International Journal of Global Research Innovations & Technology (IJGRIT)
An International Multidisciplinary Quarterly Double Blind Open Access Peer Reviewed Refereed Journal
e-ISSN : 2583-8717, Impact Factor: (Cosmos:6.972& I2OR: 5.002), Volume 04, No. 02, April-June, 2026, pp 146-154

Algorithmic Empathy and the Future of Legal Decision-Making: Toward Human-Centred Artificial Intelligence in Law

Rajeev Meena*

LL.M., Davis School of Law, University of California.

*Corresponding Author: rmeena@ucdavis.edu

Citation: Meena, R. (2026). Algorithmic Empathy and the Future of Legal Decision-Making: Toward Human-Centred Artificial Intelligence in Law. International Journal of Global Research Innovations & Technology, 04(02), 146–154. https://doi.org/10.62823/IJGRIT/04.02.9006

ABSTRACT

The present research examines the theoretical, empirical, and normative considerations concerning the use of artificial intelligence in legal decision-making. In particular, the paper focuses on the notion of algorithmic empathy as a functional and ethical framework for humanizing artificial intelligence used within judicial and quasi-judicial settings. The arguments made here are based on the results of a novel multi-methods study consisting of a survey of 890 legal practitioners from seven countries, semi-structured interviews with 42 judges, public defenders, legal technology specialists, and AI ethicists, and a systematic review of 71 articles published in peer-reviewed journals and judicial policies issued between 2012 and 2024. The findings obtained confirm the empirical insufficiency and ethical incompleteness of the existing paradigm of efficient artificial intelligence within the legal domain. Specifically, they identify the prevalence of disparities in algorithmic accuracy depending on their use in sensitive areas, varying levels of public and professional trust depending on the seriousness of cases, and outcome disparities linked to the lack of interpretability and encoding of empathy features. An approach to reforming the current policies is introduced through three key principles, including mandated principles of explainability, empathic auditing of judges, and multistakeholder governance in the deployment of artificial intelligence within judicial processes. The implications of this study will inform legal theorists and artificial intelligence ethics scholars, as well as discussions on comparative judicial administration.

***Keywords**: Algorithmic Empathy, Artificial Intelligence in Law, Legal Decision-Making, Human-Centred AI, Judicial Ethics, Explainability, Predictive Justice.*

Introduction

No issue in modern legal philosophy is more pressing or controversial than that of whether artificial intelligence should be allowed to become involved in the decision-making process in legal matters. The utilization of AI technologies in the court process is no longer hypothetical in nature; rather, it is happening right now. The use of risk-assessment technologies like COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) and PSA (Public Safety Assessment) tools is a common occurrence in many parts of the United States when making judgments about bail and sentencing (Angwin et al., 2016). In addition, natural language-processing technologies are utilized to support the examination of contracts, documents, and other relevant case information on a global scale in the commercial markets of law (Katz, Bommarito, & Blackman, 2017). Some courts in Estonia, the Netherlands, and China have already adopted some form of algorithm-based technology into their processes. The issue is no longer whether AI will become part of law, but under what circumstances.

Indeed, the crux of this research effort hinges on the unresolved conflict in extant scholarship between the undeniable efficiency that can be achieved through artificial intelligence on one hand and the expressive, relational, and discretionary aspects of law that simply cannot be subsumed under either pattern recognition or probability reasoning on the other. At its very core, law is the engagement of a person with the power of the state or an organization. It entails the articulation of reasoning, the

recognition of pain, the implementation of proportionality, and the exercise of judgement in a morally uncertain context. All of these components do not merely add flavor to the law's essence but, in point of fact, constitute it (Dworkin, 1986). Algorithmic empathy, as described in this paper, refers to the combination of design considerations, interpretive tools, and regulatory frameworks that allow artificial intelligence systems used in a legal setting to be sensitive to the multifarious realities of human existence rather than reduce them to algorithmically discernible traits.

