

The Roles AI Is (and Should Be) Playing in Dispute Resolution

On the Legitimacy of AI Legal Applications

Talia Schwartz-Maor and Orna Rabinovich-Einy

4.1 INTRODUCTION

While the transformation of legal processes and roles by new technologies has been the subject of a heated debate for some time,¹ the rise of large language models (LLMs) represents a ‘game changer’ in terms of the capabilities of artificial intelligence (AI) in the professional domain.² ChatGPT-4 has demonstrated human-level performance on an array of academic and professional tasks, including the Law School Admission Test, the Medical Knowledge Self-Assessment Program, an advanced macroeconomics course designed for high school students, and passing a bar exam.³ Other LLMs were quick to follow, including Claude, Gemini, and Perplexity.⁴

Focusing on the legal domain, LLMs have shown competence on a wide range of tasks, including assistance in writing a legal paper,⁵ passing first-year law exams with a C+,⁶ preparing cross-examination questions for a case (performing well, though inferior to lawyers),⁷ drafting a complaint in a securities cryptocurrency class action lawsuit,⁸ determining

This study was supported by the European Union (ERC-2021-COG and 101044195). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

¹ While some have long envisioned dramatic changes to the legal professions (see Richard Susskind and Daniel Susskind, *The Future of the Professions* (Oxford University Press 2015); Herbert M. Kritzer, ‘The Professions Are Dead, Long Live the Professions: Legal Practice in a Postprofessional World’ (1999) 33(3) *L & Soc’y Rev* 713; M. Ethan Katsh, *Electronic Media and the Transformation of Law* (Oxford University Press, 1989), others were more sceptical of the idea that machines could substitute for the more challenging aspects of professional work. See Frank Pasquale and L. A. Review of Books, 15 March 2016, ‘Book Review: Automating the Professions?’ (2016) U of Maryland Legal Studies Research Paper No. 2016–21 <<https://ssrn.com/abstract=2775397>> accessed 3 November 2024; Dana Remus and Frank S Levy, ‘Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law’ (2015) *SSRN Electron J* <<http://doi.org/10.2139/ssrn.2701092>> accessed 3 November 2024.

² Rishi Bommasani et al., ‘On the Opportunities and Risks of Foundation Models’ (2022) <<https://doi.org/10.48550/arXiv.2108.07258>> accessed 3 November 2024.

³ OpenAI et al., ‘GPT-4 Technical Report’ (2024) <<https://doi.org/10.48550/arXiv.2303.08774>> accessed 3 November 2024.

⁴ David Brown, ‘2024 AI Showdown: GPT-4o, Perplexity, Google Gemini, and Claude 3 Compared’ (*perplexity*, 24 July 2024) <www.perplexity.ai/page/2024-ai-showdown-gpt-4o-perple-OU.CI7U_RxKW9NaE3WmEhQ> accessed 9 November 2024.

⁵ Andrew Perlman, ‘The Implications of OpenAI’s Assistant for Legal Services and Society’ (2022) Suffolk University Law School Research Paper No. 22-14 <<http://doi.org/10.2139/ssrn.4294197>> accessed 3 November 2024.

⁶ Jonathan H. Choi et al., ‘ChatGPT Goes to Law School’ (2023) 71(3) *J Leg Ed* 387, 387–400.

⁷ Kwan Yuen Iu and Vanessa Man-Yi Wong, ‘ChatGPT by OpenAI: The End of Litigation Lawyers?’ (2023) *SSRN Electron J* <<http://doi.org/10.2139/ssrn.4339839>> accessed 3 November 2024.

⁸ Arianna Trozze, Toby Davies, and Bennett Kleinberg, ‘Large Language Models in Cryptocurrency Securities Cases: Can ChatGPT Replace Lawyers?’ (2023) <<http://dx.doi.org/10.48550/arXiv.2308.06032>> accessed 4 November 2024.

employment status with over 90 per cent accuracy,⁹ and exhibiting fair or above-average performance in legal reasoning tasks.¹⁰ Specifically, one study found that GPT-4 performed well, with an average score of 77.32 out of 100 combined, on all legal reasoning tasks.¹¹ Researchers have found that ChatGPT's strength lies in its performance on a procedural level, specifically its 'ability to understand and reason about users' needs'.¹² LLMs have also been tested for their ability to resolve disputes, with some studies awarding them high ratings for their choice of intervention strategy and message wording as mediators, in comparison with humans.¹³ Another recent study, in which researchers built an AI-based arbitration model, points to excellent performances.¹⁴ These capabilities are particularly exciting as they bring to the fore what have traditionally been considered human 'soft skills' that have to do with emotional intelligence.

