal research performed by a legal researcher/special computer program, decides that Martin violated/did not violate the rule.[4]

Participants were asked to indicate their agreement with the sentence, "The court's decision is legitimate", on a 5-point Likert scale.

Part 5 Results

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

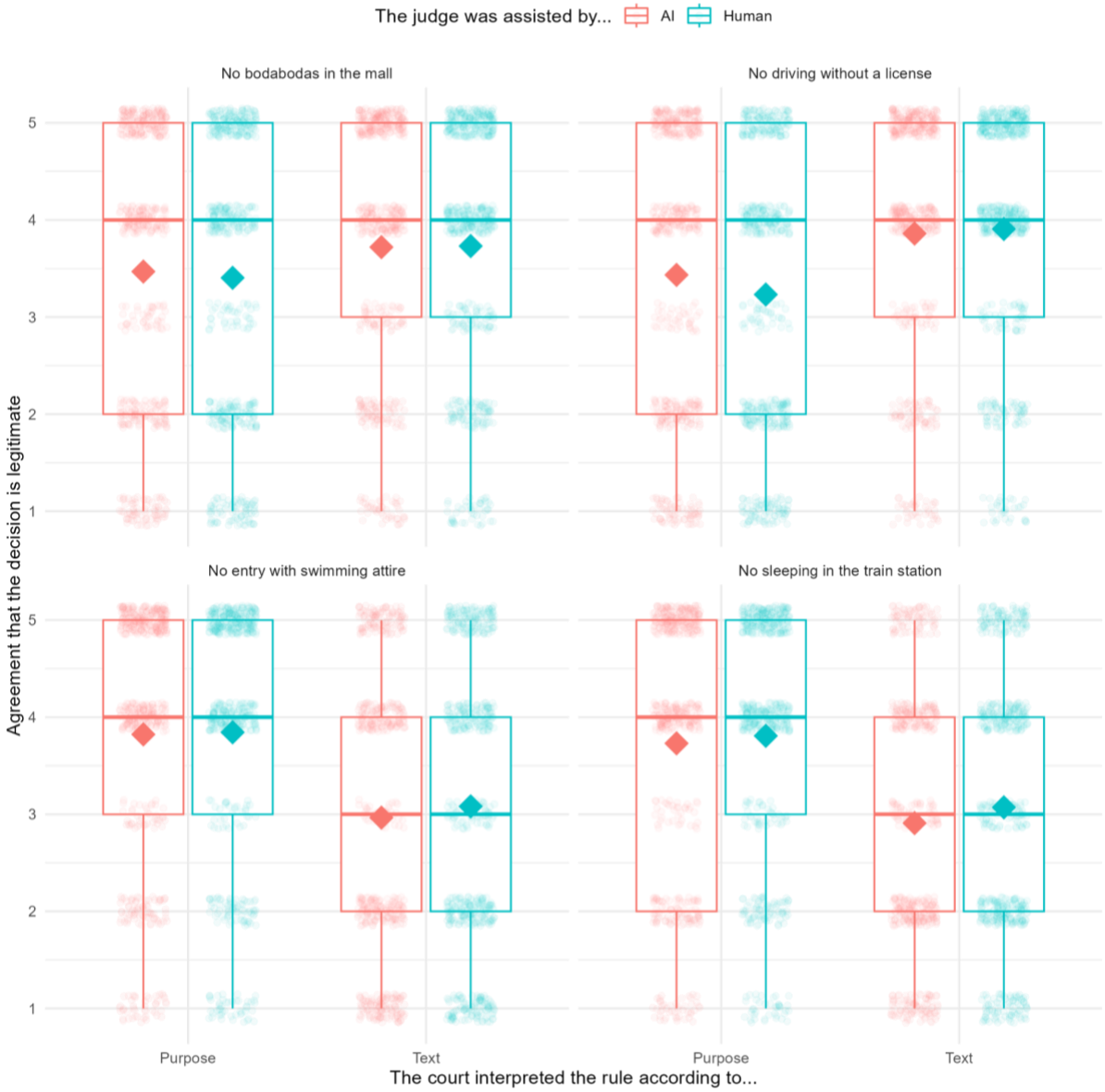

Figure 1: Agreement with the statement that the court's decision is legitimate by case-type and assistance-type, faceted by scenarios. Diamonds represent the mean.

