NYU | LAW

Journal of Intellectual Property & Entertainment Law

VOLUME 15 NUMBER 2

Statement of Purpose

Consistent with its unique development, the New York University Journal of Intellectual Property & Entertainment Law (JIPEL) is a nonpartisan periodical specializing in the analysis of timely and cutting-edge topics in the world of intellectual property and entertainment law. As NYU's first online-only journal, JIPEL also provides an opportunity for discourse through comments from all of its readers. There are no subscriptions or subscription fees; in keeping with the open-access and free discourse goals of the students responsible for JIPEL's existence, the content is available for free to anyone interested in intellectual property and entertainment law.

The New York University Journal of Intellectual Property & Entertainment Law is published up to three times per year at the New York University School of Law, 139 MacDougal Street, New York, New York, 10012. In keeping with the Journal's open access and free discourse goals, subscriptions are free of charge and can be accessed via www.jipel.law.nyu.edu. Inquiries may be made via telephone (212-998-6101) or e-mail (submissions.jipel@gmail.com).

The Journal invites authors to submit pieces for publication consideration. Footnotes and citations should follow the rules set forth in the latest edition of The Bluebook: A Uniform System of Citation. All pieces submitted become the property of the Journal. We review submissions through Scholastica (scholasticahq.com) and through e-mail (submissions.jipel@gmail.com).

The Journal is also available on WESTLAW, LEXIS-NEXIS and HeinOnline.

NEW YORK UNIVERSITY

JOURNAL OF INTELLECTUAL PROPERTY AND ENTERTAINMENT LAW

VOLUME 15 SPRING 2026 NUMBER 2

404 JUSTICE NOT FOUND: ON THE STRUCTURAL INCOMPATIBILITIES BETWEEN AI AND JUDICIAL DECISION-MAKING

ELISABETH PAAR*

It sounds like a judge. It presents the facts of a case like a judge. It summarizes the evidence like a judge. It states the law like a judge. It offers a legal opinion like a judge. Artificial Intelligence seems to tick all the boxes when it comes to judicial decision-making. And yet: Are we persuaded by AI? Can we rely on AI? Should we let AI judge us? Some scholars think so.

This article challenges the notion that AI systems can effectively assume the role of judges. Despite AI's impressive advancements, particularly in natural language processing, this article illustrates that the essence of judicial decision-making extends far beyond the mere production of legal texts. The judicial function encompasses complex cognitive processes, evaluative skills, and a nuanced understanding of societal contexts as well as legal knowledge that AI systems fundamentally lack.

Drawing on prevailing normative frameworks governing the role of the judge, this article demonstrates that AI falls short when it comes to all core capabilities required to fulfill the judicial task: social skills for fact-finding, legal skills for assessing the established facts, and the ability to provide reasoned justifications. Contrary to a popular narrative in scholarship, these failings are not merely quantitative but rather display qualitative differences between human and machine intelligence. This, in turn, results in insurmountable structural incompatibilities between AI and the judicial decision-making process.

* Postdoctoral Researcher, Department of Public Law and Political Science, Faculty of Law, University of Graz. The author thanks Gilad Abiri, Jack Balkin, Christoph Bezemek, Rafael Bezerra Nunes, Robert Post, Tom Tyler, and Gideon Yaffe for helpful discussions and valuable feedback during the development of this Article.

The article concludes that continuously tweaking AI in hope that it will eventually become "good enough" to overcome these structural incompatibilities merely sets such efforts up for failure. Instead, any attempt to recalibrate legal systems to accommodate AI judges would require nothing short of a radical reconceptualization of the judiciary's role in democratic societies, and thus of law itself.

Introduction

In 2019, Eugene Volokh posed the question, "What would happen if an AI program could write legal briefs and judicial opinions?[1] Only a few years later, it seems we are closer than ever to finding out. Given the triumphant march of Large Language Models (LLMs), we are increasingly confronted with AI-generated text reading as though it were written by legal experts.[2] ChatGPT made headlines by passing bar exams and producing legal opinions indistinguishable from those drafted by experienced judges.[3] This rapid technological advancement has sown seeds of uncertainty throughout the legal community and beyond.[4]

How should we respond to these revolutionary developments? Is it time to pass the judicial torch on to AI?[5] Are we really, as Volokh suggests, equally

---

[1] Eugene Volokh, *Chief Justice Robots*, 68 Duke L.J. 1135, 1137 (2019).

[2] Andrew Perlman, *The Implications of ChatGPT for Legal Services and Society*, 30 Mich. Tech. L. Rev., 2 (2024).

[3] Daniel Martin Katz et al., *GPT-4 Passes the Bar Exam* (2023), https://papers.ssrn.com/abstract=4389233 [https://doi.org/10.1098/rsta.2023.0254]; Rupert Macey-Dare, *ChatGPT & Generative AI Systems as Quasi-Expert Legal Advice Lawyers - Case Study Considering Potential Appeal Against Conviction of Tom Hayes* (2023), https://papers.ssrn.com/abstract=4342686 [http://dx.doi.org/10.2139/ssrn.4342686]; Molly Bohannon, *Lawyer Used ChatGPT In Court—And Cited Fake Cases: A Judge Is Considering Sanctions*, Forbes (June 8, 2023), https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/ [https://perma.cc/9PB2-RU4X]; Pallavi Pundir, *This Court Used ChatGPT to Decide Bail in a Murder Case*, Vice (Mar. 29, 2023), https://www.vice.com/en/article/india-court-chatgpt-bail-murder-case/ [https://perma.cc/C9WP-EZ9S]; Luke Taylor, *Colombian Judge Says He Used ChatGPT in Ruling*, The Guardian (Feb. 3, 2023), https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling [https://perma.cc/5EP4-FY5W].

[4] Even Chief Justice Roberts himself came to the defense of human judges, stating "AI Won't Replace Human Judges." *See* Amy Howe, *AI Won't Displace Human Judges, but Will Affect Judiciary, Roberts Says in Annual Report*, SCOTUSBlog (Dec. 31, 2023), https://www.scotusblog.com/2023/12/ai-wont-displace-human-judges-but-will-affect-judiciary-roberts-says-in-annual-report/ [https://perma.cc/YGT2-LTEZ]. *See generally* John G. Roberts, Jr., *2023 Year-End Report on the Federal Judiciary* (2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf [https://perma.cc/C3BV-KFSV].

[5] This question has been at the center of numerous scholarly pieces, particularly in recent years. *See, e.g.*, Benjamin Minhao Chen, Alexander Stremitzer & Kevin Tobia, *Having Your Day in Robot Court*, 36 Harv. J.L. & Tech. 127 (2022); Tania Sourdin, *Judge v. Robot?: Artificial Intelligence and Judicial Decision-Making*, 41 Unsw L.J. 1114 (2018) [hereinafter Sourdin, *Judge v Robot?*]; Tania Sourdin, *Robo Justice: Constitutional Issues with Judge AI*, 30 Ind. J. Global Legal Stud. 293 (2023); John Morison & Adam Harkens, *Re-Engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making*, 39 Legal Stud. 618 (2019); Anthony J. Casey & Anthony Niblett, *Will Robot Judges Change Litigation and Settlement Outcomes? A First Look at the Algorithmic Replication of Prior Cases*, MIT Comput. L. Rep. (2020), https://law.mit.edu/pub/willrobotjudgeschangelitigationandsettlementoutc

persuaded by AI-drafted legal opinions, merely because we increasingly fail to distinguish them from the reasoning of human judges? And if so, should we accept AI judges as no less reliable than their human counterparts?[6]

The article argues that drawing these conclusions solely on the basis of AI[7] output grossly misrepresents the role of the judge and their decision-making process in a democracy. A text generated by AI may sound like a legal opinion by a human judge. Even legal experts may have trouble telling them apart. This superficial resemblance, however, is neither proof nor a suitable proxy for determining whether AI is capable of taking over judicial tasks, let alone the function of a judge as a whole.[8]

The essence of judicial reasoning lies not in its presentational form but in the underlying cognitive and evaluative processes of the judge. What matters for a text to qualify as a legal opinion is not only *what* judges produce, but also *how* they produce it—namely, the method of their decision-making—and, perhaps even

omes/release/1 [https://perma.cc/SQJ9-UNFX]; Simon Chesterman, Lyria Bennett Moses & Ugo Pagallo, *All Rise for the Honourable Robot Judge? Using Artificial Intelligence to Regulate AI: A Debate*, 2023 Tech. & Regul. 5 (2023).

[6] Volokh, *supra* note 1, at 1135.

[7] AI, short for Artificial Intelligence, is an elusive term lacking a universally recognized definition. The origin of AI is usually considered to be a conference in 1956 in preparation for which John McCarthy and his collogues defined AI as "making a machine behave in ways that would be called intelligent if a human were so behaving", thereby establishing the conceptual link between AI and human intelligence prevailing up to this day. *See* John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 27 AI Mag. 12 (2006). Many decades later, John McCarthy published an updated version of what AI is, describing it as "the science and engineering of making intelligent machines, especially intelligent computer programs". He therein relativized the connection to human intelligence, stating that AI "is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable" and that AI is only "sometimes but not always or even usually" about simulating human intelligence. Nevertheless, the definitions surrounding AI remain vague, particularly due to the reference to the term "intelligence" which also lacks a universally acknowledged definition, regardless of whether one refers to human or artificial intelligence. McCarthy himself—though defining intelligence as "the computational part of the ability to achieve goals in the world" —admits that it is not yet possibly to "characterize in general what kinds of computational procedures we want to call intelligent". *See* John McCarthy, *What Is Artificial Intelligence?*, 2–3 (2007), http://jmc.stanford.edu/articles/whatisai.html [https://perma.cc/XR65-N6BR].

[8] Similarly, though against the backdrop of much less advanced versions of AI, Cass R. Sunstein argues in *Of Artificial Intelligence and Legal Reasoning*, U. Chi. Pub. L. & Legal Theory Working Paper No. 18, at 4–5 (2001), that one should reject the idea that AI can reason legally as any claim like that is "based on an inadequate account of what legal reasoning is", and "misses a central point about analogical reasoning: its inevitably evaluative, value-driven character."

more importantly, their understanding of *why* they do what they do, all of which are constitutive characteristics of legal opinions.[9]

This inward-looking,[10] substantive concept of judicial decision-making, I argue, is not contingent upon the theoretical position one adopts. It is also not merely one of many ways in which certain legal systems construct judicial decision-making, while others opt for alternative models. Rather, this understanding of the judicial decision-making process results directly from how judges are currently conceptualized in virtually any democracy founded on the rule of law.[11]

The common baseline of this concept of judges is the task assigned to them: to administer justice in the name of the state. Judges fulfill this task, conceptually speaking, by resolving legal disputes through a predetermined procedural process. Legal disputes are, in essence, everyday social occurrences between humans which are subject to legal norms. The decisions of judges are grounded in these legal norms, applied to the established facts of the case.[12] All judicial rights and duties exist solely to enable judges to best fulfill their role as agents of the state, inextricably linking them to the task assigned to them within the overall structure of a democratic society.[13]

Who may serve as a judge, therefore, depends on who can cumulatively fulfill the capabilities required to perform the task of judicial decision-making. The

[9] This can ultimately be traced back to the distinction of Wittgenstein between "following a rule" and merely "behaving in accordance with a rule." *See* Ludwig Wittgenstein, Philosophical Investigations 138–242 (1973). For a contextualization of Wittgenstein's approach through the lens of AI and its attempts at legal reasoning, see Ian R. Kerr & Carissima Mathen, *Chief Justice John Roberts Is a Robot*, 23–26 (Apr. 1, 2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3395885 [https://perma.cc/9N7J-6TNX].

[10] Though not fully congruent, this has a certain overlap with Hart's distinction between adopting an external vs an internal point of view. *See generally* H.L.A. Hart, The Concept of Law 87 (1961).

[11] Hildebrandt refers to this approach as investigating "law as constitutive of societal defaults" and "law-as-we-know-it". *See* Mireille Hildebrandt, *Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics*, 68 U. Toronto L.J. 12, 15 (2018). Hildebrandt herself builds her assessment of the conceptual foundations of the law Bruno Latour's mode of existence of the law. *Id.* at 15. *See generally* Aharon Barak, The Judge in a Democracy (2006) (discussing how judges are conceptualized in democracies).

[12] This is one of the core features of the rule of law. *See, e.g.*, David Boies, *Judicial Independence and the Rule of Law*, 22 Wash. U. J.L. & Pol'y 57 (2006); Frederick Schauer, *Rules and the Rule of Law*, 14 Harv. J.L. & Pub. Pol'y 645 (1991).