Alongside the excitement and hype, certain studies point to the underperformance of foundation models in the legal domain, including errors in statutory reasoning resulting from imperfect prior knowledge,¹⁵ as well as weak legal reasoning capabilities in certain contexts, such as cryptocurrency securities and taxes.¹⁶ In addition, LLMs have been found to perform poorly in complex legal tasks, such as legal research and reasoning in certain models (GPT-3.5, PaLM 2, and Llama 2),¹⁷ or yield outputs that exhibit a lack of concision and citation.¹⁸ Such findings reinforce concerns regarding AI bias that predated LLMs and were strongly associated with legal tools such as COMPAS (a recidivism prediction system widely used in the US).¹⁹ Hallucinations have also become a concern, with LLMs occasionally producing outputs that are false, with high certainty.²⁰ At the same time, more recent studies on LLMs' performance in fulfilling complex tasks and making predictions in the legal domain demonstrate more

- ⁹ M. C. Cohen, S. Dahan, W. Khern-Am-Nuui, H. Shiao, and J. Touboul, 'The Use of AI in Legal Systems: Determining Independent Contractor vs. Employee Status' (2023) 30 March *Artif Intell Law* 10.1007/s10506-023-09353-y.
- ¹⁰ L. Martin, N. Whitehouse, S. Yiu, L. Catterson, and R. Perera (2024). Better Call GPT, Comparing Large Language Models Against Lawyers. arXiv preprint 2401.16212.
- ¹¹ Neel Guha et al., 'LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models' (2023) <<https://doi.org/10.48550/arXiv.2308.11462>> accessed 3 November 2024.
- ¹² Jinzhe Tan, Hannes Westermann, and Karim Benyekhlef, 'ChatGPT as an Artificial Lawyer?', *Workshop on Artificial Intelligence for Access to Justice* <<https://ceur-ws.org/Vol-3435/short2.pdf>> accessed 3 November 2024.
- ¹³ Jinzhe Tan et al., 'Robots in the Middle: Evaluating LLMs in Dispute Resolution' (2024) <<https://doi.org/10.48550/arXiv.2410.07053>> accessed 3 November 2024. See also Hannes Westermann, Jaromir Savelka, and Karim Benyekhlef, 'LLMediator: GPT-4 Assisted Online Dispute Resolution' (2023) <<https://doi.org/10.48550/arXiv.2307.16732>> accessed 3 November 2024 (showing promising results for LLM mediator).
- ¹⁴ Jack Kieffaber, Kimo Gandall, and Kenny McLaren, 'We Built Judge.ai. And You Should Buy It' (28 January 2025). Available at SSRN: <<https://ssrn.com/abstract=5115184>> or <<http://doi.org/10.2139/ssrn.5115184>> (finding zero hallucinations and zero incomplete responses in an AI arbitration model).
- ¹⁵ Andrew Blair-Stanek, Nils Holzenberger, and Benjamin Van Durme, 'Can GPT-3 Perform Statutory Reasoning?', *ICAIL 2023: Nineteenth International Conference on Artificial Intelligence and Law* (ACM 2023) <<http://doi.org/10.1145/3594536.3595163>> accessed 3 November 2024.
- ¹⁶ For cryptocurrency securities, see Trozze, Davies, and Bennett (n 8). For taxes, see John Nay et al., 'Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence' (2023) Minnesota Legal Studies Research Paper No. 23-15, Northwestern Public Law Research Paper No. 23-48 <<http://doi.org/10.2139/ssrn.4476325>> accessed 3 November 2024 (finding LLMs to perform at high levels of accuracy but not as high as an expert tax lawyer would).
- ¹⁷ Matthew Dahl et al., 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' (2024) 16(1) *J Legal Anal* 64.
- ¹⁸ R. Bhambhoria, S. Dahan, J. Li, and X. Zhu, 'Evaluating AI for Law: Bridging the Gap with Open-Source Solutions' (2024) arXiv preprint 2404.12349.
- ¹⁹ Julia Angwin et al., 'Machine Bias' (*ProPublica*, 23 May 2016) <www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 3 November 2024.
- ²⁰ Dahl (n 17).

promising results and greater accuracy,²¹ which could explain why an ever-growing number of professionals and laypeople are relying on such tools for a wide array of uses.²² In some cases, professionals choose to rely on AI tools trained, tweaked, and refined for different legal tasks and contexts using various methods, for example Retrieval-Augmented Generation (RAG).²³

In addition to the use of LLMs and other legal tech products by professionals and laypeople, a growing number of AI-based applications are being developed and implemented in the justice system. Uses of AI in these settings are expanding and can include legal information retrieval and assessment tools, case management, legal research, summary and translation, writing applications, document retrieval, analysis and drafting, trial preparation functions, case analytics, negotiation, predictions, and co-pilot tools in judicial decision-making or alternative dispute resolution (ADR) processes.²⁴ While most courts currently shy away from AI-based decision-making,²⁵ and are still in the experimentation phase with many of the above-mentioned tools, a host of professional decision-making functions in both public administrative and private settings are already performed by AI, with varying degrees of human involvement and/or oversight.²⁶

²¹ Daniel Martin Katz et al., 'GPT-4 Passes the Bar Exam' (2024) 382 *Phil Trans R Soc A* 10.2139/ssrn.4389233; Nay et al., 'Large Language Models as Tax Attorneys' (n 18); X. Wang, X. Zhang, V. Hoo, Z. Shao, and X. Zhang, 'LegalReasoner: A Multi-Stage Framework for Legal Judgment Prediction via Large Language Models and Knowledge Integration' (2024) 12 *IEEE Access* 166843–166854; see also Francesco Contini, 'Unboxing Generative AI for the Legal Professions: Functions, Impacts and Governance' (2024) 15(2) *Int'l J Court Admin* 1 (discussing some of the changes to Gen AI introduced as of 2023 that improved accuracy in domain-specific tasks).