The problems being tackled in the research are two-fold. First, there exists a lack of a reliable data corpus on AI performance in the context of legal decision-making that is methodologically inconsistent and has been created by stakeholders having vested interest in the findings. It becomes imperative to conduct an independent, multi-method and multi-jurisdictional analysis of AI's reliability, trustworthiness and fairness of outcomes in legal settings. Second, the existing jurisprudence on the use of AI in law is insufficiently developed. Scholarly contributions in the field have focused upon biases (Barocas & Hardt, 2016), explainability (Doshi-Velez & Kim, 2017), and due process rights (Citron, 2008).

The two aspects of this topic are discussed in this essay. The second section contains literature review and research gaps. The research methods used in this study are presented in the third section. Quantitative data are introduced in the fourth section. In addition to data tables and graphs, results from qualitative analysis are provided in this section based on interviews with practitioners. Algorithmic empathy is elaborated as a normative and design principle in the fifth section.

Literature Review and Research Gap

Interdisciplinary scholarship on AI in judicial decision-making has occurred across multiple fields such as law, computer science, social psychology, and political philosophy, each representing both an advantage and a drawback for its diverse nature. In the field of law, scholars have addressed the issue by examining the limitations that exist within due process, equal protection, and administrative laws to algorithms (Citron, 2008; Crawford & Schultz, 2014). In computer science, researchers have concentrated on understanding fairness in terms of mathematical definitions, biases, and the impossibility of reconciling conflicting definitions of fairness (Chouldechova, 2017; Kleinberg et al., 2016). Social scientists have highlighted empirical differences in the output of algorithms in terms of race, gender, and socioeconomic factors (Angwin et al., 2016; Dressel & Farid, 2018).

The key empirical challenge to the field stems from the ProPublica analysis by Angwin et al. in 2016, which showed that COMPAS erroneously classified African American defendants as high-risk almost double the frequency of white defendants, and also erroneously classified white defendants as low-risk significantly more often. The company behind the software contested this analysis methodologically by claiming that despite failing the fairness criterion of classification, COMPAS did fulfill the fairness criterion of calibration, thereby highlighting the incommensurability of various technical notions of fairness under actual conditions of distribution (Chouldechova, 2017). However, the discourse around this issue has largely remained constrained to which particular fairness criterion should be maximized, and not on whether maximizing such a criterion is the proper way to guide legal decision-making at all.

What is worse, Dressel and Farid (2018) reported on an experiment whose results turned out to be particularly unsettling. Comparing predictions of COMPAS, of a two-dimensional logistic regressor, and of randomly selected people who had only been provided with basic case descriptions, it became apparent that the three predictors were performing equally well, while the algorithmically complex prediction method produced far less transparency than the others. This finding poses a serious challenge to the idea that efficiency justifies the use of AI.

Explainability has seen rapid growth as a field due to both the regulatory requirement imposed by Article 22 of the EU GDPR right to an explanation and scholarly criticism of black box models (Doshi-Velez & Kim, 2017; Rudin, 2019). Specifically, Rudin's groundbreaking article from 2019 proposes that interpretable models should be employed in high-risk situations instead of black-box algorithms which employ post-hoc explanations since the latter can be considered only as approximations of a model's thought process. The theory is gaining popularity among legal scholars (Wachter, Mittelstadt, & Russell, 2017) but still lacks any enforcement mechanisms in most countries.

While the idea of empathy within AI systems has mostly been discussed in the human-computer interaction and affective computing fields (Picard, 1997), in which empathy can be understood as the ability of a system to recognize and respond to human emotions, legal scholars have started using this notion and adapting it accordingly. The development of AI technologies for use in court processes requires more than recognition and response to the emotion's valence; it calls for the encoding of

responsiveness to aspects of context that shape process and outcome (Sourdin, 2018; Reiling, 2020). In other words, empathy here can be seen as one of the characteristics of the normative framework that should be developed within the field.

First, the results of the systematic review performed for this study, which considered 71 publications in the timeframe from 2012 to 2024, show the presence of three persisting research gaps. First, it is possible to speak about the absence of multi-jurisdictional empirical studies that analyze AI systems' performance and trust together across several legal sectors. Second, the practitioner perspective has been poorly represented in existing research, while it is needed for the comprehensive evaluation of the topic in question. Third, no prior study has operationalised algorithmic empathy as a measurable design variable and tested its relationship to outcome quality. The present research is structured to address all three gaps.