[13] Edzard Schmidt-Jortzig, *Aufgabe, Stellung und Funktion des Richters im Demokratischen Rechtsstaat* [*Role, Position, and Function of the Judge in the Democratic Constitutional State*], Neue Juristische Wochenschrift [NJW] 2377, 2377–2378 (1991) (Ger.).

appointment of an AI system as a judge may thus not be categorically excluded.[14] However, as is the case for a human serving as a judge, it presupposes that AI can cumulatively fulfill all the required judicial capabilities. This paper argues that even the most advanced AI tools are currently nowhere near doing so. Moreover, there is ample reason to believe that AI has inherent, immutable limitations and will most likely never overcome the structural incongruence between our current concept of judicial decision-making and AI's operational principles.

So, what are the required judicial capabilities? The core judicial capabilities, as well as their breadth and complex nature, are best illustrated by considering scenarios judges may be confronted with. Consider, for example, the following fictitious case brought before a judge: Small-town residents file a class action lawsuit against a chemical manufacturing company, claiming that groundwater contamination has led to increased cancer rates in their community. The case involves complex scientific evidence concerning chemical compounds, conflicting epidemiological studies, and disputes over the geographical scope of contamination. Both national environmental groups and industry associations seek to intervene in the case.

Building on the different aspects of judicial challenges illustrated through this fictitious case, this paper finds that, broadly speaking, the required judicial capabilities can be divided into three categories:[15]

1. Social skills required for fact-finding

[14] Contrary to claims sometimes made in legal scholarship (*e.g.,* Volokh, *supra* note 1, at 1158 (arguing that "Art III of the US Constitution is best understood as contemplating human judges, and likewise for similar state constitutional provisions")), any value ascribed to a human serving as a judge does not stem from a constitutional reservation of judicial office for humans simply because of being humans. Constitutional legislators, of course, implicitly assumed human judges. They did so, however, merely due to the absence of viable alternatives at the time of enactment. This assumption, hence, should not be equated with an entity-based human reservation for the judicial function—that is, a constitutional human reservation merely for the sake of the judge being human. *See, e.g.*, Kerr and Mathen, *supra* note 9, at 13 ("It is important to bear in mind the difference between the question of JR-R's constitutional eligibility to be a judge, and the question of JR-R's qualifications—assessed on its own merits or relative to other candidates—to perform judicial tasks.").

[15] In addition to these three categories of judicial capabilities, one may consider a strong judicial character rooted in the principle of judicial independence as the fourth category. However, since the framing of this article centers the judicial decision-making process as such and not whether and how judges may be (negatively) impacted by external pressures of all sorts, the core concerns of judicial independence are considered as interwoven with the other three categories.

2. Legal skills required for ascertaining legal norms and applying them to the facts of the case

3. The ability to justify the outcome of the decision-making process by providing a judicial opinion

The importance of all three core capabilities is evident across cases in distinct ways. In environmental class action proceedings like the one outlined, social skills help, for instance, to manage emotional testimony from affected families while maintaining professional discourse among competing scientific experts. With regard to legal expertise, this type of cases showcases the need for expertise in both environmental regulations and class action requirements. When it comes to drafting judicial opinions, the fictitious case illustrates the requirement of judicial reasoning capabilities to offer an explanation of witness credibility determinations and the weighting of expert testimony. Furthermore, it demands clear articulation of scientific causation standards for lay audiences.

Parts I-III of this article delve into the substance of each of these three capability categories and carve out their distinct features in a more holistic manner. On the basis of the findings of this assessment, I then analyze the extent to which AI meets the capabilities necessary for judicial decision-making. Part I, which examines the social skills required by a judge, concludes that the capacity for comprehensive human communication and a model of the real world are necessary for a judge to fulfill the task of legal fact-finding.[16] While AI demonstrates potential in isolated aspects of judicial fact-finding, fundamental limitations persist. This is especially the case in areas requiring deep language understanding, emotional intelligence, and holistic reasoning.[17] The inability to fully grasp communicated

[16] *See, e.g.*, Robert C. Post, *The Internet, Democracy, and Misinformation*, *in* Disinformation, Misinformation, and Democracy: Legal Approaches in Comparative Context 37, 47 (Ronald J. Krotoszynski, Jr., András Koltay & Charlotte Garden eds., 2025) (stressing that it is "necessary to understand facts about the world to know whether a statement is true or false" and that "AI is a poor instrument for determining in the first instance whether statements are true or false; we require human judgment, typically exercised in the form of law").

[17] Whenever this article refers to deep understanding or true language comprehension, it is based on the way in which Gadamer conceives of understanding, and interpretation, namely "a mode of insight that has its own rationality irreducible to any simple rule or set of rules, that cannot be directly taught, that necessarily implicates an understanding of self as well as other, and that is always oriented to the particular case at hand". Gadamer's idea of phronesis ("practical wisdom") is, in turn, influence by Heidegger who considered the concept of phronesis central, amongst others, when it comes to "constituting a mode of insight into

content or to interpret nonverbal cues precludes AI from wholly supplanting human judges in this domain.[18]

Part II, focusing on the legal skills necessary for judicial decision-making, shows that the judicial process of legal assessment is far from being a simple algorithmic procedure.[19] Rather, it is a complex, multi-faceted endeavor that resists comprehensive replication by AI systems. Even in cases where AI systems demonstrate high accuracy rates in predicting judicial outcomes, this superficial success belies a deeper failure to grasp the underlying legal norms and methodological requirements that guide judicial legal assessment.[20]

Part III assesses judicial reasoning as showcased in the opinions of the court. It refers to a continuous form of argumentation which is committed to the rationality standards of the legal community. Despite often being considered a black box, AI can occasionally elucidate its decision-making process, particularly when making use of various explainable AI (XAI) techniques.[21] Nevertheless, the comprehensive

---

our own concrete situation (both our practical situation and, more fundamentally, our existential situation, hence phronesis constitutes a mode of self-knowledge)." *See* Jeff Malpas, *Hans-Georg Gadamer*, *in* The Stanford Encyclopedia of Philosophy (Edward N. Zalta & Uri Nodelman eds., Winter 2022 ed.), https://plato.stanford.edu/archives/win2022/entries/gadamer/ [https://perma.cc/PEM9-TVRW].

[18] *See, e.g.*, Rodrigo González, *Classical AI Linguistic Understanding and the Insoluble Cartesian Problem*, 35 A.I. & Soc'y 441 (2020) (pointing to fundamental limitations of AI in this regard: "knowledge and awareness of meaning require a first-person viewpoint which is irreducible to the decomposition of algorithmic mechanisms"). One may in this regard also refer to Thomas Nagel who famously argued that "to know what it is like to be a bat, you have to be a bat." *See* Thomas Nagel, *What Is It Like to Be a Bat?*, 83 Phil. Rev. 435 (1974). Similarly, one may argue, that to know what it is like to be human requires to be a human.

[19] Richard A. Posner, How Judges Think 5 (2008) ("The judicial mentality would be of little interest if judges did nothing more than apply clear rules of law [. . .] to facts that judges and juries determined without bias or preconceptions.").

[20] Kerr and Mathen, *supra* note 9, at 23 ("[T]he very activity of judging requires following rules."), 25 (stating that AI cannot be "a true participant in [. . .] rule-following" because it lacks a normative commitment to the rules, or to the lived experience that makes the rules significant). *See also id.* at 23 ("AI might predict with perfect frequency the correct outcome in a mathematical series or respond correctly every time when generating legal outcomes—and yet the machine may only be acting in accordance with the rule rather than following it."), 25 ("[AI] may learn how to behave in accordance with the rules of the legal system. But . . . [it] certainly did not share the same social or environmental history as any other participant in the entirety of customs, usages and social practices that constitute our legal system.").

[21] Rudresh Dwivedi et al., *Explainable AI (XAI): Core Ideas, Techniques, and Solutions*, 55 ACM Comput. Surv. 194:1 (2023); Jan Zacharias et al., *Designing a Feature Selection Method Based on Explainable Artificial Intelligence*, 32 Elec. Mkt. 2159 (2022); Muhammad Suffian et al., *FCE: Feedback Based Counterfactual Explanations for Explainable AI*, 10 IEEE Access 72363 (2022).

transfer of judicial reasoning to AI systems is currently infeasible. AI, particularly LLMs, may generate text that superficially resembles judicial reasoning. However, it fundamentally lacks the cognitive capabilities required for judicial reasoning in any substantively valuable sense.[22]

With regard to all three categories of the required judicial capabilities, the analysis shows that our current conception of judicial decision-making is inextricably linked to human cognitive processes and societal structures.[23] AI, in contrast, operates on fundamentally different principles than humans, having its very own distinct architecture.[24] Thus, imposing a human-centric[25] framework onto AI systems is simply not feasible. The limitations of AI are not merely quantitative—that is, a matter of processing power or data volume—but qualitative,[26] leading to what we might term insurmountable structural incompatibilities with the judicial process.[27] The great mismatch between judicial decision-making on the one hand and AI-based decision-making on the other requires a conceptual shift by scholars. The object of investigation should not

[22] Henrik Palmer Olsen, Jacob Livingston Slosser & Thomas Troels Hildebrandt, *What's in the Box?: The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration*, *in* Constitutional Challenges in the Algorithmic Society 219, 226 (Amnon Reichman et al. eds., 2021).

[23] Post, *supra* note 16, at 47–48 (citing Ronald Beiner, Political Judgement (1983)) ("Judgment, including legal judgment, depends upon our common participation in a shared community[, and as] we exercise judgment, we participate in and shape the nature of that community."). For an assessment through the lens of legal realism, see Jerome Frank, *Are Judges Human? Part 2: As Through a Class Darkly*, 80 U. Pa. L. Rev. 233 (1931); Jerome Frank, *Are Judges Human? Part 1: The Effect on Legal Thinking of the Assumption That Judges Behave Like Human Beings*, 80 U. Pa. L. Rev. 17 (1931).

[24] AI "thinks" in ways that are fundamentally alien to human legal reasoning. One may even go as far as saying that AI does not think at all. *See, e.g.*, Rebecca Crootof, Margot Kaminski & W. Price, *Humans in the Loop*, 76 Vand. L. Rev. 429, 465 (2023). *See also*, Kerr and Mathen, *supra* note 9, at 29 (regarding the question whether AI can adopt an internal point of view of the law, never ever having been a citizen of anywhere, let alone a human being: "AI operates in the realm of the empirical, not the normative", thereby "robots miss out a whole dimension of the social life.").

[25] Ian Kerr and Carissima Mathen stress that the "primary skill involved in judging—legal reasoning—is a human endeavour that benefits from acquired technical skills over time." Kerr & Mathen, *supra* note 9, at 39.

[26] Brent Mittelstadt, Sandra Wachter & Chris Russell, *To Protect Science, We Must Use LLMs as Zero-Shot Translators*, 7 Nature Human Behaviour 1830, 1831 (2023) ("The concept of 'truth' has been highly simplified in LLM development and equated with accuracy measured against the 'ground truth' of the training data.").

[27] Post, *supra* note 16, at 48 ("AI cannot be a member of any human community. It cannot participate in, and hence construct a dialectical relationship with, any human community. AI therefore cannot pronounce law. At most AI can report factual determinations about the way that actual humans regard law.").

be when AI will attain the competencies of judicial decision-making but instead, whether AI is meaningfully capable of judicial decision-making in the first place.

The article concludes in Part IV by demonstrating the importance of these findings beyond the status quo. It does so by showcasing the risks of decentering the current human-centric concept of judicial decision-making and shift our mechanistic benchmarks of AI. While AI may have immutable limitations, legal frameworks in democratic societies are malleable. Reconceptualizing the judicial role without a deep understanding of the capabilities and limitations of AI presents fundamental challenges to our basic conception of law. Simultaneously, a deep understanding of the prevailing judicial mode within democracies is essential, not least to enable informed comparisons in the future.

I
Social Abilities for Establishing the Facts

Judges decide legal disputes and justify those decisions by giving reasons. Doing so requires two distinct yet interrelated processes, namely fact-finding and legal assessment of established facts.[28] Each function requires distinct competencies, including social skills for fact-finding and legal skills for the subsequent ascertainment and application of legal norms.[29] This Part focuses on the social skills necessary for fact-finding.

A. The Process of Establishing the Facts of a Case: Taking and Assessing Evidence

The fact-finding process demands what colloquially may be termed "interpersonal skills" that enable judges to interact with and assess human behavior. Such interpersonal skills include not only cognitive and sensorimotor aptitudes but also emotional and social intelligence. While social skills are an often-underappreciated competency, a judge's ability to mediate tense interactions or manage diverse courtroom participants fundamentally shapes the administration of justice. Consider, for instance, a family law proceeding in which emotional

[28] This article focuses on the decision-making process of a trial judge. Higher courts, by contrast, typically review the factual findings and methods of the trial judge.

[29] Naturally, the processes of fact-finding and legal assessment in judicial decision-making are frequently intertwined in practice and separated for analytical purposes. *See infra* Section II.A. The oscillation of these functions necessitates consistent application of the respective skillsets throughout the entire process of judging a case. Both social and legal skills are thus integral to the judicial profile.

intelligence can transform a potentially adversarial encounter into a constructive dialogue that centers the best interests of children.[30]

To further elucidate these social skills, it is beneficial to distill the fact-finding process into two stages: (1) the taking of evidence, and (2) the assessment of evidence and subsequent determination of facts.