²² 'The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023' (GOV.UK, 1 November 2023) <www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> accessed 3 November 2024, acknowledging that 'AI systems are already deployed across many domains of daily life including housing, employment, transport, education, health, accessibility, and justice, and their use is likely to increase'; Kathrin Eidenmueller, Conor McLaughlin, and Horst GM Eidenmueller, 'Expanding the Shadow of the Law: Designing Efficient Judicial Dispute Resolution Systems in a Digital World – An Empirical Investigation' (2024) 29(2) *Harv Negot L Rev* 10.2139/ssrn.4686785 (findings that the usage of AI tools across different functions by UK legal practitioners has more than doubled within the period from 2019/2020 to 2022/2023).

²³ N. Wiratunga et al., 'CBR-RAG: Case-based Reasoning for Retrieval Augmented Generation in LLMs for Legal Question Answering', in *International Conference on Case-Based Reasoning* (Springer Nature, 2024) 445–460.

²⁴ See chapter on 'law' and legal applications in Bommasani et al. (n 2); Hadar Yoana Jabotinsky and Michal Lavi, 'AI in the Courtroom: The Boundaries of RoboLawyers and RoboJudges' (2025) 35 *Fordham Intell Prop Media & Ent L J* 286.

²⁵ But see Vikas Mahendra and Arunima Athavale, 'Use and Regulation of AI in Dispute Resolution: Focus on the United Kingdom, Singapore and India' (2024) 18(1) *DRI* 5 (describing the integration of AI into judiciary in several countries, specifically Singapore); Tania Sourdin, 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making' (2018) 41(4) *UNSWLJ* 1114 <<http://doi.org/10.53637/zgux2213>> accessed 3 November 2024; 'CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment – European Commission for the Efficiency of Justice (CEPEJ) – www.coe.int' (European Commission for the Efficiency of Justice (CEPEJ)) <www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment> accessed 3 November 2024.

²⁶ As regards public administrative settings, finding that nearly half of the studied federal agencies (45 per cent) have experimented with AI and ML in a wide array of use types from enforcement tasks to analysis use-cases, and from provision of services to the public to internal management applications (David Freeman Engstrom et al., 'Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies' (2020) NYU School of Law, Public Law Research Paper No. 20-54 <<http://doi.org/10.2139/ssrn.3551505>> accessed 3 November 2024). As regards decision-making functions in private settings, see Jack Kelly, 'How Companies Are Hiring and Reportedly Firing With AI' (*Forbes*, 4 November 2023) <www.forbes.com/sites/jackkelly/2023/11/04/how-companies-are-hiring-and-firing-with-ai/> accessed 3 November 2024 (stating a 2023 survey that found 98 per cent of surveyed HR executives said they plan to employ software and algorithms to reduce labour costs); Rachel Curry, 'Microsoft, Amazon Among the Companies Shaping AI-enabled Hiring Policy' (CNBC, 11 October 2023) <www.cnbc.com/2023/10/11/microsoft-amazon-among-the-companies-shaping-ai-enabled-hiring-policy.html> accessed 3 November 2024; Vijay Pereira et al., 'A Systematic Literature Review on the Impact of Artificial Intelligence on Workplace Outcomes: A Multi-Process Perspective' (2023) 33 *Hum Resour Manag Rev* 100857.

In addition, there is evidence that some judges are using these tools on their own initiative, with some courts attempting to regulate such use.²⁷ Parties to a legal dispute (as well as their counsel) may also rely on their own AI tools in assessing the strength of their court case, developing their legal strategy, and their resolution efforts.

The expansion in AI capabilities and uses in the legal domain raises pressing and difficult questions about the potential and the limits of automation.²⁸ In this chapter, we look to the perceptions of laypeople and professionals on these questions as a significant prism for guiding societal decisions on the desirability of AI use in legal and dispute resolution settings. Such perceptions also constitute important means for assessing the legitimacy-related implications of the adoption of various AI-based legal tools and functions. We, therefore, begin with a review of legal AI tools and functions in Section 4.2 of this chapter. In Section 4.3, we draw on empirical research examining the perceptions of laypeople and professionals regarding the legitimacy of AI use by courts and in alternative dispute resolution processes. Section 4.4 offers concluding thoughts stemming from the juxtaposition of the review of AI capabilities and existing empirical research.

4.2 AI LEGAL TOOLS AND FUNCTIONS: AN OVERVIEW

Two key developments lie at the heart of the current trend of AI-based legal tools: the evolution of the ‘fourth party’ as part of the rise of online dispute resolution (ODR) on the one hand and the growth of a legal-tech industry on the other.²⁹ ODR emerged in the mid 1990s in the e-commerce setting with the spread of internet communication. As remote transactions proliferated, it soon became clear that there was a need for online avenues of dispute resolution, with both courts and ADR avenues typically being unavailable or impractical for distant parties involved in small-scale disputes.³⁰ eBay quickly became a pioneer, developing a first-of-its-kind ODR system, involving a tiered process that made successful use of automation, offering disgruntled buyers and sellers an avenue of redress.³¹ The role of technology in the resolution of disputes was termed by Ethan Katsh and Janet Rifkin as the ‘fourth party’, either supporting or supplanting the human ‘third party’.³² In the decades that followed, ODR processes expanded from private settings to courts and tribunals and from online small-scale

²⁷ Ralph Losey, ‘Eleventh Circuit Judge Admits to Using ChatGPT to Help Decide a Case and Urges Other Judges and Lawyers to Follow Suit’ (JDSUPRA, 7 June 2024) <www.jdsupra.com/legalnews/eleventh-circuit-judge-admits-to-using-1737989/> accessed 13 January 2025. Jennifer Kay, ‘Delaware Judges, Law Clerks Cleared to Use Generative AI Tools’ (22 October 2024) <<https://news.bloomberglaw.com/artificial-intelligence/delaware-judges-law-clerks-cleared-to-use-generative-ai-tools>> accessed 1 February 2024.