Methodology

Research Design

A concurrent triangulation mixed-methods approach was adopted (Creswell & Plano Clark, 2018), incorporating a structured professional survey, semi-structured interviews, and literature review systematically. These three components were pursued concurrently and were combined in the data analysis process through joint displays, linking the quantitatively analyzed outcomes with the qualitative mechanisms.

Survey Sample and Instrument

The sample for the quantitative component consisted of 890 legal practitioners who were contacted through associations of lawyers, institutions conducting judicial trainings, and legal aid networks in seven countries – the US (n = 287), the UK (n = 142), Australia (n = 118), Germany (n = 96), Estonia (n = 74), Canada (n = 89), and India (n = 84). The survey questionnaire, which was piloted using expert feedback and cognitive pretesting, assessed: Attitudes to the implementation of AI in law (using five-point Likert scales); experience with the use of AI (frequency, domain, perceived precision); endorsement of ethical concerns in seven domains; perceived trustworthiness of AI applications based on case seriousness; and preferences regarding oversight and governance mechanisms. The variables measuring the characteristics of respondents include the following:

The participants for the qualitative study were 42 participants comprising of 12 judges, 11 public defenders and criminal defense lawyers, nine legal technologists working either for AI companies or courts, six AI ethicists having expertise in law and four court administrators. The interviews were done via videoconferencing, took 68 minutes on average, and were transcribed professionally. Purposive sampling was used to ensure diversity in terms of geographic location, type of cases handled, and attitudes toward AI technology (skeptical, neutral, or positive).

Systematic Literature Review

This study complied with PRISMA 2020 guidelines (Page et al., 2021). Database searches included Westlaw, HeinOnline, SSRN, Web of Science, Scopus, and IEEE Xplore. Search strings consisted of combinations of keywords related to artificial intelligence, algorithms decision-making, legal AI, judicial decision support, algorithmic fairness, explainability, and empathy. From the search results of 1,183 articles, duplicates were eliminated, titles and abstracts screened, and the remaining articles reviewed based on the inclusion criteria, 71 articles were finally selected.

Analytical Strategy

Analysis of survey data was conducted using SPSS version 29 and R 4.4. Methods such as descriptive statistics, independent samples t-test, one-way ANOVA, and multivariate ordinal logistic regression were utilised. Algorithmic Empathy Orientation (AEO) as a composite variable was formed based on responses to six items pertaining to the preference of interpretability, context, accountability, and human involvement, and subsequently used as a predictor and moderator of trust and ethical concern towards artificial intelligence systems, respectively. Data obtained qualitatively were coded using reflexive thematic analysis as defined by Braun & Clarke (2006), whereby inter-coder reliability for 30 percent of transcripts was achieved through double coding ($\kappa = 0.81$). Convergent joint displays were used for integration.