To analyze 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 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 with its text (M = 3.40 [3.28, 3.53]). However, participants rated

courts 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]).

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 the 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 preference has 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 ($b = -0.12$, $t = -2.07$, $p = .039$). These trends are represented in Figure 1.

Part 6 General Discussion

In this paper, we tested whether the administration of justice through judicial reliance on AI-generated legal analysis coheres with our folk concept of law. Confirming our hypothesis, the study revealed no overall difference in the perceived legitimacy of AI- and human-assisted legal interpretations. With the exception of a small bias against AI law clerks in one specific scenario ("No sleeping in the station"), 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 prioritised the law's text or its purpose - a factor known to impact rule application - did significantly influence its perceived legitimacy.

Legal dispute resolution systems have been shown to facilitate economic development by promoting competitive credit markets and firm productivity (e.g., Ahsan, 2013; Sequeira, 2016), and by spurring investment in the business environment (e.g., Lichand and Soares 2014; Chemin 2012). But an estimated 1.5 billion individuals globally struggle to secure justice in respect of administrative, criminal, or civil challenges (World Justice Project, 2019). By analogy with the positive contribution of human law clerks (Peverall 2020), the use of LLMs

to assist judicial legal research promises to significantly enhance the efficiency of formal adjudication. Notably, judges who have pioneered the use of LLMs in their own research have emphasized this potential dividend: 'it provides judges… with an inexpensive research tool' (Newsom 2024, 15; similarly, Judge Juan Padilla 2023). Justice delayed is justice denied: whether LLMs can increase access to adjudication systems 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" (Rachlinski and Andrew Wistrich 2017, 223; similarly, Spamann and Klohn 2016, 274). 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. Notably, busy judges' chambers are liable to silently rely on the legal analysis of user-generated internet content of unknown origin, namely, that of anonymous Wikipedia editors (Thompson et al 2024). 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 then offers a better alternative; more positively, its judicial use might no more compromise the rule of law than the employment of a human clerk. On this question - on whether reliance on a machine's analysis is as acceptable as reliance on that of a qualified human - the results of our survey suggest that, intuitively, the answer may be affirmative.

In principle, the capacity to draft legal judgments explaining the law's application to particular disputes can serve the functions of advice 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, a prominent objection is that the process by which LLMs operate is simply too opaque to litigants to count as 'an instance of adjudication' (Lucy 2023, 19; similarly, Sarid and Ben-Zvi 2024). The objection's logic might extend to the human judge's reliance on an artificial law clerk, for 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' (Remington 2017). Indeed, the absence of any formula for systematically mapping some entity's ostensible beliefs onto their causal determinants has been taken to qualify that entity as an agent in its own right, whether the entity in question is an individual human or a group of humans (List and Pettit 2011), including a collective legislature (Ekins 2012). If, 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' 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. Opacity is not the only objection to computer judges, however.

Unlike a human, a robot has nothing at all at stake in adjudicating someone's legal rights - neither retention of their job, advancement of their promotion prospects nor 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…' (Endicott and Yeung 2022, 398; similarly, Wendel 2019, 42). Perhaps for this reason, even advocates concede that the legitimacy of AI judges 'is likely to be… counterintuitive' (Volokh 2019, 1142). One goal for future research would be to test this prediction.

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 find that Kenyans consider judicial reliance on the legal opinion of an AI program 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

how the integration of legal LLMs might make justice systems more efficient, accessible, and trustworthy in practice.

References

Acemoglu D, Johnson S and Robinson J A., 2001 'The colonial origins of comparative development: An empirical investigation' *American Economic Review* 91(5) 1369 https://www.aeaweb.org/articles?id=10.1257/aer.91.5.1369.