²⁸ Amy J. Schmitz, ‘Chapter 1. Responsible Use of AI in Civil Dispute Resolution’ (2024) Ohio State Legal Studies Research Paper No. 870 <<https://doi.org/10.2139/ssrn.4903238>> accessed 4 November 2024 (elaborating on the benefits and risks of AI in dispute resolution).

²⁹ For ODR see, Ethan Katsh and Janet Rifkin, *Online Dispute Resolution: Resolving Conflicts in Cyberspace* (John Wiley & Sons, 2001). For the growth of a legal-tech industry see David Freeman Engstrom and Jonah B Gelbach, ‘Legal Tech, Civil Procedure, and the Future of Adversarialism’ (2021) *U Pa L Rev* 1001–1099; Richard Susskind, *Tomorrow’s Lawyers: An Introduction to Your Future* (Oxford University Press, 2023) 136; Robert Dale, ‘Law and Word Order: NLP in Legal Tech’ (2018) 25(1) *Nat Lang Eng* 211, 211–217 <<http://doi.org/10.1017/S1351324918000475>> accessed 4 November 2024; Qian Hongdao et al., ‘Legal Technologies in Action: The Future of the Legal Market in Light of Disruptive Innovations’ (2019) 11(4) *Sustainability* 1015 <<http://doi.org/10.3390/su11041015>> accessed 4 November 2024.

³⁰ Ethan Katsh and Orna Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (Oxford University Press, Incorporated, 2017) 35–36.

³¹ *Ibid.*

³² Katsh and Rifkin (n 29) 93.

disputes to a wide variety of claims, ranging from small claims and traffic cases to family and eviction claims.³³

In the post-COVID-19 years, various mediums inside and outside the courts became increasingly common in addressing claims and disputes, ranging from asynchronous text-based processes to video and hybrid ones and from processes involving human facilitators and decision-makers to automated resolution avenues.³⁴ With the advancements in AI, it was only a matter of time before such functions would be incorporated into an online algorithm-driven process.³⁵

Another key development was the growth of the legal tech market, rooted in the spread of digital communication and the internet. These changes that took place towards the end of the twentieth century were seen as both an opportunity and a threat to legal (and other) professionals.³⁶ As access barriers to professional information were lifted with the growth of the internet and the advent of search engines, laypeople were exposed to knowledge that was previously available only through professional publications and required specialised expertise.³⁷ In addition to access to legal information, tools were later developed to assist in the generation of legal documents, such as contracts and wills, and in the filing of claims.³⁸ Similarly, AI-based tools for professional use by lawyers for contract drafting and review, due diligence, discovery, and more soon became available. Recently, law firms have turned to RAG systems, described earlier, hoping for enhanced performance and fewer errors and hallucinations.³⁹ The new capacities offered by these various AI tools have helped lawyers increase their efficiency and workload capacity, but they have also created pressures for reduced costs,⁴⁰ as well as strengthened the notion that, in certain cases, litigants could now manage some functions on their own.⁴¹ While some highlighted that in the wake of these developments lawyers would need to transform their practice,⁴² the profession has often resisted such calls on the grounds of ‘unauthorized practice of law’ related to the introduction of legal technologies.⁴³

³³ Orna Rabinovich-Einy, ‘The Past, Present, and Future of Online Dispute Resolution’ (2021) 74(1) *CLP* 125, 125–148.

³⁴ Orna Rabinovich-Einy, ‘Process Pluralism in the Post-Covid Dispute Resolution Landscape’ (2022) 10(1) *Texas A&M Law Rev* 55.

³⁵ See for example Amica, offering AI-supported negotiation for divorce cases in Australia: ‘Amica – The Simple, Low Cost, Smart Way to Separate or Divorce Online’ (*amica*) <<https://amica.gov.au/>> accessed 4 November 2024.

³⁶ Katsh (n 1) 224–226 (using the term ‘Trojan horse’ in this context). Orna Rabinovich-Einy and Ethan Katsh, ‘The New New Courts’ (2017) 67 *Amer U L Rev* 165, 213.

³⁷ Katsh, *Electronic Media and the Transformation of Law* (n 1) 224–226.

³⁸ Engstrom and Gelbach (n 28) (referring to LegalMation, RocketLawyer, CaseText, and Lawyaw in the Document Assembly and Creation category).