Taking evidence involves direct interaction between the judge and various individuals, along with the objects relevant to the case. The specific skills employed may vary depending on the type of material presented. In the case of testimonial evidence, proficiency in verbal and non-verbal communication is required. Documentary and demonstrative evidence, by contrast, necessitates the direct sensory observation of situations, objects, or persons. Throughout this stage, the judge should maintain an active, guiding role by engaging with parties and overseeing proceedings in a holistic fashion.[31]

A judge must then evaluate the soundness of the evidence taken and determine the facts available for legal assessment.[32] This intellectual exercise is both pragmatic and subjective. Insofar as it relies on the personal experience of a judge, it bears an indirect relationship to legal standards. Scholarship investigating the application of interpersonal skills often vaguely directs judges to assess evidence based on life experience or their personal understanding of human nature without providing a definitive instructive scheme.[33]

---

[30] *See, e.g.*, Carlos Manuel Rosales, Danitza Morales Gómez & Araceli Rodríguez Saro Vargas, *The Emotional Intelligence of Judges as a Fundamental Component for The Administration of Justice*, 15 Mex. L. Rev. 115 (2022); Terry Maroney, *The Emotionally Intelligent Judge: A New (and Realistic) Ideal*, 49 Ct. Rev. 100 (2013).

[31] For a comparative overview of establishing the facts, different types of evidence, and the role of the judge during this stage of the proceeding, see John Langbein, *The German Advantage in Civil Procedure*, 52 U. Chi. L. Rev. 823 (1985). *See also* Anna Carpenter et al., *Judges in Lawyerless Courts*, 110 Geo. L.J. 509 (2022) (arguing that judges should take a more proactive role engaging unrepresented parties who lack familiarity with court proceedings).

[32] Susan Haack, *Epistemology and the Law of Evidence: Problems and Projects*, *in* Evidence Matters: Science, Proof, and Truth in the Law 1-26, 1 (2014) ("Every legal system needs, somehow, to determine the truth of factual questions.").

[33] Legal scholarship mostly focuses on the role of emotional intelligence or social skills (sometimes interchangeably). They do so, however, in a rather abstract manner without specific attention to judges. *See, e.g.*, Christine C. Kelton, *Clients Want Results, Lawyers Need Emotional Intelligence*, 63 Clev. St. L. Rev. 459, 474 (2014) (proposing a model of emotional intelligence for lawyers); Cyril Jaksic & Katja Schlegel,

The absence of statutory guidance in evidence evaluation does not confer unfettered discretion upon the judge. Rather, creating a factual foundation demands a meticulous consideration of the entire evidentiary record acquired in the initial stage of fact finding.[34] In essence, the judge must possess the capacity to:

1. Determine the credibility of persons and the reliability of objects (e.g., a document, a piece of cloth) serving as evidence;
2. Comprehend the information gathered during the taking of evidence;
3. Relate that information to lived experience; and
4. Articulate truth conditions in the real world that demonstrate genuine understanding of the evidence presented.[35]

In short, a judge's understanding of the real world must be functional. Knowledge of logical and natural laws enables the identification of contradictions within witnesses' narratives or litigants' claims. Moreover, this understanding also integrates less quantifiable elements such as life experience and knowledge of human nature. These pragmatic-subjective components are indispensable in evaluating the demeanor and credibility of parties or witnesses.[36]

In conclusion, judicial fact-finding necessitates a capacity for comprehensive human communication. This foundational capacity includes both the ability to perceive symbolic, acoustic, and visual signals during the taking of evidence and the capability to understand, evaluate, and describe these signals during evidence

Accuracy in Judging Others' Personalities: The Role of Emotion Recognition, Emotion Understanding, and Trait Emotional Intelligence, 8 J. INTELL. 34 (2020).

[34] Judges are obligated to assess evidence to recognize certain claims as facts of the case. Review courts refer to this competency as discernment of "substantial evidence." According to *Maynard v. City of S. J.*, 37 F.3d 1396, 1404 (9th Cir. 1994), substantial evidence is "such relevant evidence as reasonable minds might accept as adequate to support a conclusion even if it is possible to draw two inconsistent conclusions from the evidence." *Cf.* Louis L. Jaffe, *Judicial Review: Constitutional and Jurisdictional Fact*, 70 HARV. L. REV. 953 (1957); Louis L. Jaffe, *Judicial Review: "Substantial Evidence on the Whole Record,"* 64 HARV. L. REV. 1233 (1951).

[35] In addition, Volokh names in this regard the ability to accurately evaluate a witness's demeanor; the ability to reveal inconsistencies; and the ability to evaluate human biases, perception, and memory. Volokh, *supra* note 1, at 1154.

[36] On the complexity of determining the credibility of evidence presented to a judge, see, e.g., Stephen Porter & Leanne ten Brinke, *Dangerous Decisions: A Theoretical Framework for Understanding How Judges Assess Credibility in the Courtroom*, 14 LEGAL & CRIMINOLOGICAL PSYCH. 119 (2009); Ronald Taft, *The Ability to Judge People*, 52 PSYCH. BULL. 1 (1955).

assessment and fact determination. Underpinning these abilities is the judge's internalized model of the real world, which serves as the foundation for credible and just decision-making.

B. Judicial Fact-Finding by AI? Assessing and Producing Text

For AI to serve as a judge, it must demonstrate the social skills necessary to perform the judicial task of fact-finding. This section examines the capabilities and limitations of AI in replicating judicial language comprehension, beginning with the capability to assess and produce text. This ability is necessary for evaluating the accuracy of documents and the reception and evaluation of written expert reports or opinions. Additionally, textual comprehension may be necessary to receive and evaluate textual elements of visual artefacts.

Natural Language Processing (NLP), a subfield of AI, has made substantial strides in text analysis and generation. NLP encompasses a variety of applications aimed at enabling human-machine dialogue through natural language, including text.[37] The field is currently undergoing rapid development, particularly with the advent of large language models (LLMs) like GPT-3[38] and its successors. The increased accessibility and capabilities of these models have significantly advanced NLP's performance, making it stand out as a potentially transformative technology for the legal profession.[39]

However, a critical distinction must be made between language processing and genuine comprehension. This dichotomy lies at the heart of ongoing debates in linguistics, philosophy of language, and cognitive science. To fully appreciate the capabilities and limitations of AI in this domain, it is necessary to examine the

[37] *See generally* Chandrika G, *Natural Language Processing (NLP)*, 12 Int'l J. Rsch. Applied Sci. & Eng'g Tech. 1092 (2024), https://www.researchgate.net/publication/381844702_Natural_Language_Processing_NLP [https://perma.cc/UE74-P9VU]; *Natural Language Processing*, Sciencedirect.com, https://www.sciencedirect.com/topics/engineering/natural-language-processing [https://perma.cc/PY3K-XX36] (last visited Mar. 10, 2026).

[38] *See* Tom B. Brown et al., *Language Models Are Few-Shot Learners*, arXiv (July 22, 2020), http://arxiv.org/abs/2005.14165 [https://doi.org/10.48550/arXiv.2005.14165].

[39] *See, e.g.*, K. R. Chowdhary, *Natural Language Processing*, *in* Fundamentals of Artificial Intelligence 603 (K.R. Chowdhary ed., 2020); Prakash M Nadkarni, Lucila Ohno-Machado & Wendy W. Chapman, *Natural Language Processing: An Introduction*, 18 JAMA 544 (2011); Arkaitz Zubiaga, *Natural Language Processing in the Era of Large Language Models*, 6 Frontier A.I. (2024), https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1350306/full [doi:10.3389/frai.2023.1350306].

technological underpinnings of NLP and the philosophical debates surrounding machine understanding of language.

1. Technological Underpinnings of NLP: From Rule-Based Systems to Deep Learning

The evolution of NLP begins with knowledge-based models, which represent the initial attempt to imbue machines with linguistic comprehension capabilities. These systems rely on pre-programmed rules, lexicons, and semantic frameworks to parse and interpret text. The process typically unfolds in three distinct phases: This initial stage (crawling) involves the systematic indexing and structuring of relevant documents, extracting metadata and core textual content. Then, the system employs parsing algorithms to identify language components, breaking down pages, paragraphs, and sentences into discrete tokens for analysis (tokenization and syntactic preparation). The final stage (semantic evaluation) involves the application of parsing rules to assign meaning to words and phrases based on their context and relationship to other textual elements.[40]

These knowledge-based NLP applications utilize dictionaries and knowledge models to identify synonyms and process ambiguous words, thereby offering a certain level of transparency in their decision-making process. They do, however, often struggle with the nuances and complexities inherent in natural language. Recognizing the limitations of rule-based systems, the field of NLP has increasingly embraced Machine Learning (ML) techniques, with a particular emphasis on Deep Learning (DL) methods. This shift represents a fundamental change in approach, moving from explicitly programmed rules to AI systems that can learn linguistic patterns and relationships from vast corpora of text data.[41]

The advent of Large Language Models (LLMs), exemplified by OpenAI's GPT series, marks a significant milestone in this evolution. These models, trained on unprecedented volumes of textual data, demonstrate remarkable capabilities in generating human-like text and performing a wide array of language-related tasks. The release of ChatGPT has catapulted NLP into the public consciousness,

[40] On the foundations of knowledge-based NLP, see, e.g., Wendy G. Lehnert, *Knowledge-Based Natural Language Understanding*, *in* Exploring Artificial Intelligence 83 (Howard E. Shrobe & the American Association for Artificial Intelligence eds., 1988).

[41] Ilias Chalkidis & Dimitrios Kampas, *Deep Learning in Law: Early Adaptation and Legal Word Embeddings Trained on Large Corpora*, 27 A.I. & L. 171 (2019).

showcasing its potential to engage in seemingly intelligent dialogue across diverse domains, including law.[42]

2. The Dichotomy Between Language Processing and Genuine Comprehension

Despite these advancements, a fundamental question persists: Do these sophisticated AI systems truly understand language,[43] or are they merely engaging in highly advanced pattern recognition?

Noam Chomsky and other prominent scholars have articulated a fundamental critique of current AI approaches to language. They argue that systems like ChatGPT, despite their sophisticated pattern recognition capabilities, lack a genuine understanding of language's underlying structure and meaning. The crux of their argument rests on the observation that these AI systems cannot explicate the syntactic rules governing the languages they appear to use fluently.[44]

Conversely, other scholars like Steven T. Piantadosi and Daniel Everett argue that NLP systems can acquire necessary language structures.[45] However, their arguments focus on opposing Chomsky's theory of Universal Grammar rather than proving AI's genuine language comprehension. They stress that AI systems learn patterns and associations without explicit grammar or syntax rules, suggesting that natural language is acquired through interaction rather than inherent principles.

---

[42] *See, e.g.*, S. Balasubramaniam et al., Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks (2024); B. Yadav, *Generative AI in the Era of Transformers: Revolutionizing Natural Language Processing with LLMs*, 4 J. Image Proc. & Intel. Remote Sensing 54 (2024).

[43] Whenever this article refers to deep understanding or true language comprehension, it is based on the way in which Gadamer conceives of understanding, and interpretation, namely "a mode of insight that has its own rationality irreducible to any simple rule or set of rules, that cannot be directly taught, that necessarily implicates an understanding of self as well as other, and that is always oriented to the particular case at hand". *See* Malpas, *supra* note 17.

[44] *See* Noam Chomsky, Ian Roberts & Jeffrey Watumull, *The False Promise of ChatGPT*, N.Y. Times (Mar. 8, 2023), https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html [https://perma.cc/V69E-UXCL]. *See also* Noam Chomsky & Ramin Mirfakhraie, *ChatGPT and Human Intelligence: Noam Chomsky Responds to Critics*, MRonline (Apr. 23, 2023), https://mronline.org/2023/04/24/chatgpt-and-human-intelligence-noam-chomsky-responds-to-critics/ [https://perma.cc/GWC2-P7X9].

[45] *See* Steven Piantadosi, *Modern Language Models Refute Chomsky's Approach to Language* (2024), https://lingbuzz.net/lingbuzz/007180 [https://doi.org/10.5281/zenodo.12665933]; Mohammad Mazhari, *Exclusive: Linguist Says ChatGPT Has Invalidated Chomsky's 'Innate Principles of Language,'* Tehran Times (2023), https://www.tehrantimes.com/news/483187/Exclusive-Linguist-says-ChatGPT-has-invalidated-Chomsky-s-innate [https://perma.cc/N27X-3EUB].