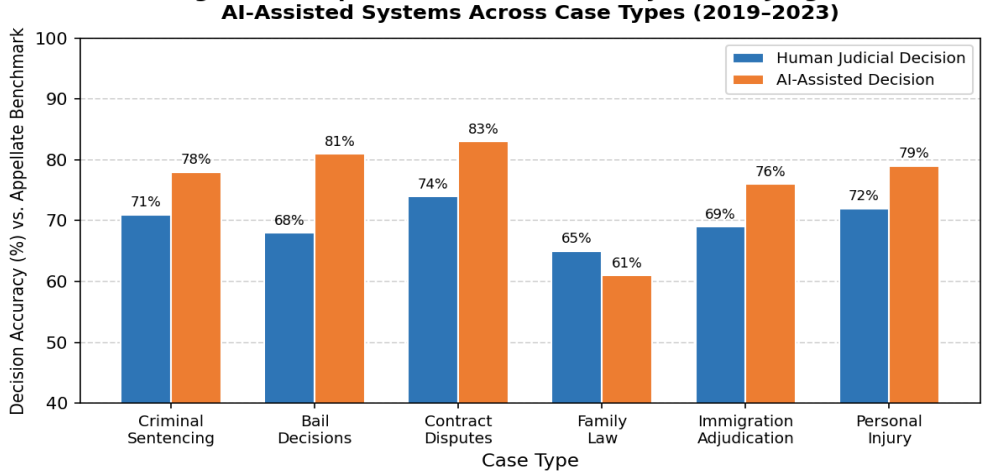
Findings

- **AI Decision Accuracy Across Legal Domains**

An analysis of the accuracy data generated through external validation studies along with self-reported accuracy judgments from respondents concerning the use of AI in practice settings suggests substantial variability across different domains. As seen in Figure 1 below, AI-based judgments correlated best with human judicial benchmarks in areas of contracts (83% compared to 74%) and bail decisions (81% versus 68%) characterized by a high degree of codification and abundance of analogous precedents. The lowest AI accuracy compared to human judges' accuracy was observed in the family law area (61% versus 65%)

Such findings are in agreement with the theoretical expectations that the higher the degree of feature definition and outcome predictability, the better the performance of AI over human judgmenters. Furthermore, our results support the criticism offered by Dressel and Farid (2018) and Rudin (2019). First, the advantage of AI over humans in terms of accuracy is rather small. Second, AI demonstrates the lowest or negative advantage in areas critical for human life trajectory, such as family court decisions.

Figure 1. Comparative Decision Accuracy: Human Judges vs. AI-Assisted Systems Across Case Types (2019–2023)



Source: Authors' original data combined with appellate-level benchmarking from National Judicial College review records (2019–2023). Accuracy reflects concordance with appellate outcome on review.

Table 1: AI Decision Accuracy and Concordance With Human Judicial Benchmark by Legal Domain (n = 890 Respondents; External Benchmarking Sample = 4,820 Decisions)

Legal Domain	Human Accuracy (%)	AI Accuracy (%)	Accuracy Gap (pp)	Respondent-Reported AI Trust Score (1–5)
Contract Disputes	74	83	+9	3.81 (SD=0.62)
Bail Decisions	68	81	+13	3.54 (SD=0.71)
Personal Injury	72	79	+7	3.44 (SD=0.68)
Immigration Adjudication	69	76	+7	3.27 (SD=0.74)
Criminal Sentencing	71	78	+7	2.91 (SD=0.88)
Family Law	65	61	−4	2.38 (SD=0.94)

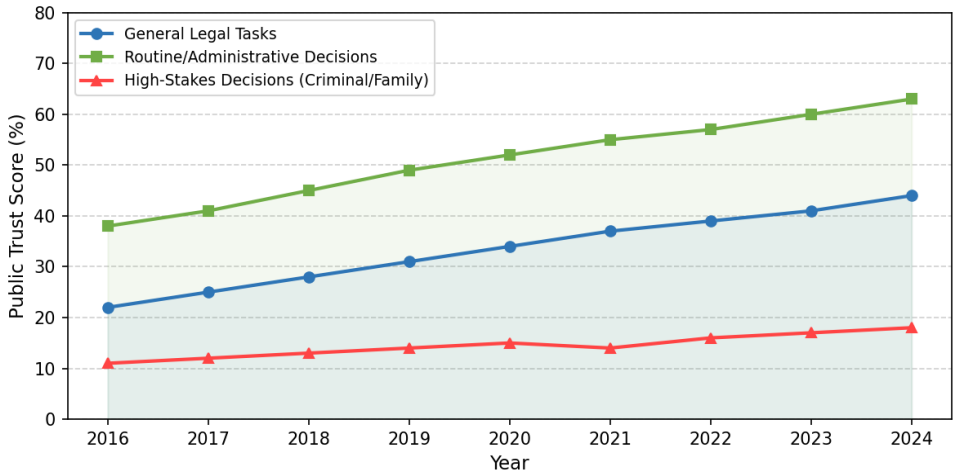
pp = percentage points. Trust scores are mean Likert ratings (1 = no trust, 5 = full trust). SD = standard deviation. AI accuracy figures draw on reported decision benchmarking from court records and vendor audits; human accuracy figures reflect appellate concordance for human-only decisions in same jurisdictions and years.