³⁹ There is supporting evidence that RAG outperforms GenAI. R. Lakatos, P. Pollner, A. Hajdu, and T. Joo, ‘Investigating the Performance of Retrieval-Augmented Generation and Fine-tuning for the Development of AI-driven Knowledge-based Systems’ (2025) 7 *Machine Learn & Knowl Extract* 15. However, RAG models have been also shown to suffer from errors and hallucinations. See V. Magesh, F. Surani, M. Dahl, M. Suzgun, C. D. Manning, and D. E. Ho, ‘Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools’ (2025) 22 *J Empir Legal Stud* 216–242 (finding Lexis +ai and westlaw ai to hallucinate at considerable rates).

⁴⁰ John O. McGinnis and Russell G. Pearce, ‘The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services’ (2014) 82 *Fordham L Rev* 3041 (discussing the effect of machine-aided legal services on making services cost-effective as well as putting pressure on experts to cut costs).

⁴¹ Amy J. Schmitz and John Zeleznikow, ‘Intelligent Legal Tech to Empower Self-Represented Litigants’ (2022) 23 *Colum Sci & Tech Law Rev* 142; J. J. Prescott, ‘Using ODR Platforms to Level the Playing Field’, in David Freeman Engstrom *Legal Tech and the Future of Civil Justice* (Cambridge University Press, 2023) 286–304.

⁴² Susskind (n 29) 136. See also Paul D. Callister, ‘Generative AI and Finding the Law’ (2025) 117 *L Libr J* 1.

⁴³ See for example FTC’s action and settlement in the matter of DoNotPay, a platform that offers, among other things, the filing of legal claims (‘DoNotPay’ (Federal Trade Commission) <www.ftc.gov/legal-library/browse/cases-proceedings/donotpay> accessed 4 November 2024).

Today, we find a myriad of AI-based legal tools and functions being developed for the public sector (e.g. courts and tribunals),⁴⁴ private settings (e.g. ADR service providers and law firms), and individual users – both professionals (e.g. lawyers, mediators, and arbitrators) and laypeople (typically those seeking redress or a defence).⁴⁵ Some of these tools have to do with the technical-procedural elements, such as case management,⁴⁶ or document translation and transcription.⁴⁷ Other functions have to do with document summaries or drafting capabilities.⁴⁸ Some tools offer legal research and analysis capabilities, such as jury evaluation and displaying relevant past decisions.⁴⁹ Certain AI tools help parties explore their interests and options, either in the form of a tool guiding them in assessing their legal issue and strategies for addressing it or as part of an ADR process;⁵⁰ other tools offer outcome predictions or recommendations to parties, lawyers, or third parties (e.g. judges, arbitrators, or mediators).⁵¹

Over the years, scholars and practitioners have offered various categorisations of AI usage in courts and ADR, distinguishing tools according to the scope of human involvement,⁵² the stage of the process into which AI is introduced,⁵³ the types of legal functions performed,⁵⁴ and what is at stake (simple versus complex, technical versus emotional).⁵⁵ Our focus is on the nature of the capabilities – those drawing on traditional AI-related strengths, such as efficiency and capacity, as opposed to those relating to ‘soft skills’, shaping the quality of communication with disputants or their procedural experiences. Such ‘soft skill’ capabilities (such as empathy and creativity) have been seen, until not long ago, as the sole purview of human communication. AI was initially seen as a means of streamlining and automating high volumes of cases in an efficient and accurate manner by handling mass, simple, and easily categorised disputes. We are now moving towards the handling of more complex, high-stake, sensitive, and emotional conflicts, with AI performing in ways that require more sophisticated language and emotional capabilities. This shift towards a more sophisticated use of AI in the legal domain raises a host of legal and normative questions: can machines satisfy the needs created in these more complex arenas? Are people willing to have algorithms resolve such disputes? Do the more advanced machine capabilities we are now seeing in LLMs satisfy human expectations in procedural terms? Will people feel heard by such machines, or that

⁴⁴ David Freeman Engstrom, *Legal Tech and the Future of Civil Justice* (Cambridge University Press, 2023); Francesco Contini (n 21); see also Chapter 21 by Vermeys and Tan in this book.

⁴⁵ For professionals see Darla Wynon Kite-Jackson, ‘2023 Artificial Intelligence (AI) TechReport’ (ABA, 15 January 2024) <www.americanbar.org/groups/law_practice/resources/tech-report/2023/2023-artificial-intelligence-ai-techreport/> accessed 4 November 2024; Contini, ‘Unboxing Generative AI’ (n 21); Katie Wolf, ‘Current Use of Large Language Models (LLMs) in Law’ (Filevine, 11 July 2024) <www.filevine.com/blog/current-use-of-large-language-models-llms-in-law/> accessed 4 November 2024. For laypeople see Tina Seabrooke et al., ‘A Survey of Lay People’s Willingness to Generate Legal Advice Using Large Language Models (LLMs)’, *TAS ‘24: Second International Symposium on Trustworthy Autonomous Systems* (ACM, 2024).

⁴⁶ Samuel Sahagian, ‘Lawyer’ (2021) 3 *UCLA J L & Tech Dig* 24.

⁴⁷ 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) 6 *Artif Intell Law* 1–38 and references there.

⁴⁸ *Ibid.*, 5–6 and references there.

⁴⁹ Sahagian (n 45), 15–16 for jury evaluation and 10–12 on legal research for displaying relevant past decisions.

⁵⁰ *Ibid.*, 20–24 on legal advice.