Critics, in turn, respond that NLP applications only describe or predict text based on their training data model, thus producing what some characterize as pseudoscientific outputs.[46] By heavily relying on statistical learning from vast datasets, NLP therefore faces inherent limitations in its approach to language. These systems, while adept at predicting likely word sequences or generating contextually appropriate responses, operate without a grounded understanding of the real-world concepts and relationships that language describes.[47]

For AI to truly understand natural language, it must seek explanations independently and consider improbable explanatory models. Current AI systems, including LLMs, lack this capability, having no personal perspective, subjective experiences, or self-awareness, resulting in their inability to reliably generate meaningful content in the way human judges do (so-called deep understanding).[48]

3. Implications for AI in Judicial Fact-Finding

This critique, particularly the disconnect between statistical language models and genuine understanding, has profound implications for the use of NLP in legal contexts. If AI systems are merely engaging in sophisticated pattern matching rather than true language comprehension, their reliability in interpreting text in the way required of a judge comes into question.

It is evident that the complexity of text processing required varies depending on the specific scope of the judicial tasks. Broadly speaking, these judicial tasks can be divided into two main categories: organizational and evaluative tasks. With regard to organizational tasks, the performance level of NLP

---

[46] *See* Dan Milway, *A Response to Piantadosi (2023)*, Dan Milway (Apr. 26, 2023), https://milway.ca/2023/04/26/a-response-to-piantodosi-2023/ [https://perma.cc/3QFV-S3TK]; Jordan Kodner, Sarah Payne & Jeffrey Heinz, *Why Linguistics Will Thrive in the 21st Century: A Reply to Piantadosi (2023)*, https://lingbuzz.net/lingbuzz/007485 [https://perma.cc/E83N-HB6V].

[47] This results potentially in AI hallucinating. *See, e.g.*, Razvan Azamfirei, Sapna R. Kudchadkar & James Fackler, *Large Language Models and the Perils of Their Hallucinations*, 27 Critical Care 120 (2023); Vipula Rawte et al., *The Troubling Emergence of Hallucination in Large Language Models – An Extensive Definition, Quantification, and Prescriptive Remediations*, arXiv (July 30, 2025), http://arxiv.org/abs/2310.04988 [https://perma.cc/LG48-WKY9].

[48] Mathematically speaking, the lack of deep understanding follows from the fact "(i) that computation cannot realize understanding, (ii) that computation cannot realize mathematical insight, and (iii) that computation cannot realize raw sensation, and hence that computational syntax will never fully encapsulate human semantics." J. Mark Bishop, *Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It*, 11 Front. Psych. (2021).

is promising. NLP systems typically excel in tasks that involve organizing, categorizing, and summarizing large volumes of textual data.[49] In the context of judicial fact-finding, these capabilities could support strategic case scheduling, efficient document management, and the development of systematic approaches to evidence compilation and analysis. For instance, in complex multi-party litigation, a judge could utilize NLP tools to rapidly parse thousands of documents, quickly identifying critical themes and potential evidentiary inconsistencies that traditionally would require weeks of manual review. Current NLP technology can already adequately handle these judicial tasks, potentially even enhancing the efficiency of the fact-finding process.[50]

However, current NLP systems struggle with nuanced evaluative tasks that demand sophisticated contextual understanding. Assessing written expert testimony presents a particularly challenging domain: NLP technologies struggle to holistically evaluate logical consistency, detect subtle rhetorical maneuvers, and gauge real-world plausibility. Contractual interpretation represents another example where current NLP technologies reveal substantial interpretative constraints when assessing written text. While they can identify specific textual patterns, these systems often fail to capture the complex interplay between contractual language, industry-specific norms, parties' implicit intentions, and broader contextual considerations that human legal professionals routinely navigate.[51]

While NLP can assist with initial document organization and preliminary textual analysis, independently evaluating their content's validity and coherence, particularly in the context of a specific case, is beyond its current capabilities. Such tasks would demand not only linguistic processing but a complex interplay

[49] *See, e.g.*, Meg Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 Vand. J. Ent. & Tech. L. 77, 105 (2015).

[50] *See, e.g.*, Farid Ariai & Gianluca Demartini, *Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges*, arXiv (July 30, 2025), http://arxiv.org/abs/2410.21306 [https://perma.cc/U8YH-EF2C]. Specifically with regard to judicial decision-making, see Beniamino Di Martino et al., *A Semantic-Based Methodology for the Management of Document Workflows in e-Government: A Case Study for Judicial Processes*, 66 Knowledge Info. Sys. 3959, 3959–3982 (2024).

[51] *See*, Ariai & Demartini, *supra* note 50, at 1449 (discussing limitations of NLP systems). *See, e.g.*, Marco Siino et al., *Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches*, IEEE Access 1 (2025) (overviewing AI-based contract analysis as well as contract review).

of world knowledge, logical reasoning, and contextual understanding—capabilities that current AI systems have not yet convincingly demonstrated.

C. Judicial Fact-Finding by AI? Verbal Communication through Spoken Language

Judges, of course, require not only the skills to comprehend written language, but also spoken natural language. The AI technology most analogous to human verbal communication falls again within the domain of NLP. These AI-based speech processing systems share fundamental operational principles with text-based NLP, differing primarily in their input and output interfaces. As such, the limitations inherent to NLP in text processing, as discussed in the previous section, apply equally to speech processing systems.

1. The Interplay of Text and Speech Processing in the Course of Judicial Fact-Finding

Judicial text and speech processing are inextricably linked, both in their technological counterparts and within the judicial fact-finding process itself. A judge's verbal interactions in court are grounded in the thorough analysis of the written language in various case files and documents. This preparatory phase requires the synthesis of textual information to inform subsequent verbal inquiries. An AI system designed to replicate this process would need to seamlessly integrate text analysis capabilities with speech generation functions.[52]

To develop such a system, one could envision training an AI on a dataset comprising of former case files, their transcripts, and the questions posed by judges during the corresponding proceedings. The training data could potentially extend to include judges' factual findings and legal assessments, thereby addressing

[52] The technology behind it is usually called "text-to-speech" or "speech-to-text". *See, e.g.*, Yogesh Kumar, Apeksha Koul & Chamkaur Singh, *A Deep Learning Approaches in Text-to-Speech System: A Systematic Review and Recent Research Perspective*, 82 Multimedia Tools Applications 15171 (2023); Mahmoud Gaber, Gloria Corpas Pastor & Ahmed Omer, *Speech-to-Text Technology as a Documentation Tool for Interpreters: A New Approach to Compiling an Ad Hoc Corpus and Extracting Terminology from Video-Recorded Speeches*, Trans: revista de traductología 263 (2020). Well-known examples of AI-based speech processing tools include the speech to text tools from OpenAI, https://platform.openai.com/docs/guides/speech-to-text [https://perma.cc/R3FM-PHW9] and Google Cloud, https://cloud.google.com/speech-to-text [https://perma.cc/Y9XC-MSW5].

the challenges posed by the intertwined nature of fact-finding and legal analysis inherent in judicial decision-making.[53]

2. Communication and Interaction with Parties and Witnesses

Moving to the interactions between judges and witnesses or parties as the core judicial task involving skills to communicate verbally, it becomes clear that replicating this facet of the judicial fact-finding process requires more from an AI than merely following a predetermined list of questions. Rather, it involves conducting a dialogue, with the judge not only asking questions but also understanding the responses in the context of the specific case and continuously adjusting the content of the examination accordingly. AI needs to be able to engage in real-time analysis of courtroom proceedings, actively comparing new information against the pre-existing case file. Gaining this ability would enable AI to identify unexplored areas of inquiry relevant to the case, formulate additional questions based on emerging testimony, recognize inconsistencies or contradictions in witness statements, and adapt its line of questioning based on new information or unexpected responses.

3. Once again: Absence of Genuine Language Comprehension as AI's Fundamental Challenge

Structurally, AI seems capable of replicating the requirements for verbal communication skills in judicial fact-finding. However, a fundamental limitation persists: the absence of genuine language comprehension. As explored above when discussing text-based NLP, current AI technologies, including LLMs, lack understanding of the content they process in any deeper sense.[54]

This deficiency is particularly problematic in the context of judicial fact-finding for several reasons. First, judicial proceedings can encompass a virtually unlimited range of subject matters, from complex financial transactions to nuanced interpersonal disputes. An AI system would need to possess not just broad general knowledge but also the ability to apply this knowledge contextually to specific cases—a task that requires more than pattern recognition or statistical

[53] This specific challenge will be discussed in Section II.A.

[54] *See, e.g.*, Melanie Mitchell & David C. Krakauer, *The Debate Over Understanding in AI's Large Language Models*, 120 Proc. Natl. Acad. Sci. U.S.A. e2215907120 (2023).

analysis. Second, judges do not merely engage in general communication; they filter information through a sophisticated legal lens. This requires not only language comprehension but also a deep grasp of legal principles, precedents, and their application to novel situations. The integration of legal reasoning with verbal communication presents a challenge that current AI systems are ill-equipped to handle.[55]

Third, while AI systems excel at identifying patterns across large datasets, they struggle to treat each case as unique, with its own specific context and nuances. AI tends to apply "average" solutions based on their training data, which can lead to inappropriate generalizations in specific cases and potentially compromise the fairness and accuracy of the judicial process.[56] These shortcomings are closely tied to AI systems' lack of intuitive understanding of the world (often referred to as "common sense") and inability to reason about their own knowledge and limitations (meta-knowledge) in the way that human judges can.[57] Additionally, current AI systems identify correlations rather than causal relationships.[58] This limitation can likewise result in logical inconsistencies when addressing the complex, real-world scenarios presented in court.

Against this backdrop, the result is similar to the one regarding text-based judicial tasks: While AI systems—particularly those based on advanced NLP models—can generate syntactically correct and contextually relevant questions and responses, they fundamentally differ from human judges in their

[55] *See infra* Section II.E.

[56] The "world" of an AI system thus begins and ends with the data that is made available to it. That is why even in the case of LLMs like ChatGPT, the users most likely will find a disclaimer that the program has, for example, "limited knowledge of the world and events after 2021". *See, e.g.*, *What is ChatGPT?*, OpenAI, https://help.openai.com/en/articles/6783457-what-is-chatgpt [https://perma.cc/2ZYK-K8RY] (last visited Aug. 28, 2024). This, in turn, results in the AI system not knowing anything about the "world" beyond the date of their training data cut-off.

[57] On the resulting (lack of) understanding of Al of the world, see, e.g., Mittelstadt, *supra* note 26, at 1831 ("The concept of 'truth' has been highly simplified in LLM development and equated with accuracy measured against the 'ground truth' of the training data.").

[58] This is a characteristic of AI considered as one of the most central challenges when delegating human-centric decision-making processes to AI. This limitation has been flagged by scholars for decades. *See, e.g.*, Nils J. Nilsson, *Artificial Intelligence: Engineering, Science, or Slogan?*, 3 AI Magazine 2 (1982).

underlying reasoning processes.[59] The output of these systems, while sometimes impressively accurate, can also be unpredictably erroneous or nonsensical.[60] This unpredictability stems from the AI's lack of genuine understanding of the content it processes.[61] As the requirements for judicial verbal communication extend beyond mere form to encompass substantive understanding and reasoned decision-making, current AI technologies, despite their advancements, therefore fall short of meeting these comprehensive requirements. The limitations in AI-based speech processing, particularly the lack of genuine content comprehension, preclude the full replication of a judge's verbal communication capabilities in the fact-finding process.

D. Judicial Fact-Finding by AI? Nonverbal Communication

Perhaps the most challenging aspect for AI in replicating judicial social skills is the interpretation of nonverbal cues. When examining the requirement of a judge's nonverbal communication skills, we shift our focus from the substance, the objective content of testimony, to the nuanced manner in which it is delivered—what linguists term "paralanguage."[62] These nonverbal cues often provide crucial insights into witnesses' and parties' subjective feelings and motives. Therefore, they play a pivotal role in judicial assessments of credibility, a cornerstone of effective fact-finding.[63]

The judicial task of recognizing and evaluating subjective feelings through nonverbal cues presents several fundamental challenges. Unlike verbal content, subjective feelings are not directly perceptible to outside observers, including

[59] Jocelyn Maclure fittingly refers to AI having an "atomistic bias which makes it blind to the social and institutional dimension of human reasoning processes." Jocelyn Maclure, *AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind*, 31 Minds & Machines 421, 421 (2021).

[60] On the "weirdness" of AI, particularly output of ML, see, e.g., Janelle Shane, You Look Like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place 11–12 (Illustrated ed. 2019) (on file with NYU Journal of Intellectual Property and Entertainment Law).

[61] This lack of genuine understanding of the content it processes has been traced back to lack of a "deep understanding of the structure of the world" comparable to the ones of humans based on their "prior experience." *See* Douglas Heaven, *Why Deep-Learning AIs Are so Easy to Fool*, 574 Nature 163 (2019).

[62] Tim Wharton, *Paralanguage*, *in* The Routledge Handbook of Pragmatics 69 (Anne Barron, Yueguo Gu, & Gerard Steen eds., 2017).