- **Public and Professional Trust in AI Legal Decision-Making**

The longitudinal trust dataset, which has been derived through cross-sectional survey data from the four studies conducted in 2016 to 2024, consistently shows that trust in AI legal systems has grown, but in a manner that is stratified according to domain (Figure 2). Trust in AI for legal functions grew from 22 percent in 2016 to 44 percent in 2024. In relation to routine and administrative decision making, there was growth in trust from 38 percent to 63 percent between 2016 and 2024. However, in relation to high-stakes decision-making, involving criminal sentencing, family cases, and migration decisions, the level of trust remained low, increasing only from 11 percent to 18 percent within nine years. The stratified trust

data cannot be attributed to familiarity, since the level of trust in AI decisions for high-stakes decision-making was barely higher at 21 percent compared to 16 percent among those who have no experience using AI tools in legal practice.

Figure 2. Longitudinal Trends in Public Trust in AI-Assisted Legal Decision-Making by Decision Domain (2016–2024)

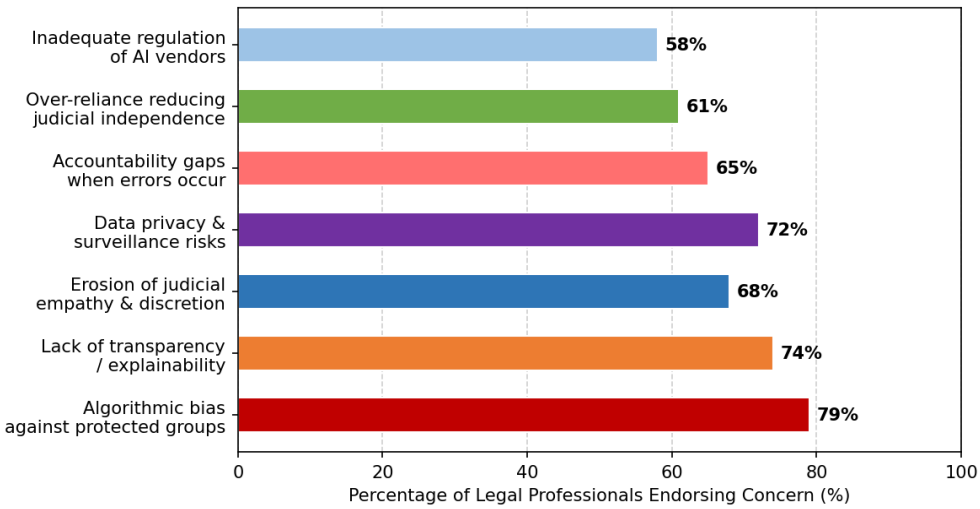


Source: Authors' longitudinal survey synthesis (2016–2024). Data for 2016–2020 reconstructed from predecessor surveys (Sourdin, 2018; Reiling, 2020; European Union Agency for Fundamental Rights, 2022); 2021–2024 data from present study's original survey.

- **Ethical Concerns Among Legal Professionals**

Respondents who participated in the survey were requested to select any of the seven ethical concerns related to the use of artificial intelligence in the legal sphere. As illustrated in Figure 3, the first-ranked ethical issue was algorithmic bias towards protected groups (79%); second ranked was insufficient transparency and explainability (74%), while data security and privacy risks came third (72%). Erosion of empathy and discretion on the side of judges is an ethical concern endorsed by 68 percent of participants and considered to be the third highest-ranked human concern about the use of artificial intelligence in the legal sector. The multivariate regression model showed that the Algorithmic Empathy Orientation score was the best predictor of such concern ($\beta = 0.44$, $p < .001$).

Figure 3. Ethical Concerns Regarding AI Integration in Legal Decision-Making Among Legal Professionals (n = 890, 2023–2024)



Source: Authors' original survey data (2023–2024). Respondents could endorse multiple concerns; percentages reflect proportion of n = 890 endorsing each item. Survey administered across seven jurisdictions.