⁵¹ *Ibid.*, 14–15 on predicting case outcomes. See also Chapter 22 by Giacalone in this book.

⁵² Leah Wing and Chris Draper, ‘Parameters of Online Dispute Resolution: Introducing a New Framework for ODR’ (2022) 9(2) *Int’l J Online Disp Res* 112.

⁵³ Bommasani et al. (n 2).

⁵⁴ Sahagian (n 45).

⁵⁵ Jabotinsky and Lavi (n 23).

they have been treated with respect and dignity? Will the outcomes of such processes seem fair and acceptable?

In Section 4.3 we draw on existing empirical research on perceptions of fairness and legitimacy of AI in courts and ADR, generating insights on the implications and limitations of introducing a growing array of AI tools. As we will see, generally, the research shows that AI is preferred based on qualities such as accuracy, while humans are valued for their discretion and empathy.⁵⁶ Much of the research predates the widespread launch of advanced LLMs, leading to the questions whether resistance to the use of AI in certain domains will persist and whether the considerations that drive preference for AI over human decision-making will persevere.

4.3 SELECT RESEARCH FINDINGS ABOUT AI AND LEGITIMACY: AN EMPIRICAL PERSPECTIVE

In recent years, there has been a growing understanding that trust in and perceived legitimacy of AI are important measures of AI fairness and acceptability. Therefore, alongside the significance of the ‘objective’ evaluation of algorithms’ performance, a substantial number of studies have sought to capture how users of such systems (potential or actual, lay or professional) and the general public view the effectiveness, fairness, and desirability of AI use in various domains, including law and dispute resolution.⁵⁷ While many studies focus on using AI in various spheres of life, only a small portion relate to legal and dispute resolution settings.

Despite variations in research methods and in the types of perceptions they capture, and the often-vague descriptions of AI tools provided to study participants, studies examining the fairness and legitimacy of AI in the legal and dispute resolution arena offer important insights into the role AI can and should play in these domains. An important lesson from these studies is that while many of them uncover algorithmic aversion, meaning that people perceive human decision-making as preferable and more fair or legitimate than that of AI,⁵⁸ a nuanced analysis shows that under certain conditions AI may not only be as acceptable as human third parties,⁵⁹ but it may even be preferable to them (algorithm appreciation).⁶⁰

⁵⁶ Outside the legal context, see Yochanan E. Bigman and Kurt Gray, ‘People Are Averse to Machines Making Moral Decisions’ (2018) 181 *Cognition* 21 (finding overall machine aversion – i.e. preference towards a human agent – due to perceived lack of moral competence in machines, across decision domains).

⁵⁷ David Gefen, Elena Karahanna, and Detmar Straub, ‘Trust and TAM in Online Shopping: An Integrated Model’ (2003) 27(1) *MIS Quarterly* 51 (defining trust as a key predictor of intended use of technology in e-commerce settings); Ella Glikson and Anita Williams Woolley, ‘Human Trust in Artificial Intelligence: Review of Empirical Research’ (2020) 14(2) *Acad Manag Ann* 627 (meta-analysis on trust in AI across use-cases and domains); Christopher Starke et al., ‘Fairness Perceptions of Algorithmic Decision-Making: A Systematic Review of the Empirical Literature’ (2022) 9(2) *Big Data & Soc* 205395172211151 (meta-analysis on perceptions of algorithms’ fairness across decision-domains); Matthew Katsaros et al., ‘Procedural Justice and Self Governance on Twitter’ (2022) 1(3) *J Online Trust & Safety* (on the applicability of procedural justice to online realms); A. Morin-Martel, ‘Machine Learning in Bail Decisions and Judges’ Trustworthiness’ (2024) 39(4) *AI & Soc* 2033–2044.

⁵⁸ See for example A. J. Wang, ‘Procedural Justice and Risk-Assessment Algorithms’ (2018) SSRN *Electronic Journal* (where algorithms received the lowest score on preference, fairness and legitimacy compared with human decision-maker and guidelines-based risk assessment in bail hearings).

⁵⁹ Ryan P. Kennedy, Philip D. Waggoner, and Matthew Ward, ‘Trust in Public Policy Algorithms’ (2021) 84(2) *J Pol* 1132, 1132–1147; Ric Simmons, ‘Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System’ (2018) 52 *U C Davis Law Rev* 1067, 1067–1118.

⁶⁰ Theo Araujo et al., ‘In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence’ (2020) 35(3) *AI & Soc* 611, 611–623; Eike Schneiders et al., ‘Objection Overruled! Lay People Can Distinguish Large Language Models from Lawyers, but Still Favour Advice from an LLM’ (2024) Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–14.