Ahsan R N., 2013. 'Input Tariffs, Speed of Contract Enforcement, and the Productivity of Firms in India' 90(1) *Journal of International Economics* 181 https://ideas.repec.org/a/eee/inecon/v90y2013i1p181-192.html.

Aidid A and Alarie B, 2023. The Legal Singularity: How Artificial Intelligence Can Make Law Radically Better (University of Toronto Press).

Almeida, G., Nunes, JL,. Engelmann, N,. Wiegmann, A., de Araújo, M., 2024. Exploring the Psychology of LLMs' Moral and Legal Reasoning. *Artificial Intelligence* 333 https://doi.org/10.1016/j.artint.2024.104145

Barrett, HC., 2020 Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution and Human Behavior* 41(5) 445-453.

Bench-Capon, T., Prakken, H., Sartor, G. (2009). Argumentation in Legal Reasoning. In: Simari, G., Rahwan, I. (eds) Argumentation in Artificial Intelligence. Springer, Boston, MA.

Birss C. 2023. https://www.lawgazette.co.uk/news/solicitor-condemns-judges-for-staying-silent-on-woeful-reforms/5117228.article

Buchanan B and Headrick T., 1970. Some Speculation about Artificial Intelligence and Legal Reasoning, Stanford Law Review 23(1) 40-62

Casey, A. and Niblett, A., 2019 "A Framework for the New Personalization of Law," 86 University of Chicago Law Review 333-58.

Chemin M, (2012) 'Does court speed shape economic activity? Evidence from a court reform in India' 28(3) Journal of Law, Economics, & Organization 460.

Chemin, M., Chen, D., Di Maro, V., Kimalu, P., Mokaya, M., and Ramos-Maqueda, M., 2022. Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya, TSE Working Papers 22-1391 https://ideas.repec.org/p/tse/wpaper/127593.html.

Chirimuuta, M. (2023). Rules, judgment and mechanisation. *Journal of Cross-disciplinary Research in Computational Law*, *1*(1) https://journalcrcl.org/crcl/article/view/22

Choi, J. and Schwarcz, D., (forthcoming). AI Assistance in Legal Analysis: An Empirical Study. *Journal of Legal Education* 73 http://dx.doi.org/10.2139/ssrn.4539836

Choi, J, Monahan, A, and Schwarcz, D. (forthcoming) Lawyering in the Age of Artificial Intelligence. Minnesota Law Review 109 http://dx.doi.org/10.2139/ssrn.4626276

Claudy MC, Aquino K and Graso M (2022) Artificial Intelligence Can't Be Charmed: The Effects of Impartiality on Laypeople's Algorithmic Preferences. *Frontiers in Psychology* 13:898027.

Cohen, J. 1995. *In the Shadow of the Law Clerk: Assessing the Roles of Law Clerks in the Judicial Process*, Long Term View 3, 99.

Contos G, Guyton J, Langetieg P, Vigil M. 2011. Individual taxpayer compliance burden: the role of assisted methods in taxpayers response to increasing complexity. In *IRS Research Bulletin: Proceedings of the IRS Research Conference 2010*, ed. ME Gangi, A Plumley, Washington, DC: Intern. Revenue Serv. 191–220.

Criddle C, 2023. 'Law firms embrace the efficiencies of artificial intelligence' *Financial Times* https://www.ft.com/content/9b1b1c5d-f382-484f-961a-b45ae0526675.

Crowe, J. (2014) Between Morality and Efficacy: Reclaiming the Natural Law Theory of Lon Fuller' *Jurisprudence* 5(1) 109.

Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, 7(3), 269–279.

Deng J, 2019. 'Should the Common Law System Be Intelligentized?: A Case Study of China's Same Type Case Reference System' Georgetown Law Technology Review 3, 223.

Dickerson, R. 1963 Some Jurisprudential Implications of Electronic Data Processing, *Law & Contemporary Problems* 28, 53-70.