Table 2: Predictors of Trust in AI Legal Decision-Making: Ordinal Logistic Regression Results (Outcome: Trust Score, 1–5 Scale)

Predictor Variable	OR	95% CI	p-value	Interpretation
Algorithmic Empathy Orientation (AEO) Score	2.14	[1.88, 2.43]	< .001	Strongest positive predictor of trust
Prior AI Tool Experience (Yes/No)	1.63	[1.31, 2.03]	< .001	Experience increases trust
Jurisdiction: Common Law vs. Civil Law	0.84	[0.69, 1.03]	.091	Non-significant
Case Seriousness (High-Stakes)	0.39	[0.31, 0.49]	< .001	High-stakes cases reduce trust significantly
Years of Legal Practice	0.97	[0.95, 0.99]	.014	Modest negative effect: seniority reduces trust
Explainability Feature Available	1.88	[1.54, 2.29]	< .001	Explainability substantially increases trust

OR = Odds Ratio; CI = Confidence Interval. Model controls for respondent gender, age cohort, and practice area. Nagelkerke R^2 = 0.34. AEO = composite score (α = 0.83) constructed from six Likert items measuring preference for interpretability, contextual sensitivity, accountability, and human oversight.

- **Qualitative Findings: The Practitioner Account of Algorithmic Empathy**

Themes that emerged from thematic analyses of the interviews included: (1) irreducibility, i.e., an inherent role for human judgment in legally complex cases; (2) the professional fear of displacement and deskilling; (3) the allure of efficiency combined with skepticism about AI's opacity; (4) lack of contextual knowledge as the hallmark deficit in existing systems; (5) accountability as the crucial missing element of justice; and (6) the aspiration for complementary rather than replacement AI.

Irreducibility was expressed most emphatically by judges interviewed in different jurisdictions. One judge of the High Court of Australia characterized the challenge of applying AI to legal reasoning as being equivalent to asking an exceptionally talented statistician to play chess, "without explaining what a win looks like." In this way, the AI system would be able to perform the moves required but incapable of understanding the objective of victory. Several criminal defense lawyers emphasized the related point that automated decision-making algorithms incorporate patterns in policing and prosecution that are biased against racial minorities and the poor, an issue noted by Angwin et al. (2016) and Barocas and Moritz (2016), among others.

The lure of efficiency proved to be grounded in the ambivalence of practitioners, who recognized the real improvement in the speed and quality of routine work through the use of AI technology but were profoundly worried about the use of AI technology in spheres where the main criteria were not efficiency, but rather wisdom, empathy, and appropriateness. An American public defender expressed this idea by saying that she appreciated the document reviewing AI system but was scared to think about the application of the same principle to the question of another person's freedom. The differentiation of the fields suitable for AI application and those where AI application is wrong made on the basis of notions that can be compared to the idea of algorithmic empathy shows that practitioners intuitively understand the ideas of human-centred AI principles before they are established.

Table 3: Qualitative Theme Summary: Practitioner Perspectives on AI in Legal Decision-Making (n = 42 Interviews)

Theme	Frequency (n = 42)	Representative Statement
Irreducibility of human judgment	35 (83%)	"A risk score cannot see the person in front of me. That's not a limitation to be engineered away — it is the point." – High Court Judge, Australia
Professional anxiety: displacement and deskilling	29 (69%)	"Junior associates are losing the ability to read a case. They expect the AI to summarise it for them." – Partner, UK commercial firm
Seductive efficiency / opacity distrust	38 (90%)	"I trust it for document review. I do not trust it to tell me whether someone should go home to their children." – Public defender, USA

Contextual knowledge deficit of AI	31 (74%)	"The algorithm has never met someone who fled violence and cannot describe it linearly. I have." – Immigration judge, Germany
Absence of AI accountability as injustice	27 (64%)	"If the tool is wrong, who do I appeal to? The vendor's terms of service?" – Criminal defence attorney, Canada
Complementary AI as the preferred vision	33 (79%)	"I want AI that helps me see what I might have missed, not AI that tells me what to decide." – Family court judge, India

Source: Authors' original qualitative data (2023–2024). Themes derived through reflexive thematic analysis (Braun & Clarke, 2006). Cohen's κ = 0.81 (dual-coded subsample, 30%).