[63] *See generally* United States v. Scheffer, 523 U.S. 303 (1998) (emphasizing the importance of demeanor evidence in credibility assessments).

judges.[64] Consequently, the judge must rely on proxies—external manifestations of internal states. These proxies primarily include facial expressions, voice pitch and tone; physiological indicators like pulse rate, gestures, and posture.[65]

1. AI-based Emotion Recognition

Recent advancements in AI have led to the development of systems aimed at replicating nonverbal communication skills as required of judges.[66] These AI applications typically follow a three-stage process:

1. Sensors record signals such as facial expressions or voice pitch;
2. The system performs feature abstraction and classification, assigning observations to predefined basic emotions; and
3. The system infers the overall emotional state based on this classification.

In the context of judicial fact-finding, three types of external manifestations seem particularly promising for emotion recognition: facial expressions (visual), prosody (acoustic), and overall behavior, particularly gestures.

Facial expression recognition aims to infer emotional states from facial muscle movements and eye contact patterns.[67] While subtle and fleeting expressions pose challenges for human judges, AI systems excel at capturing these micro-expressions. AI-based facial expression recognition[68] employs either 2D or 3D modeling, with three-dimensional approaches offering greater accuracy

[64] Antonio Damasio, The Feeling of What Happens: Body and Emotion in the Making of Consciousness 54 (1999) (pointing out that only in very few cases "we can predict with some success that certain stimuli will produce certain emotions").

[65] On how to detect deceit from these external manifestations, see Paul Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage 80-161 (2001).

[66] *See, e.g.*, The Oxford Handbook of Affective Computing (Rafael A. Calvo et al. eds., 2015).

[67] Yingli Tian, Takeo Kanade & Jeffrey F. Cohn, *Facial Expression Recognition*, *in* Handbook of Face Recognition 487 (Stan Z. Li & Anil K. Jain eds., 2011); Peter Enders, *Einsatz Künstlicher Intelligenz bei juristischer Entscheidungsfindung* [*Usage of Artificial Intellligence in Legal Decision-Making*], 50 Juristische Arbeitsblätter [JA] 721 (2018) (Ger.); Jacqueline Melzer, *Auswirkungen von Künstlicher Intelligenz auf Völkerrecht, insbesondere die Gewährleistung der Garantien des Art. 6 EMRK* [*Implication of Artificial Intelligence on Public International Law, particularly the Guarantee of Art. 6 ECHR*], Zeitschrift zum Innovations- und Technikrecht [ZZIT] 145 (2020) (Ger.).

[68] For an overview, see Hakan Boz & Utku Köse, *Emotion Extraction from Facial Expressions by Using Artificial Intelligence Techniques*, 9 BRAIN: Broad Rsch. in A.I. & Neuroscience 5 (2018).

and pose independence.[69] While these systems have shown promise in medical contexts, such as diagnosing mental health disorders, concerns persist about the potential conflation of correlation and causation in AI-generated insights, as well as the generalizability of results across diverse populations.[70]

Acoustic emotion recognition focuses on speech prosody—the "tone of voice" beyond verbal content.[71] This approach analyzes sound properties such as intensity, intonation, and speech rhythm. However, cultural variations in intonation patterns and the heterogeneous nature of acoustic features complicate the development of universally applicable AI systems for prosody analysis.

Comprehensive analysis of a witness's or party's overall behavior, especially gestures, represents another avenue for AI-assisted emotion recognition in the courtroom. While AI systems capable of analyzing complex physical movements exist—such as those used in Olympic gymnastics scoring[72]—applying similar technology to courtroom gesture analysis presents unique challenges due to the absence of a comprehensive "evaluation catalog" for linking gestures to specific emotions.

2. Fundamental Limitations of AI in Emotional Understanding

Despite the apparent sophistication of AI systems in capturing and processing various manifestations of nonverbal communication, they face significant—and potentially insurmountable—hurdles in evaluating the meaning of these cues vis-à-vis emotional states. This limitation is not merely technological but reflects a profound epistemological challenge.

---

[69] *See, e.g.*, Abhinav Chaubey & Vijay Maheshwari, *Facial Emotions Recognition Using Deep Learning Technology*, 9 Int'l J. for Rsch. in Applied Sci. & Eng'g Tech. 1245 (2021).

[70] "Lack of generalization" und "lack of trust in automated decision-making" are usually identified as the main challenges when it comes to AI-based emotion recognition. *See, e.g.*, Smith K. Khare et al., *Emotion Recognition and Artificial Intelligence: A Systematic Review (2014–2023) and Research Recommendations*, 102 Info. Fusion 1, 12 (2024).

[71] Hiroya Fujisaki, *Prosody, Models, and Spontaneous Speech*, *in* Computing Prosody 27 (Yoshinori Sagisaka, Nick Campbell & Norio Higuchi eds., 1997).

[72] Takuya Matsumoto, *New AI Scoring System Eyed for 2024 Paris Olympics*, Japan News (Oct. 12, 2023), https://japannews.yomiuri.co.jp/sports/other-sports/20231011-142383/ [https://perma.cc/ZN3P-R2V2].

The interpretation of nonverbal cues presents a twofold conundrum: first, the deficiencies of human judges in accurately decoding emotional states,[73] and second, the inherent difficulty of objectively verifying the accuracy of such interpretations, whether performed by human or AI. The deficiencies of human judges in accurately identifying others' emotional states might, prima facie, suggest a compelling case for AI replacement in this domain. However, this human limitation creates a paradoxical impediment to assessing AI's efficacy in this realm. The lack of a "ground truth"[74] against which to measure AI performance renders the evaluation of its capabilities in correctly detecting emotional states of humans inherently challenging.[75]

The crux of this dilemma, therefore, lies in the absence of an objective, reliable methodology for verifying emotional states of humans. While self-reporting remains a cornerstone of psychological inquiry,[76] its application in the adversarial context of judicial proceedings is fraught with complications. The inherent tensions between truth-seeking and strategic testimony render such self-reports unreliable at best and potentially hazardous at worst.

Moreover, AI systems themselves cannot truly experience emotions in the way humans do, therefore lacking the experiential foundation necessary for genuine emotional understanding. They can merely recognize and potentially mimic emotional expressions.[77] AI, however, lacks profound understanding of the

[73] Rainer Banse & Klaus R. Scherer, *Acoustic Profiles in Vocal Emotion Expression*, 70 J. Personality & Soc. Psych. 614 (1996) (concluding that human judges do detect emotions from nonverbal acoustic signals "with much-better-than-chance accuracy"). The authors, nevertheless, stress that this aspect still requires further empirical research. Porter & ten Brinke, *supra* note 82 (pointing to various challenges in this regard).

[74] For an example of how to approach tasks without a ground truth like emotion recognition, see Ramya Srinivasan & Ajay Chander, *Crowdsourcing in the Absence of Ground Truth—A Case Study* (June 17, 2019) (unpublished manuscript) (presented at 2019 ICML Workshop on Human in the Loop Learning (HILL 2019)), http://arxiv.org/abs/1906.07254 [https://perma.cc/7MFA-6QAL].

[75] This is a particularly challenging case resulting from the general difficulty for humans to identify and assess failures of AI. Conversely, AI may obscure human intent and responsibility, when the human is placed in the loop. *See* Crootof, Kaminski & Price, *supra* note 24, at 467.

[76] Especially in the course of determining to what extent a person is capable of empathy and by what characteristics this can be measured, such self-disclosures (so-called "self-reports") are used. *See, e.g.*, Brett A. Murphy & Scott O. Lilienfeld, *Are Self-Report Cognitive Empathy Ratings Valid Proxies for Cognitive Empathy Ability? Negligible Meta-Analytic Relations with Behavioral Task Performance*, 31 Psych. Assessment 1062 (2019).

[77] Alain Finet, Kevin Kristoforidis & Julie Laznicka, *The Limits of AI in Understanding Emotions: Challenges in Bridging Human Experience and Machine Perception*, F1000Research, 3 (2025), https:

embodied, subjective experience of emotions that informs human interpretation of nonverbal cues.

This absence of lived emotional experience creates an unbridgeable gap between AI analysis and human emotional interpretation.[78] It reflects the structural differences between human cognition and AI. Human emotions form the core of subjective inner experiences—what philosophers term qualia, the ineffable "what it is like" to be a specific conscious entity. To truly understand human emotions requires being human, echoing Thomas Nagel's famous argument about the subjective nature of conscious experience.[79]

3. Implications for Judicial Fact-Finding

The judiciary's role in evaluating witness credibility and emotional states is predicated on a shared human experience—resulting in a common vocabulary that allows for nuanced interpretation of nonverbal cues. This shared experiential base enables judges to respond to the individual circumstances of the disputing parties and empathetically grasp possible motives, information, and emotional needs in their social context.

AI's inability to genuinely experience emotions renders it fundamentally ill-equipped for this critical aspect of judicial fact-finding. While AI may excel in identifying surface-level nonverbal cues, it lacks the deeper contextual understanding and experiential framework necessary for meaningful interpretation of these cues.

E. Conclusion

The analysis in this Part of the article reveals a nuanced landscape where AI demonstrates potential in isolated aspects of judicial fact-finding, particularly in specific tasks such as document authentication or organizational functions. However, fundamental limitations persist, especially in areas requiring deep

//f1000research.com/articles/14-582 [https://doi.org/10.12688/f1000research.164796.1]; Angelina Chen et al., *Feels Like Empathy: How "Emotional" AI Challenges Human Essence*, Austl. Conf. Info. Sys. 1, 3 (2023), https://ssrn.com/abstract=4763428 [https://perma.cc/6ZU8-E3NU].

[78] Sadia Tariq et al., *Examining Some Serious Challenges and Possibility of AI Emulating Human Emotions, Consciousness, Understanding and 'Self,'* 1 J. NeuroPhilosophy 73 (2022) (stressing that the practical realization of AI having emotions "would be very difficult, if not impossible").

[79] Nagel, *supra* note 18, at 435, 442 ("[T]o know what it is like to be a bat, you have to be a bat.").

language understanding, emotional intelligence, and holistic reasoning about human behavior as well as internal human states.

Even advanced AI applications, such as LLMs, fall short of the judicial competencies required for comprehensive fact-finding. The inability to fully grasp communicated content or to interpret nonverbal cues precludes AI from wholly supplanting human judges in this domain.

II
Legal Abilities for Legally Assessing the Facts

In addition to a nuanced understanding of the social dimensions of adjudication, the judicial function necessitates equally sophisticated legal skills. Such abilities, encompassing a comprehensive understanding of the legal system and the ability to apply general norms to specific factual scenarios, are universally considered an essential component of the judicial function.[80] The judge's mandate to resolve legal disputes on the basis of the law necessarily presupposes a sophisticated understanding of legal principles and their application to factual scenarios.

This Part of the article aims to elucidate the key steps a judge undertakes in legally analyzing established facts. Notably, the normative foundations for conducting legal analysis are sparse. Core considerations regarding judicial approaches to applying the law primarily fall within the domain of legal methodology instead.

A. The Interplay of Facts and Law

Before elaborating on the process of judicial legal assessment, it is worth noting that the legal analysis is not confined to the post-fact-finding stage. Rather, it permeates the entire judicial process. The judge must consider potential legal interpretations ab initio, beginning with the review of party submissions and continuing throughout the fact-finding process. Judicial decision-making is, therefore, characterized by a constant, dynamic interplay between the different

[80] *See, e.g.*, Marbury v. Madison, 5 U.S. 137 (1803) (establishing the principle of judicial review and, by extension, the necessity of judicial legal competence).

stages of the proceeding.[81] The judge must navigate these steps with the flexibility demanded by due process, considering the results of each in relation to the others.

The dialectic between factual and legal considerations occurs both during oral arguments and in the drafting of the opinion. Throughout oral arguments, parties' assertions and proffered evidence serve as touchstones for potential legal interpretations for the judge, ensuring comprehensive evidence gathering and facilitating meaningful dialogue with parties on both factual and legal grounds.

The opinion-drafting stage represents the culmination of this fact-law symbiosis.[82] Here, established facts serve as the foundation for definitive legal analysis, reflecting an increasing narrowing of focus throughout the proceedings. Although opinions present factual findings and legal assessments as distinct components, they are the product of an integrated cognitive process not explicitly captured in the written decision.[83]

Against this backdrop, it is evident that effective adjudication demands more than isolated social and legal competencies from the judge. It requires an integrative capacity, a holistic understanding of the interplay between real-world occurrences and their legal implications.

B. The Application of Law in Detail

So, how do judges apply legal norms to facts—whether alleged or established—and what are the specific legal skills required for this process?

The initial step in a judge's legal analysis, when confronted with a case, is to determine the applicable area of law. This determination, which precedes the oral arguments and, consequently, the fact-finding process itself, is an integral part of the preliminary jurisdictional assessment. It demands not so much a detailed knowledge of specific statutes, but rather a fundamental understanding of the

[81] Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281, 1297 (1975) (referring to a "complex and continuous interplay between fact evaluation and legal consequence").