Donelson R., and Hannikainen, Ivar. 2020. 'Fuller and the Folk: The Inner Morality of Law Revisited' in Tania Lombrozo, Joshua Knobe and Shaun Nichols (eds), *Oxford Studies in Experimental Philosophy* (Oxford University Press 2020) 6.

Dworkin, R. 1986. *Law's Empire*, Cambridge, MA: Harvard University Press.

Ekins, R. (2012). *The Nature of Legislative Intent*. Oxford: Oxford University Press.

Endicott T and Yeung K, 'The death of law? Computationally personalized norms and the rule of law' (2022) 72(4) University of Toronto Law Journal 373.

Flanagan, B. (2010). Revisiting the Contribution of Literal Meaning to Legal Meaning, *Oxford Journal of Legal Studies*, 30(2) 255–271.

Fuller, L.L. 1964. The Morality of Law, New Haven, CT: Yale University Press.

Gitonga, N. 2024. Justice denied as over 520,000 cases pend in courts for years. *People Daily* https://www.pd.co.ke/news/justice-denied-as-over-520000-cases-pend-in-courts-for-years-224522/

Greinke, A,. 1994. Legal Expert Systems: A Humanistic Critique of Mechanical Legal Interface, Murdoch University Electronic Journal of Law 1(4).

Guha, N., et al 2023. Legalbench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. Osgoode Legal Studies Research Paper No. 4583531, http://dx.doi.org/10.2139/ssrn.4583531

Gur, N., and Jackson J., 2020. Procedure–content Interaction in Attitudes to Law and in the Value of the Rule of Law, in Meyerson D. et al (eds.) *Procedural Justice and Relational Theory* Routledge 111-40.

Hannikainen I., et al., 2021. 'Are There Cross-Cultural Legal Principles? Modal Reasoning Uncovers Procedural Constraints on Law' (2021) *Cognitive Science* 45(8) e13024.

Hart, H.L.A. 1961. The Concept of Law, Oxford: Clarendon Press.

Henrich, J., Heine, S. & Norenzayan, A. (2010). Most people are not WEIRD. *Nature* 466, 29.

Hildebrandt M, 'Law as Information in the Era of Data-Driven Agency' (2016) 79(1) Modern Law Review 1.

Jones C, and Bergen B,. 2024. Does GPT-4 Pass the Turing Test?, arXiv:2310.20216

Klatt, M., 2008. *Making the Law Explicit* (Hart, Oxford)

Katz DM, Bommarito MJ, Gao S, Arredondo P. 2024. GPT-4 Passes the Bar Exam. *Philosophical Transactions of the Royal Society A* 382: 20230254 https://doi.org/10.1098/rsta.2023.0254

Lawlor, R. C., [1960] (1972). Excerpts from FACT CONTENT OF CASES AND PRECEDENT — A Modern Theory of Precedent. *Jurimetrics Journal*, *12*(4), 245–270.

Lichand G and Soares R R., (2014) 'Access to justice and entrepreneurship: evidence from Brazil's special civil tribunals' 57(2) *Journal of Law and Economics* 459.

Lim E, 2023. 'Law by Algorithm' *Oxford Journal of Legal Studies* 43(3) 650.

Lipez, K., 2007. Judges and Their Law Clerks: Some Reflections, 22 *Maine Bar Journal* 112.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents.* New York: Oxford University Press.

Livermore, M. 2020. "Rule by rules." In *Computational Legal Studies: The Promise and Challenge of Data-Driven Research*, edited by Ryan Whalen, 238–264. Edward Elgar.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: people prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* 151, 90–103.

Lucy W., (2023) Algorithms and adjudication, *Jurisprudence* https://www.tandfonline.com/doi/full/10.1080/20403313.2023.2243712

MacCormick N. 'Defeasibility in Law and Logic' in Z Bankowski and others (eds), *Informatics and the Foundations of Legal Reasoning* (Kluwer, MA 1995) 99.

Machery, E. 2017. *Philosophy within its Proper Bounds*. Oxford: Oxford University Press.