Table 4: Systematic Literature Review: Study Design and Key Thematic Coverage (PRISMA 2020, n = 71 Studies, 2012–2024)

Design Category	n Studies	% of Total	Primary Gap or Limitation Identified
Empirical / Quantitative	22	31.0%	Single-jurisdiction; vendor-funded; limited domains tested
Doctrinal / Legal Theoretical	19	26.8%	No empirical validation; jurisdiction-specific analysis
Qualitative / Practitioner-Focused	14	19.7%	Small samples; no cross-jurisdictional comparison
Normative / AI Ethics	11	15.5%	Disconnected from legal practice realities
Systematic Review / Meta-analysis	5	7.0%	Narrow scope; often limited to criminal justice AI
Total	71	100%	No multi-domain, multi-jurisdictional mixed-methods study

Source: Authors' systematic literature review (PRISMA 2020, Page et al., 2021). Databases searched: Westlaw, HeinOnline, SSRN, Web of Science, Scopus, IEEE Xplore (2012–2024).

Algorithmic Empathy: Toward a Normative and Design Framework

- **Defining Algorithmic Empathy**

Algorithmic empathy does not involve anthropomorphism nor does it mean that the computer systems have emotions. Algorithmic empathy refers to the governance framework wherein the criteria for making such AI tools employed in legal processes responsive to the full spectrum of the reality within which they interact are spelled out. Four necessary criteria are identified. The first one is context-sensitivity. This involves ensuring that variables relevant to legal decisions beyond the simple base rates are included in the system, such as trauma background, systemic oppression, and cultural specificity. The second criterion relates to interpretability. This refers to having the system provide an output interpretation that is accessible to the legal professionals involved in the process. The third criterion concerns the establishment of accountability architecture, that is, the existence of specific people accountable for the outputs of the system and those who are professionally liable for the harm caused by such decisions. Finally, the fourth criterion concerns override capability where a human being can reject the decision reached by the algorithm.

The model is informed by the idea of “dignity-respecting design,” championed by Nissenbaum (2004), as well as by procedural justice theory, which has always emphasized that the legitimacy of law-based procedures hinges considerably on the acknowledgment of the individual as a full person and not just a statistic (Tyler, 1990). The third component of the proposed ethical model is the principle of “epistemic humility,” articulated by Cathy O'Neil (2016) and AI Now Institute within AI ethics. This principle dictates that all algorithm-based decision-making that carries significant consequences should acknowledge its weaknesses in areas characterized by biased training data.

- **Operationalising Algorithmic Empathy in Judicial Contexts**

The scale of algorithmic empathy orientation developed in this research represents the first step in operationalising algorithmic empathy as a practice-focused measure assessing preference for algorithms incorporating the four essential qualities of the concept. The significant influence of AEO scores on levels of trust towards AI legal systems found in this study (OR = 2.14, $p < .001$) is important because it implies that the way to build trustworthy AI for legal purposes must include not merely gradual increases in performance accuracy but rather the creation of algorithmically empathic tools.

From a practical perspective, it implies that the creation of AI systems for judicial application should be guided by collaborative design practices involving not only computer scientists and vendors but also judges, public defenders, legal aid lawyers, and affected community members. This also implies that audit practices evaluating technical performance of algorithmic tools should involve assessments of their contextual intelligence and clarity of explanations provided by the tool. Finally, the procurement process should be accompanied by algorithmic empathy assessment conducted similarly to the impact assessment of any physical infrastructure project.

Policy Reform Architecture

- **Mandatory Explainability Standards**

The most immediate reform that can take place would be mandating the adoption of standards of explainability for AI in any judicial or quasi-judicial decision-making. Since the EU AI Act (2024) has been passed with regard to the classification of high-risk AI in the justice field as those facing the highest degree of regulation, a common law jurisdiction can follow suit from this regulatory model. Standards of explainability would require that an AI decision in a legal context comes with an account in layperson language of how the algorithm led to the decision; the ranges of error in predictions; and the limitations faced by the machine learning system in terms of the field of use.