[82] On the complexity of dividing questions of law from questions of fact, see G. Alexander Nunn, *Law, Fact, and Procedural Justice*, 70 EMORY L.J. 1273 (2021).

[83] *See, e.g.*, Mireille Hildebrandt, *Data-Driven Prediction of Judgment. Law's New Mode of Existence?*, *in* OUP COLLECTED COURSES VOLUME EUI SUMMER-SCHOOL 3 (2019) (exemplifying this back and forth in the following way: "what counts as the criminal offence of murder depends on the double focus of looking at the facts from the perspective of the criminal law and looking at the criminal law from the perspective of the facts").

structure of the legal system as a whole, coupled with the ability to connect various legal domains to real-world events. Such comprehension is crucial for classifying the factual scenario as legally relevant and, more specifically, as falling within the subject-matter jurisdiction of the specific court.

Following this initial legal categorization of the case as within the court's jurisdiction, the next step involves identifying the specific legal principles and legal provisions potentially applicable to the case at hand.

Courts typically employ a process known as subsumption or legal syllogism, which involves applying a major premise (legal rule) to a minor premise (factual situation) to reach a conclusion.[84] In judicial reasoning, the major premise consists of legal rules of a certain level of abstraction expressed in natural language, while the minor premise is the specific fact pattern at issue.[85]

While subsumption follows an "if-then" structure, it would be overly reductive to view it as a purely mechanical process.[86] The logical structure does not account for the content of the premises, such as the meaning of terms in the legal rule. Judges must engage in a nuanced analysis, concretizing both the legal concepts and the factual elements to determine whether the facts align with the statutory elements, thereby triggering the prescribed legal consequences.[87]

The content of these legal texts is determined through legal interpretive methods. While these are typically not explicitly enumerated in positive law, their existence as well as their guiding function as tools for arguing for or against particular interpretations of legal norms are widely recognized. What these legal interpretive methods are and their weight in relation to one another rank, however, among the most debated issues—both within each legal system and between different legal traditions.[88]

---

[84] On the concept of subsumption in judicial decision-making, see, e.g., Frederick Schauer, *Balancing, Subsumption, and the Constraining Role of Legal Text*, 4 L. & Ethics Hum. Rts. 33 (2010).

[85] David Lyons, *Justification and Judicial Responsibility*, 72 Calif. L. Rev. 178, 180 (1984).

[86] Lochner v. New York, 198 U.S. 45 (1905) (Holmes, J., dissenting) (critiquing formalistic legal reasoning).

[87] Jerzy Wroblewski, *Legal Syllogism and Rationality of Judicial Decision*, 5 Rechtstheorie 33 (1974).

[88] Most notably for the US discourse is the well-known debate between originalists living constitutionalists as the two competing methods of constitutional interpretation. Despite each having a certain basic concept with defining features, both originalism and living constitutionalism are multilayered terms which can hardly be conclusively defined. In fact, there are more than just these two "schools" or theories of how to interpret the

When approaching legal interpretation, we may consider as minimal common ground to acknowledge that law is—at least for now—expressed through natural language. Therefore, its interpretation must adhere at least to general linguistic rules, considering not only the common usage of terms but also their context and broader communication practices. This process involves both semantic and pragmatic analysis. Through this lens, the traditional rules of legal interpretation can consequently be understood as specifications of the rules of natural language.[89]

As is the case with any interpretation of written natural language, the starting point for interpreting legal norms is the text itself.[90] Judges must rely on their linguistic competence as members of a certain community, focusing on the ordinary meaning of words. The textual argument is therefore an argument directly based on the text itself by considering the words used.[91] However, textual analysis is not limited to semantics; it also encompasses grammatical and syntactical considerations.[92]

Moving to the pragmatic dimension of legal interpretation, systematic interpretation as one of its methods examines a norm within the broader context of the legal system, considering its relationship to other norms.[93] This method often

constitution. On each side, different sub forms have emerged. Even among originalism, there are "different kinds of originalism", making the lines to living constitutionalism blurry to a degree that some forms of originalism become "indistinguishable from a form of living constitutionalism". *See* David A. Strauss, The Living Constitution 10–11 (2010).

[89] For an overview of legal interpretation and how it is connected to the philosophy of language, see, e.g., Brian H. Bix, *Legal Interpretation and the Philosophy of Language*, *in* The Oxford Handbook of Language and Law 146 (Lawrence M. Solan & Peter M. Tiersma eds., 2012), https://doi.org/10.1093/oxfordhb/9780199572120.013.0011.

[90] A famous textualist—though in the context of statutory interpretation—was Justice Antonin Scalia. He focused on the text alone, leaving aside the historical dimension with the argument that history is either a secret or cannot be trusted. Instead, his source and main reference point was the dictionary when it comes to determine the meaning of the text; more accurately: its words. *See, e.g.*, Antonin Scalia & Bryan A. Garner, Reading Law: The Interpretation of Legal Texts (2012).

[91] As Philip Bobbitt puts it, "the plain words of some portion of the text." *See* Philip Bobbitt, Constitutional Fate: Theory of the Constitution 38 (1982). The meaning of the words is derived from the "common use of a particular term in a particular context." *Id.* at 26.

[92] Heikki E. S. Mattila, *Legal Vocabulary*, *in* The Oxford Handbook of Language and Law 27 (Lawrence M. Solan & Peter M. Tiersma eds., 2012).

[93] In the context of interpreting the US Constitution, this approach is often referred to as "structural" argument. The aim is, however, the same: The way of determining the overall context the text is embedded in is defined by Bobbitt as "deceptively simple logic moves from the entire Constitutional text rather than from one of its parts". Bobbitt, *supra* note 91, at 74.

provides more insight into legislative intent than mere textual analysis. Historical interpretation, another method of pragmatic legal interpretation, seeks to discern the legislature's intent at the time of enactment. It considers legislative history, including committee reports, floor debates, and the legal landscape at the time of enactment. Finally, the teleological interpretation, another method of pragmatic legal interpretation, focuses on the purpose or objective of a norm, which may be explicitly stated in the text or legislative materials, or inferred from context. This method of interpretation requires judges to consider both factual and value-based experiences, including generally recognized principles of justice and economic efficiency.

The interpretive process, particularly in its pragmatic aspects, demands that judges think holistically[94] about the legal system and explore the underlying motives and objectives of legal provisions. This necessitates a combination of factual knowledge, understanding of potential consequences, and recognition of societal values.

C. Judicial Decision-Making Despite Interpretive Ambiguity

The process of legal interpretation may lead to multiple equally valid results, creating a discretionary space that cannot be fully resolved through applying recognized interpretive methods. How should a judge proceed in this scenario? Such situations lie at the heart of judicial decision-making in cases of legal ambiguity,[95] raising fundamental questions about the nature of judicial discretion and the line between applying existing law and creating new law.

[94] Jack Balkin has argued against the backdrop of the US Constitution that the question whether one considers it as a living document that adapts to changing circumstances or whether it is to be interpreted according to its original meaning is actually not one where making a choice for one or the other leads to a satisfying result. Instead, these two approaches need to be considered compatible rather than opposed. *See* Jack Balkin, Living Originalism 3, 277 (2011) (describing a holistic approach to interpretation). Similarly, in this regard, Akhil Reed Amar: "We are all textualist; we are all living constitutionalists". *See* Akhil Reed Amar, America's Unwritten Constitution xiii (2012). He argues for constitutional interpretation as a mean to grasp the constitution as encompassing as possible. Such a "comprehensive account of America's Constitution" can be seen as an "aspiration to holism." Akhil Reed Amar, America's Constitution: A Biography 465, 469 (2005).

[95] Ralf Poscher, *Ambiguity and Vagueness in Legal Interpretation*, *in* The Oxford Handbook of Language and Law 128 (Lawrence M. Solan & Peter M. Tiersma eds., 2012).

Traditionally, this aspect of judicial decision-making is addressed primarily from a legal theoretical point of view. Positivistic concepts of the law, such as H.L.A. Hart's theory of the "open texture" of law or Kelsen's Pure Theory of Law,[96] posit that judges must select from among plausible interpretations to fulfill their adjudicative function. However, such frameworks offer little guidance on the criteria for such selections, potentially relegating the choice to unfettered judicial discretion.[97] This prospect, in turn, raises concerns about democratic legitimacy and the rule of law.

To mitigate the risk of arbitrary decision-making, legal systems often establish both procedural and substantive constraints on judicial discretion. Various doctrines and canons have been developed to guide judicial decision-making in cases of ambiguity. For instance, the US Supreme Court has articulated different "tiebreaker" rules for resolving interpretive deadlocks, such as the doctrine of constitutional avoidance[98] and the rule of lenity in criminal cases.[99]

Constraints of this kind serve a dual function: guiding judges in navigating interpretive ambiguities and reinforcing the legitimacy of judicial decisions. Procedural constraints, such as the obligation to provide reasoned decisions, serve as a check on judicial discretion by requiring transparency and facilitating appellate review.[100]

Substantive constraints often manifest in prohibitions against decisions based on improper motives. These principles effectively incorporate extra-statutory normative considerations into the decision-making process, referencing concepts such as fundamental legal principles, general principles of justice and fairness and the preservation of constitutional structure.[101]

---

[96] *See* H.L.A. Hart, *supra* note 10; Hans Kelsen, Pure Theory of Law (1967).

[97] Richard S. Higgins & Paul H. Rubin, *Judicial Discretion*, 9 J. Leg. Stud. 129 (1980).

[98] *See, e.g.*, NLRB v. Catholic Bishop of Chicago, 440 U.S. 490 (1979).

[99] *See, e.g.*, United States v. Gradwell, 243 U.S. 476 (1917); McBoyle v. United States, 283 U.S. 25 (1931); United States v. Bass, 404 U.S. 336 (1971).

[100] These constraints will be examined in greater depth in Part III of this article.

[101] The Austrian Civil Code (ABGB) provides an instructive example of how statutory frameworks can guide judicial decision-making. Sections 6 and 7 of the ABGB establish a "decision program" program that not only mandates judicial interpretation but also provides guidance when even comprehensive analogy fails to yield a clear solution. Notably, Section 7 directs judges to decide according to "natural principles of law" in such cases. Allgemeines bürgerliches Gesetzbuch [ABGB] [Civil Code] § 6–7, <https://www.ris.bka.gv

Despite these constraints, there might still be cases that cannot be resolved even when considering all available interpretive tools and normative principles. Judges are nevertheless obliged to decide every case within their jurisdiction, regardless of the degree of ambiguity.[102] Therefore, it remains conceivable that judicial decisions may contain a random element that cannot be predetermined or guided by law.

D. Summary of the Required Judicial Skillset

Based on the preceding analysis of the judicial process of legally assessing the facts of the case, the following competencies form the core legal skills judges require:

- A comprehensive understanding of the structure of the legal system and the interplay between facts and law;
- Proficiency in legal interpretation, including mastery of semantic and pragmatic aspects of legal language;
- The ability to apply and synthesize various interpretive methodologies;
- The capacity to engage in meta-analysis of competing interpretations;
- An awareness of prevailing societal values and their relative weights in different legal contexts;
- The ability to navigate residual discretionary spaces within established legal and normative frameworks.

The process of legal assessment by the judge must always culminate in a decision, including a decision for one of the plausible meanings of the applicable legal norms. Only by making the judicial decision is the ambiguity resolved for the case in question, and the process of law application concluded.

E. Judicial Legal Assessment by AI? Legal Abilities of AI

In light of the judicial legal skills outlined above, this section examines the extent to which AI can fulfill these requirements.

.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001622&Artikel=&Paragraf=0&Anlage=&Uebergangsrecht=> (Austria).

[102] Lyons, *supra* note 85, at 183.

1. Legal Assessment as AI-Based Language Processing?

At first glance, the application of AI to the task of judicial legal assessment might seem like yet another use case of natural language processing. After all, the law is, as elaborated above, expressed through natural language. From statutory texts to judicial opinions, the legal profession is steeped in linguistic analysis and argumentation. However, this surface-level similarity belies the profound yet unique complexities of legal language in general and the process of judicial legal application more specifically.[103]

Legal language, while rooted in natural language, operates with a specialized logic and precision that diverges significantly from everyday communication.[104] Legal terms often carry meanings that are distinct from their colloquial usage, and the process of interpreting these terms follows specialized methodologies that are not intuitive to those outside the legal profession. Moreover, judges are bound by procedural and substantive constraints in their legal reasoning that are not readily apparent from the text alone.[105]

Thus, while proficiency in natural language processing is undoubtedly a prerequisite for any AI system aspiring to take over judicial functions, it is far from sufficient. The true challenge lies in replicating the specific processes of legal assessment as required of a judge.