Marmor, A,. *Interpretation and Legal Theory* (Hart, Oxford 2005)

Martínez, E. 2024. Re-evaluating GPT-4's Bar Exam Performance. *Artificial Intelligence and the Law* https://doi.org/10.1007/s10506-024-09396-9

Maynez, J., Narayan, BB, and McDonald R., 2020. On Faithfulness and Factuality in Abstractive Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Association for Computational Linguistics) 1906–1919.

Merken S., 2023. New York lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief, https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

Newman, D. T., Fast, N. J., and Harmon, D. J. (2020). When eliminating bias isn't fair: algorithmic reductionism and procedural justice in human resource decisions. *Organ. Behav. Hum. Decis. Process.* 160, 149–167.

Newsom K. 2024. *Snell v. United Speciality Insurance Company* No. 22-12581 11th Circuit US Court of Appeals https://www.govinfo.gov/app/details/USCOURTS-ca11-22-12581/context.

Padilla J., 2023 https://www.bluradio.com/judicial/sentencia-la-tome-yo-chatgpt-respaldo-argumentacion-juez-de-cartagena-uso-inteligencia-artificial-pr30

Pasquale F., 2019. 'A Rule of Persons, Not Machines: The Limits of Legal Automation' George Washington Law Review 87(1) 1-55.

Peverall, J., 2020. Inside State Courts: Improving the Market for State Trial Court Law Clerks, University of Richmond Law Review 55, 227.

Posner R., 1985. *The Federal Courts: Crisis and Reform*. Harvard University Press.

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

Rachlinski J,. and Wistrich A., 2017. *Judging the Judiciary by the Numbers: Empirical Research on Judges*, 13 Annual Review of Law and Society 203.

Remington, C,. 2017. In J Tashea 'Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions,' American Bar Association Journal http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole.

Rodrik D, 'Institutions for high-quality growth: What they are and how to acquire them' (2000) Working Paper 7540, *National Bureau of Economic Research*, Cambridge http://www.nber.org/papers/w7540.

Rundle, K. 2012 *Forms Liberate: Reclaiming the Jurisprudence of Lon L Fuller* (Bloomsbury).

Sarid, E., and Ben-Zvi, O., (2024) Machine Learning and the Re-Enchantment of the Administrative State 87(2) *Modern Law Review* 371–397.

Sequeira S, (2016) 'Corruption, trade Costs, and gains from tariff liberalization: Evidence from southern Africa' 106(10) American Economic Review 3029.

Simmons, R., 2018.'Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System' (2018) 52 University of California Davis Law Review 1067.

Spamann H and Klohn L, 2016. *Justice is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges*, Journal of Legal Studies 45, 255.

Surden, H,. 2014 *Machine Learning and Law*, 89 Wash L Rev 87.

Susskind, R. (1986). Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning. *Modern Law Review*, *49*(2), 168–194.

Susskind, R. 2019. *Online Courts and the Future of Justice*. Oxford University Press.

Thompson N, Luo X, McKenzie B, Richardson E and Flanagan B (2024) User-Generated Content Shapes Judicial Reasoning: Evidence from a Randomized Control Trial on Wikipedia. Information Systems Research 0(0) https://doi.org/10.1287/isre.2023.0034

Volokh E, 'Chief Justice Robots' (2019) 68 Duke Law Journal 1135.

Wendel, WB., 2019. The Promise and Limitations of Artificial Intelligence in the Practice of Law, Oklahoma Law Review 72, 21.

World Justice Project, (2019) 'Measuring the justice gap: A Peoplecentered Assessment of Unmet Justice Needs around the World' Technical report, Washington, D.C. URL: https://worldjusticeproject.org/sites/default/files/documents/WJPMeasuring%20the%20Justice%.

Zhang, Y. et al. 2023 Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models, 2023, https://doi.org/10.48550/ARXIV.2309.01219.

Zuckerman, A., Artificial Intelligence in the Administration of Justice, In A Higgins (ed.) *The Civil Procedure Rules at 20*. Oxford University Press (2020) 291-304.