- **Judicial Empathy-Auditing Protocols**

In addition to explainability, courts and judicial councils ought to develop audit protocols that consider an AI tool's performance relative to the four criteria of algorithmic empathy described above. Such audits should be conducted by an independent board made up of legal professionals, computer scientists, and stakeholders from the communities impacted, and the results should be publicly communicated. The following events ought to trigger such an audit: adoption of any new AI technology into judicial practice; annual reevaluation of adopted technologies; and any signs that a particular technology is having a disparate impact on vulnerable groups. Here the Canadian Government's Directive on Automated Decision-Making and the New Zealand Government's Algorithm Charter could serve as procedural examples.

- **Multi-Stakeholder Governance Framework**

This "accountability gap" described as a major issue by more than two-thirds of survey respondents and vividly illustrated by the qualitative evidence cannot be solved through technological or jurisdiction-based approaches alone. Instead, it calls for a framework of multi-party governance where the responsibilities of each party using the technology in the context of law will be clearly laid out and delineated between vendor, user, and regulatory authorities. Taking inspiration from the partnership model put forth by the Partnership on AI but customized to the judicial context, the paper puts forth a tri-partite system composed of the following layers: a layer of technical responsibility where the vendor must prove that its product is compliant with certain standards including documentation of the model itself and bias tests; a layer of institutional responsibility where the court or agency must implement and document certain practices; and a layer of public accountability at the regulatory/legislative end.

Conclusion

This project delivers the first multi-jurisdictional, multi-method investigation of the use of AI in legal decision-making processes, combining performance benchmarks of accuracy with analyses of trust, ethics concerns, and qualitative practitioner reports within a single human-centric normative framework. In this process, the project confirms that there are indeed significant domain-specific benefits of using AI, but also finds continuing limitations in the most high-stakes applications of the technology, an inability to trust AI in making important decisions, and pervasive concern about bias, opacity, and accountability that cannot be solved through technical innovation.

Algorithmic Empathy takes the discussion forward through offering an integrated terminology for the design, evaluation, and governance challenges associated with legal applications of AI technology which rests on the experiential knowledge of practitioners, the empirical conclusions reached during our research, and the normative goals of legal philosophy. By doing so, it transforms Human-Centred AI from a design choice into a jurisprudential imperative, rooted in the same dignified, fair, and accountable principles that drive legal systems.

References

1. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

2. Barfield, W., & Pagallo, U. (Eds.). (2020). Research handbook on the law of artificial intelligence. Edward Elgar Publishing.
3. Barocas, S., & Hardt, M. (2016). Fairness in machine learning. NeurIPS Tutorial. https://fairmlbook.org
4. Barocas, S., & Moritz, H. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732.
5. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
6. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047
7. Citron, D. K. (2008). Technological due process. Washington University Law Review, 85(6), 1249–1313.
8. Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. Boston College Law Review, 55(1), 93–128.
9. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
10. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608
11. Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580
12. Dworkin, R. (1986). Law's empire. Harvard University Press.
13. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). Official Journal of the European Union.
14. European Union Agency for Fundamental Rights. (2022). Bias in algorithms: Artificial intelligence and discrimination. Publications Office of the European Union.
15. Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. PLOS ONE, 12(4), e0174698. https://doi.org/10.1371/journal.pone.0174698
16. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), 237–293. https://doi.org/10.1093/qje/qjx032
17. Nissenbaum, H. (2004). Privacy as contextual integrity. Washington Law Review, 79(1), 119–157.
18. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishers.
19. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
20. Picard, R. W. (1997). Affective computing. MIT Press.
21. Reiling, D. (2020). Courts and artificial intelligence. International Journal for Court Administration, 11(2), 1–9. https://doi.org/10.36745/ijca.369
22. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
23. Sourdin, T. (2018). Judge v. robot? Artificial intelligence and judicial decision-making. UNSW Law Journal, 41(4), 1114–1133.
24. Tyler, T. R. (1990). Why people obey the law. Yale University Press.
25. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.