While various studies have offered explanations for factors that give rise to algorithm appreciation (or aversion), the overarching conclusion emerging from the literature is that preferences for AI in the legal domain are highly contextual. Such explanations vary and include factors such as the domain in which AI is applied (e.g. criminal sentencing versus natural disaster planning),⁶¹ the decision's complexity (e.g. monetary versus mental issues in a divorce case),⁶² outcome control and decision-making authority (e.g. mediation versus arbitration, or having 'a human in the loop'),⁶³ or the stage of the legal process in which the algorithm is used (with AI being perceived more positively than human decision-making in earlier stages, a finding that is reversed in later stages).⁶⁴

A significant number of studies find that algorithms are preferred in the legal context for their consistency,⁶⁵ accuracy,⁶⁶ ability to perform complex calculations,⁶⁷ and impartiality.⁶⁸ One study found that AI's perceived accuracy was the main reason for algorithmic appreciation (and not human-related factors, such as 'voice', the opportunity to tell one's side of the story).⁶⁹ Another study found that the vast majority (87.7 per cent) of respondents (including those with a human preference) thought that introducing machine learning would enhance the determinacy of legal outcomes, contributing to greater certainty in judicial decisions.⁷⁰

Some studies find that humans (as opposed to AI) are appreciated for their 'soft skills', such as empathy and providing an opportunity for voice,⁷¹ and that the preference for human versus AI agents is grounded in ethical considerations, such as warmth, competence,⁷² and voice.⁷³ One

⁶¹ Michael C. Horowitz and Lauren Kahn, 'What Influences Attitudes about Artificial Intelligence Adoption: Evidence from U.S. Local Officials' (2021) 16(10) *PLOS ONE* e0257732.

⁶² Gizem Yalcin et al., 'Perceptions of Justice by Algorithms' (2023) 31 *AI & L* 269.

⁶³ Anthony Joseph Casey and Anthony Niblett, 'Will Robot Judges Change Litigation and Settlement Outcomes? A First Look at the Algorithmic Replication of Prior Cases' (2020) *SSRN Electron J* <<http://doi.org/10.2139/ssrn.3633037>> accessed 5 November 2024; Eidenmueller, McLaughlin, and Eidenmueller, 'Expanding the Shadow of the Law' (n 24) 31–32 (pointing to overall preference for online systems among dispute resolution professionals, while rejecting AI for decision-making: 'The following three proposals received the strongest support' (mean value): 'online platform for communication with the court and other parties' (3.94), 'online filing and case management' (3.89), and 'greater specialisation / subject matter expertise of judges' (3.84)). In contrast, respondents (weakly) oppose the 'use of artificial intelligence to assist case handling and decision-making' (2.18); Ayelet Sela, 'Can Computers Be Fair? How Automated and Human-Powered Online Dispute Resolution Affect Procedural Justice in Mediation and Arbitration' (2018) 33 *Ohio St J on Disp Resol* 91.

⁶⁴ Dovilė Barysė and Roeë Sarel, 'Algorithms in the Court: Does It Matter Which Part of the Judicial Decision-Making Is Automated?' (2024) 32(1) *AI & L* 117; Andreia Martinho, 'Surveying Judges about Artificial Intelligence: Profession, Judicial Adjudication, and Legal Principles' (2024) 40 *AI & Society* 569–584.

⁶⁵ Benjamin Minhao Chen and Zhiyu Li, 'How Will Technology Change the Face of Chinese Justice?' (2020) 34 *Col J Asian L* 1.

⁶⁶ Wang (n 57).

⁶⁷ Glikson and Williams Woolley (n 56) 627–660.

⁶⁸ Minhao Chen and Li (n 64).

⁶⁹ Benjamin Minhao Chen, Alexander Stremitzer, and Kevin Tobia, 'Having Your Day in Robot Court' (2022) 36 *Harvard J L & Tech* 127 (finding in mediation analysis that reduction in fairness ratings that came from having an algorithm rather than a human is most significantly attributed to accuracy (29 per cent) vs. only 2 per cent coming from voice).

⁷⁰ Minhao Chen and Li (n 64).

⁷¹ Eiichiro Watamura et al., 'Empathetic Robot Judge, We Trust You' (2023) 40(18) *Int'l J Hum-Comput Int'n* 1, 1–10.

⁷² Xusen Cheng et al., 'Human vs. AI: Understanding the Impact of Anthropomorphism on Consumer Response to Chatbots from the Perspective of Trust and Relationship Norms' (2022) 59(3) *Inf Process & Manage* 102940 (examining the impact of anthropomorphism on consumer trust in chatbots in e-commerce – authors find that when consumers consider chatbots as having higher warmth or higher competence, they tend to trust them more).

⁷³ R. Binns, M. van Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'It's Reducing a Human Being to a Percentage": Perceptions of Justice in Algorithmic Decisions', in R. Mandryk and M. Hancock (eds), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (The Association for Computing Machinery, 2018) 1–14.

way of explaining such findings is that humans are perceived as superior to AI in certain ‘emotional’ qualities that are deemed essential to decision-making, such as reasoning and empathy.

One study found that algorithmic aversion was anchored in concerns over the lack of deliberation, voice, and procedural fairness.⁷⁴ Another found perceived empathy to be positively associated with trust in a judge (both human and AI), which in turn predicts positive judgment evaluation and overall acceptance of the judgment.⁷⁵ That study further found that a human judge is perceived as significantly more empathetic than an ‘AI judge’. Much in the same vein, a study researching perceptions of human versus ‘robot’ lawyers found that the latter were perceived as not being able to fulfil tasks requiring empathy or operate with creativity, nor being able to engage in psychological warfare.⁷⁶ Alternatively, the study showed that ‘robot lawyers’ were perceived as having the ability to collect and retrieve data and to analyse and predict cases.