2. AI and the Interplay of Facts and Law

One of the core challenges of judicial legal assessment for AI is replicating the dynamic interplay between facts and their legal classification. As elaborated above, judges engage in a constant cognitive back and forth between the factual circumstances of a case and the relevant legal provisions. This process is not linear or mechanical but involves continuous reassessment and refinement of both the factual and the legal dimension.

Conceptually, this process could be modeled as a data comparison between two pools: one containing factual information and another containing legal

---

[103] Mattila, *supra* note 92.

[104] Peter M. Tiersma, *A History of the Languages of Law*, *in* The Oxford Handbook of Language and Law 12 (Lawrence M. Solan & Peter M. Tiersma eds., 2012).

[105] *See, e.g.*, Hildebrandt, *supra* note 83, at 13–14.

principles. However, this simplification fails to capture the nuanced, iterative nature of judicial legal assessment. Each new piece of factual information can potentially alter the relevance or interpretation of legal provisions in light of the specific case, and vice versa.

While AI systems seem structurally equipped for this back and forth and could potentially be designed to perform iterative analyses between factual and legal data sets, replicating the subtle, intuitive aspects of this process remains a significant challenge. The ability to recognize when, *e.g.,* a seemingly minor factual detail necessitates a wholesale reevaluation of the applicable legal framework is a hallmark of legal expertise that AI systems struggle to emulate.

3. The Limits of Rule-Based AI in Legal Assessment

Moving from analyzing the compatibility of AI with the overall framework and structural requirements underlying the process of judicial legal assessment to the replicability of the process itself, I will now examine the extent to which AI is able to replicate the substantive dimension of this core judicial task.

At first glance, traditional rule-based AI systems (often referred to as GOFAI—Good Old-Fashioned AI) might seem well-suited to replicate the process of judicial legal assessment, given law's reliance on rules and logical structures. Indeed, in highly standardized legal processes, such as certain tax calculations or administrative procedures, rule-based AI can be effective.[106]

However, these standardized processes are the exception rather than the rule in judicial decision-making. As elaborated above, the vast majority of cases that come before judges involve complex, nuanced scenarios that resist reduction to simple "if-then" statements. Judicial discretion, the weighing of competing principles, and the need to interpret ambiguous or conflicting legal sources are all crucial aspects of the judicial function that rule-based AI struggles to replicate.

[106] On the challenges of using GOFAI for legal assessment, see Burkhard Schafer, *Formalising Law, or the Return of the Golem*, *in* Research Handbook on Law and Technology 59 (Bartosz Brożek, Olia Kanevskaia, & Przemysław Pałka eds., 2023).

4. Machine Learning and Legal Assessment: A Partial Solution with New Challenges

In contrast to these traditional rule-based AI systems, ML-based approaches offer a potentially more flexible solution for modeling judicial legal assessment, particularly in capturing its evaluative dimensions. Conceptually, ML could be applied to the process of legally assessing facts by perceiving the latter as a classification task. Supervised learning algorithms could be trained on datasets of past cases, using the established facts as input and the legal decisions as output. In theory, such a system could derive the rules by which judges move from factual inputs to legal outputs, and then apply these derived rules to new fact patterns.[107]

However, this approach faces several significant limitations, too: Most importantly, unlike human judges who apply deductive reasoning to explicit legal norms, an ML system would need to inductively derive the rules of legal subsumption from the provided data. The system would not be pre-programmed with established methods of legal interpretation or the general legal norms to be applied; instead, it would need to independently identify these as patterns within the data.[108] This inversion of the deductive reasoning process that judges are expected to employ raises serious doubts as to whether ML-based legal skills can ever truly replicate the judicial process of legal assessment as envisioned in our legal system.[109] Even with high-quality training data, no AI system to date has successfully derived the comprehensive rules of legal subsumption used in judicial reasoning, even for limited case types.[110]

---

[107] For a use case of a ML-based approach to modeling judicial legal assessment, see Riya Sil, Alpana & Abhishek Roy, *Machine Learning Approach for Automated Legal Text Classification*, 13 Int'l J. Comput. Info. Sys. & Indus. Mgmt. Applications 10 (2021).

[108] On the problem the one cannot rely upon ML algorithms employing inductive methods to learn valid generalizations, see Daniel Harris & Darin Dunham, *Induction in Machine Learning* (2021), https://ieeexplore.ieee.org/abstract/document/9438325 [https://perma.cc/2TKM-98N5].

[109] Olsen, Slosser & Hildebrandt, *supra* note 22, at 226.

[110] Cor Steging, Silja Renooij & Bart Verheij, *Rationale Discovery and Explainable AI*, *in* Frontiers in Artificial Intelligence and Applications, 229. (Erich Schweighofer ed., 2021). Katie Atkinson, Trevor Bench-Capon & Danushka Bollegala, *Explanation in AI and Law: Past, Present and Future*, 289 A.I. 1, 15 (2020), https://www.sciencedirect.com/science/article/pii/S0004370220301375?via%3Dihub [https://doi.org/10.1016/j.artint.2020.103387].

Furthermore, this approach relies on the assumption that past decisions are inherently correct and should be replicated—a premise that ignores the evolving nature of law and the occasional need for courts to overturn precedent.[111]

5. Hybrid AI Systems: A Promising Approach

Given the technological limitations of both GOFAI and ML approaches, scholars increasingly advocate for hybrid AI systems that combine elements of both. This approach aims to leverage the strengths of both AI paradigms to more closely approximate judicial legal assessment.[112]

Hybrid systems offer several advantages. The ML component allows for consideration of a wide variety of case scenarios and the identification of patterns that may not be easily formalized a priori.[113] This addresses the challenge that law, expressed in natural language, cannot be fully captured by formal logical rules alone. Simultaneously, the GOFAI component introduces explicit legal knowledge and logical rules derived from applicable legal norms, helping to legally categorize the patterns identified by the ML component.

While such a hybrid system would still ultimately produce outputs of a statistical nature, the integration of rule-based reasoning would allow these outputs to be generated following explicit methodological guidelines. This approach, therefore, represents a more accurate replication of judicial reasoning processes than either GOFAI or ML alone could achieve.

6. Limitations of AI in Legal Interpretation

Despite the potential of hybrid AI systems, significant challenges remain in fully replicating the process of judicial legal assessment. These challenges arise both in determining the content of general legal norms and contextualizing specific facts in light of these norms, as well as in filling any remaining interpretive gaps.

---

[111] Simon Deakin & Christopher Markou, *Evolutionary Interpretation: Law and Machine Learning*, 1 J. Cross-disciplinary Rsch. Comput. L. 1 (2022).

[112] Jack Mumford, Katie Atkinson & Trevor Bench-Capon, *Combining a Legal Knowledge Model with Machine Learning for Reasoning with Legal Cases*, *in* Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law 167 (2023).

[113] The way ML works may therefore be referred to as "case-based reasoning" in contrast to GOFAI as "rule-based reasoning." For more on this conceptual difference, see Atkinson, Bench-Capon & Bollegala, *supra* note 110, at 3.

6.1. *The Dichotomy of Formal and Substantive Legal Application*

A fundamental challenge in AI-based legal assessment lies in the nature of AI cognition itself. While AI systems, particularly those based on ML, excel at identifying patterns and making predictions based on those patterns, they lack the capacity for substantive reflection on the content and implications of their predictions. Even hybrid AI systems, despite being potentially more sophisticated in their approach, do not engage in the kind of reflective reasoning that characterizes human judicial legal assessment. An AI system can determine with high accuracy that fact X fulfills element Y of legal norm Z, but it cannot engage in a meaningful reflection on *why* this is the case or consider the broader implications of this conclusion. This distinction between formal pattern matching and substantive understanding once more raises questions about the extent to which AI-driven legal reasoning can truly capture the nuanced, context-dependent nature of human judicial decision-making.[114]

6.2. *Challenges in Semantic Interpretation*

To truly replicate the process of judicial legal assessment, an AI system would need to master the intricate process of legal interpretation outlined above, employing recognized hermeneutical methods to uncover the substantive content of applicable legal norms. This requirement demands a level of interpretive sophistication that, at minimum, satisfies the threshold for methodological rigor in legal interpretation.

Semantic interpretation, being the most formalizable aspect of legal interpretation, initially appears to be the low-hanging fruit for AI systems within the process of judicial legal assessment. An AI system could theoretically utilize a combination of dictionaries, legal commentaries, and large-scale textual corpora to determine the meaning of words and interpret legal language. This approach

[114] *See* Crootof, Kaminski & Price, *supra* note 24, at 466 (pointing out that "algorithms lack the social conditioning and tacit knowledge that humans have and thus miss crucial unarticulated (even inarticulable) aspects of human decisionmaking" and that "algorithms that try to 'learn' ethics or social norms from human behavior can import the nastier elements along with the good" due to a lack of understanding for the norms or ethics in question).

mirrors, to some extent, the resources that human judges might consult in their interpretive process.[115]

As elaborated above, particularly recent advancements in NLP have demonstrated an impressive ability to capture semantic nuances in natural language, making it seem technologically possible to replicate semantic interpretation through AI. However, despite these advancements, AI still falls short of replicating the level of language competence that judges possess as members of the linguistic and legal community—a competence required even at the level of semantic interpretation. The human capacity for understanding subtle connotations, recognizing cultural and historical context, and appreciating the evolving nature of language usage remains unmatched by current AI systems.[116]

Furthermore, the semantic interpretation of legal language often requires an understanding of specialized legal terminology and concepts that may not be fully captured by general-purpose NLP models.[117] The development of domain-specific legal language models presents a potential solution to this challenge, but goes hand in hand with raising the converse question about the generalizability and adaptability of such AI models. The latter seems especially tricky in the face of the application of these AI systems to a diverse set of cases, as well as the required ability to handle evolving legal language and concepts.

6.3. *The (Even More) Complex Landscape of Pragmatic Interpretation*

While semantic interpretation presents significant challenges, the pragmatic dimension of legal interpretation poses an even greater hurdle for AI systems. As elaborated above, pragmatic interpretation encompasses a range of interpretive approaches, including systematic, subjective-historical, and objective-teleological perspectives on legal norms. This multifaceted approach to interpretation requires not just an understanding of the literal meaning of legal texts, but also a

[115] Tommaso Agnoloni & Giulia Venturi, *Semantic Processing of Legal Texts*, *in* Handbook of Communication in the Legal Sphere 109 (Jacqueline Visconti ed., 2018).

[116] *See, e.g.*, Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 Ga. St. U. L. Rev. 1305, 1325 (2019) ("[F]or many problem areas there is no easy way to identify or capture the relevant knowledge. In some cases, key concepts or abstractions cannot be meaningfully encoded in a computer-understandable form.").

[117] Gregor Behnke & Niklas Wais, *On the Semantic Difference of Judicial and Standard Language*, *in* Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law 382 (2023).

deep appreciation of their broader context, historical development, and intended purposes. To fully replicate this complex process of pragmatic interpretation, with its various methods and contextual considerations, AI would therefore, among others, require the ability to contextualize individual legal norms within the broader framework of a legal system, consider the historical context and legislative intent behind legal provisions, evaluate the potential consequences and societal impacts of different interpretations, balance competing legal principles and policy considerations, and adapt interpretive approaches based on the specific facts and circumstances of each case.[118]

This level of contextual understanding and adaptive reasoning stretches far beyond the current capabilities of AI systems. While ML algorithms can identify patterns in large datasets of legal decisions, they struggle to capture the nuanced reasoning processes that underlie these decisions.[119] Moreover, the inherently open-textured nature of many legal concepts and the evolving nature of legal interpretation present additional challenges.[120] Legal interpretation often requires judges to apply established principles to novel situations or to reevaluate existing interpretations in light of changing societal norms and values. This dynamic aspect of legal reasoning is particularly difficult to capture in AI systems, which rely on historical data and established patterns.[121]

7. Limitations of AI in Exercising Judicial Discretion

Even if all these various methods of legal interpretation could be successfully modeled and applied by AI systems, yet another fundamental challenge remains:

---

[118] Hildebrandt, *supra* note 83, at 13 (stressing that applying the law to a specific set of facts happens within a "shared world" dimension between the human judge and those subject to the decision, with this "shared world" dimension being connected with empathy and common sense, tacit knowledge and the need to anticipate the institutional world we navigate; she further stresses that "[t]his dimension, however, is not shared by the NLP systems that cannot but restrict themselves to text as data").

[119] *See, e.g.*, Sourdin, *Judge v Robot?*, *supra* note 5, at 1130 ("The technology possesses the ability to process and manipulate these symbols, but it does not understand the meaning behind these processes. In other words, the machine does not understand the information that it is processing.").

[120] Clement Guitton et al., *The Challenge of Open-Texture in Law*, A.I. L. (2024).

[121] Since ML models derive their function from historical training data, they are based on the assumption that the future will be the same as the past (so-called "stationarity assumption"). This assumption may, however, be challenged and consequently falsified by various kinds of changes within society, including changes in law and its application (so-called "concept/population drift"). *See* Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 Harv. J.L. & Tech. 1, 21–22 (2018).