As we can see, most of the features driving positive perceptions towards AI in these studies are the types that one would typically associate with AI strengths, such as accuracy and predictability. However, with advancements in LLM capabilities we may find that some of the distinctions between human and AI fortes are becoming less stark.

A recent study demonstrated LLMs’ ability to mediate between people with divergent opinions on divisive issues and generate a consensus.⁷⁷ In that study, AI-generated statements reflecting common ground on divisive issues among group members were consistently preferred by participants over those prepared by trained humans and were deemed by external experts as being comparatively more clear, informative, and fair. Most striking was the finding that the machine not only produced consensual statements that were preferred by group members over those prepared by humans, but it was also successful in moving participants’ views closer while showcasing minority opinions and providing them with the space in which they could sway others in their direction.

Along similar lines, a study comparing LLMs and human mediators found that the LLM outperformed human mediators. Specifically, the LLM’s choice of intervention strategies was rated equal or superior to human-selected interventions in 62 per cent of cases. Moreover, in 84 per cent of cases the messages crafted by the LLM were judged to be better than, or at least on par with, those written by human mediators (evaluation was guided by rubric items including understanding and contextualisation, neutrality and impartiality, empathy awareness, and resolution quality).⁷⁸ While the evaluators in the study were not expert mediators, and therefore relied on comprehension and tone assessment rather than professional criteria, the findings are striking. Another study explored the use of LLMs in the context of ODR, employing an experimental platform (LLMediator). Researchers demonstrated certain facilitative capabilities, including the LLMediator’s ability to reformulate inflammatory messages to a more neutral tone, assist mediators by providing suggested interventions, and autonomously intervene in the negotiation process with the aim of guiding the parties towards an amicable resolution.⁷⁹

⁷⁴ Anna Fine, Stephanie Le, and Monica K. Miller, ‘Content Analysis of Judges’ Sentiments Toward Artificial Intelligence Risk Assessment Tools’ (2023) 24(2) *J Criminol Crim Just Law & Soc* 31, 31–46.

⁷⁵ Watamura et al. (n 70); Sela (n 62).

⁷⁶ Ni Xu and Kung-Jeng Wang, ‘Adopting Robot Lawyer? The Extending Artificial Intelligence Robot Lawyer Technology Acceptance Model for Legal Industry by an Exploratory Study’ (2019) 27(5) *J Manag Organ* 867, 867–885.

⁷⁷ Michael Henry Tessler et al., ‘AI Can Help Humans Find Common Ground in Democratic Deliberation’ (2024) 386 *Science* 6719.

⁷⁸ Jinzhe (n 13).

⁷⁹ Westermann (n 13)

Clearly, with the advancement in LLM capabilities and the increased creativity they have garnered, these tools seem to be much more capable of building rapport and conveying empathy and support. Indeed, it may very well be that LLMs' hallucinations are less of a 'bug' and more of a 'feature', a cost one needs to pay if they are to enjoy these other extraordinary abilities.⁸⁰ Aside from the expected technological changes in upcoming years (model upgrades and technological advancements), habits and social norms surrounding AI services are expected to transform.⁸¹ As we study this moving target, it remains to be seen how such changes will shape perceptions of human and AI capacities in the legal domain, as well as the underlying reasons driving people's preferences for human versus automated legal decision-making.

4.4 CONCLUDING THOUGHTS

With the dramatic leap in LLM capabilities, we are witnessing legal AI functions improve and expand from the technological automation domain to new frontiers. Research findings, perhaps due to the inherent gap between the pace of technology and the research publication process, demonstrate an overall algorithm aversion for certain legal functions, but an appreciation for algorithmic performance in particular contexts, frequently where traditional algorithmic qualities and strengths are most pronounced.

Will these findings persist despite the profound changes we are witnessing in the capabilities of LLMs? How will perceptions of AI shift now that AI capabilities in the legal arena expand beyond more traditional functions to include enhanced communication skills and emotional intelligence-like features that have long been considered hallmarks of the human skill set? Finally, even if algorithms *can* perform as well as humans in what have traditionally been areas of human strength, *should* we, as a society, fully delegate such functions to AI from a moral standpoint? The concern underlying this question has less to do with the outcomes that will be reached through such AI-led processes and more to do with what it means for us as individuals and as a society to undergo legal and dispute resolution processes without leaving some space in which we interact with human 'third parties'.

⁸⁰ Minhyeok Lee, 'A Mathematical Investigation of Hallucination and Creativity in GPT Models' (2023) 11(10) *Mathematics* 2320 <<http://doi.org/10.3390/math11102320>> accessed 5 November 2024; Tanmoy Chakraborty and Sarah Masud, 'The Promethean Dilemma of AI at the Intersection of Hallucination and Creativity' (Communications of the ACM, 2024) <<http://doi.org/10.1145/3652102>> accessed 5 November 2024.

⁸¹ Schneiders et al. (n 59) (the authors suggest that algorithmic aversion is shaped by social norms, or the expectation to trust human more than AI. The authors' conclusion results from an interesting pattern in findings in a series of experiments on willingness to accept legal advice depending on its source: when the source of the advice was unknown to participants, they were more willing to act on the LLM-generated advice than the lawyer-generated advice. However, when the source of advice was known – i.e. respondents were explicitly told it was a human lawyer – higher reliance on humans was observed).