As elaborated above, different interpretive approaches can lead to conflicting results, and there is no fixed hierarchy or predetermined formula for resolving these conflicts. Additionally, the methods of legal interpretation are themselves dynamic and highly context dependent.

When the interpretative process yields no unequivocal result, judges must exercise discretion to fulfill their decision-making obligation. In these scenarios, an AI system would require the capability to make a decision from within, just like the human judge. Before making this decision, it would need to conduct a case-specific evaluation that considers a wide range of factors, including the specific facts and circumstances of the case, the broader legal and social context in which the case arises, the potential future implications of the decision as well as considerations of justice, equity, and public policy. This decision-making process is not merely a matter of applying predefined rules or weighing quantifiable factors. Instead, it is a complex social and cognitive practice that draws upon a judge's accumulated knowledge, experience, and understanding of legal principles and social norms.[122]

7.1. *Inherent Unsuitability of Rule-Based AI*

From a structural perspective, only ML-based AI systems could potentially replicate exercising judicial discretion. Rule-based AI systems (GOFAI) which solve problems using rigid if-then schemas, are, in contrast, inherently unsuitable for the process of judicial decision-making beyond legal interpretation and the navigation of discretionary spaces.

The "rules" governing these discretionary decisions—if they exist at all—are neither complete nor objectively representable in a form accessible to AI.[123] This lack of direct, unambiguous instructions to the judge is precisely what characterizes this aspect of judicial decision-making in the first place. Cases of this type typically involve multiple (more or less concrete, explicit, or implicit) objectives or interests that cannot be fully reconciled with each other. The weight a judge should assign to each objective or interest in the balancing process is rarely predetermined and depends on the specific circumstances of the case. Ultimately, determining the

[122] *See supra* Section II.C.

[123] This is a case of "tacit knowledge," i.e., "knowledge that can't always readily be translated into code." *See* Crootof, Kaminski, and Price, *supra* note 24, at 462; *see also* Frank Pasquale, New Laws of Robotics 24 (2020) ("We know more than we can explain.").

relationship between different interests always involves a value judgment entrusted to the judge.

7.2. *Can AI Autonomously Navigate Discretionary Spaces?*

In contrast to GOFAI, ML systems could be capable of learning from training data how to handle discretionary spaces within the judicial decision-making process and subsequently exercise discretion independently to reach a legal assessment of a novel factual situation. However, this approach faces, yet again, two fundamental technological challenges:

First, training data rarely capture how the specific discretionary space in question has been filled by judges in past cases. This is because the discretionary space is itself a consequence of the specific circumstances of each case, making it highly unlikely that a past case with an identical discretionary space has been decided. This creates a factual obstacle for an AI system in deriving how to fill the discretionary space in a specific case, even if the underlying pattern identification process were theoretically amenable to technological replication.[124]

Second, even in the unlikely event that the training data include a case with the same discretionary spaces as the case the AI system is set to decide, the AI system would not be legally bound by how other judges in the past exercised discretion in the previous case. Classifying the AI system's "blind adoption" of another judge's exercise of discretion as "voluntary" does not solve this discrepancy, either, since it would not meet the requirements placed on judges in relation to the exercise of discretion. Filling such discretionary spaces is entrusted specifically to the judge responsible for deciding the concrete case, not just any judge. This judge—here: the AI system assigned to deciding the specific case—must fill this space independently from within, after appropriate reflection and consideration, in light of their own innate sense of justice.[125] While it is conceivable that they reach the

[124] Frank Fagan & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. Cal. L. Rev. 1, 11 (2019) ("The very process of assigning a standard or a discretionary decision to a machine requires detailed rules; true standards are for human decisionmakers. The same might be said about the ways in which humans make decisions when instructed with standards, but the critical difference is that human judges can rely on intuition and, in any event, (like machines) may be unable to identify their own basis for decisionmaking.").

[125] *See supra* Section II.C.

same result as another judge would or has in the past, the decision on how to fill the discretionary space must be substantively made by the competent judge.

An AI system, by its very nature, is incapable of making such a decision from within itself.[126] Even if ML systems may appear to be able to create genuinely new outputs, these are always merely compositions of already known elements. This holds true even when the specific way in which the AI system has combined different preexisting elements is not explicitly represented in the training data.[127]

Moreover, AI lacks awareness of the distinction between law application and law creation. It also does not grasp the significance of discretionary spaces and the corresponding responsibility of judges in the democratic law-creation process in any meaningful way. This is yet another consequence of AI's general lack of fundamental understanding of human concepts. From a technological perspective, this "translation problem" of discretionary spaces in judicial decision-making into an AI-comprehensible form manifests in AI systems interpreting such spaces as noise, i.e. occurrences in the training data that should be filtered out as much as possible when identifying decision rules to achieve a model that most accurately reflects the decision-making process. At best, AI "circumnavigates" these spaces. However, it does not learn how to navigate these spaces as such.[128]

Even if an AI system were capable of recognizing discretionary spaces within the law and determining that such spaces require to be filled in by the judge, the AI system's approach would fundamentally diverge from the exercise of judicial discretion. AI's decision-making process would likely default to a simplistic heuristic: "In the absence of explicit rules or discernible patterns, insert the statistically most probable outcome or a random selection." This algorithmic

[126] Stuart Russell, Human Compatible: Artificial Intelligence and the Problem of Control 10 (2019) ("Because machines, unlike humans, have no objectives of their own, we give them objectives to achieve. In other words, we build optimizing machines, we feed objectives into them, and off they go.").

[127] This aspect is much discussed when it comes to AI "creating" art, e.g. by composing music. *See, e.g.*, Arthur I. Miller, The Artist in the Machine: The World of AI-Powered Creativity (2019); Corey Ford et al., *Reflection Across AI-Based Music Composition*, C&C '24 Proc. of the 16th Conf. on Creativity & Cognition 398 (2024), https://doi.org/10.1145/3635636.3656185.

[128] Stephan Dreyer & Johannes Schmees, *Künstliche Intelligenz als Richter? Wo keine Trainingsdaten, da kein Richter—Hindernisse, Risiken und Chancen der Automatisierung gerichtlicher Entscheidungen* [*Artificial Intelligence as a Judge? No training data, no judge - obstacles, risks and opportunities of automating judicial decisions*], 35 Computer und Recht [CUR] 758, 762 (2019) (Ger.).

approach stands in stark contrast to the nuanced, deliberative, and value-laden process of judicial discretion,[129] as outlined above.

F. Conclusion

The judicial process of legal assessment is far from being a simple algorithmic procedure. Rather, it represents a complex, multi-faceted endeavor that resists comprehensive replication by AI systems. While narrow, standardized legal procedures may be amenable to automation, the vast majority of judicial legal assessment involves a degree of nuance and complexity that current AI technology is ill-equipped to handle.

The limitations of AI in this regard are not merely quantitative—that is, a matter of processing power or data volume—but qualitative, stemming from fundamental differences in cognitive architecture between human and machine intelligence. Because AI "thinks" in ways that are fundamentally alien to human legal reasoning,[130] its modes of reasoning remain structurally misaligned with the judicial process.

The process of legal assessment is inextricably linked to law as a human construct, expressed through natural language and interpreted through culturally-informed hermeneutics.[131] AI can, at best, imitate aspects of this process but falls

[129] For an illustrative example, see Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI*, 41 Comput. L. & Sec. Rev. 1, 7 (2021) on EU discrimination law ("Due to the disparate nature of algorithmic and human discrimination, the EU's current requirements are too contextual, reliant on intuition, and open to judicial interpretation to be automated. Many of the concepts fundamental to bringing a claim, such as the composition of the disadvantaged and advantaged group, the severity and type of harm suffered, and requirements for the relevance and admissibility of evidence, require normative or political choices to be made by the judiciary on a case-by-case basis … we show that automating fairness or non-discrimination in Europe may be impossible because the law, by design, does not provide a static or homogenous framework suited to testing for discrimination in AI systems.").

[130] One may even go as far as to say that AI does not think at all. *See, e.g.*, Crootof, Kaminski, and Price, *supra* note 24, at 465.

[131] This human-centric nature of law was recognized by Niklas Luhmann in his 1997 work on administrative decision-making, where he argues that the ultimate aim of legal interpretation is to ensure human decision-making rationality. He stresses in this regard that "system autonomy, however, does not mean arbitrariness, and certainly not personally motivated arbitrariness of the system members" (translation by author). *See* Niklas Luhmann, Recht und Automation in der öffentlichen Verwaltung: eine verwaltungswissenschaftliche Untersuchung [Law and Automation in Public Administration: an Administrative Legal Study] 35 (1997). Thus, for an AI system to truly perform legal assessment,

short of replicating this level of contextual understanding and reflective, context-sensitive, adaptive reasoning. Even in cases where AI systems demonstrate high accuracy rates in predicting judicial outcomes—as seen in some assessments of "predictive justice" tools[132]—this superficial success belies a deeper failure to grasp the underlying legal norms and methodological requirements that guide judicial legal assessment.[133] The inability of AI to comprehend these elements, even at a rudimentary level, underscores the qualitative gap between statistical correlation and true legal reasoning. Additionally, AI might perpetuate or even amplify discriminatory patterns present in training data. Unlike humans, who can potentially recognize and counteract their biases, AI systems lack an understanding of why certain criteria are discriminatory and should not be considered in legal assessments.[134]

AI therefore lacks the necessary capabilities to fully replicate the legal abilities required of a judge, precluding the comprehensive transfer of judicial fact-based legal assessment to AI systems.

III
Ability of Legal Reasoning to Justify Judicial Decisions

Once a judge has rendered a decision, both in terms of establishing the facts and their legal assessment, thus concluding the decision-making process, their task is not yet fully accomplished. In the vast majority of cases, judges are obligated to articulate the rationale underpinning their decisions.[135] This judicial reasoning,

---

it must not only produce results that fall within acceptable interpretive boundaries but also arrive at these results through processes that mirror human legal reasoning.

[132] Predictive justice refers to the "analysis of large amounts of judicial decisions by artificial intelligence technologies in order to make predictions for the outcome of certain types of specialised disputes." *See* The European Commission for the Efficiency of Justice, European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment 74 (2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c [https://perma.cc/2222-8TGP].

[133] *See, e.g.*, *id.* at 29 (*id.* at 29 ("In most occasions, the objective of these systems is not to reproduce legal reasoning but to identify the correlations between the different parameters of a decision … and through the use of machine learning, to infer one or more models. Such models would be used to 'predict' or 'foresee' a future judicial decision.").

[134] *See generally* Paola Lopez, *Reinforcing Intersectional Inequality via the AMS Algorithm in Austria*, Conf. Proc. of the STS Conf. Graz 2019 302 (2020).

[135] The concept of the rule of law provides a central basis for the requirement of judicial reasoning as it the essence of the rule of law principle culminates in the requirement that all acts of state organs must be grounded in law and that a system of legal protection institutions guarantees that only those acts that have

while inextricably intertwined with factual findings and their legal assessment, constitutes a distinct component of the judicial activity. Its aim is to communicate the judge's decision in a comprehensible manner to the parties involved and the broader legal community. The discursive nature of judicial reasoning necessitates a separate examination of the skills required for this task.

A. The Architecture of Judicial Reasoning: The Subsumption Model as a Foundational Framework

At its core, judicial reasoning aims to establish a logical connection between the legal rule, the facts, and the legal consequence. The subsumption model serves as the primary structural scaffold for this endeavor, thereby providing a framework for both the application of law[136] and the justification of the resulting decision. In essence, the subsumption model dictates that the judge's decision (legal consequence) is derived by deduction from the grounds for the decision. As already elaborate above, the concept of subsumption falls, however, short in addressing what is often the most challenging aspect of judicial decision-making: the justification of premises. The subsumption model thus offers a useful starting point for conceptualizing judicial reasoning, but it is not a comprehensive framework.

B. The Relationship Between the Judicial Decision and Its Justification: Production vs. Presentation or Back-and-Forth 2.0?

An examination of the relationship between the judicial decision as such and its justification provides further insight into the judge's approach to reasoning—both structurally and substantively. Contrasting judicial decision-making with its justification, the former can be understood as the process of producing the judgment, while the latter represents its presentation.[137] Simultaneously, it is critical to recognize that the decision-making process and the justification of the decision as its outcome cannot be fully disaggregated. Rather, there is an intrinsic

---

been issued in accordance with the higher-level acts conditioning them become permanently part of the legal order.

[136] *See supra* Section II.A.

[137] *See* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 636 (1995). Schauer's understanding of "reason" is a structurally similar, generalized version as he is concerned "not with *having* a reason, but instead with *giving* a reason, and with what follows from the very act of giving one."

connection between judicial decision-making and its reasoning, and thus between the production and presentation of the judicial decision.[138]

This connection manifests in two distinct ways: First, the necessity of justification directly constrains the judge's decision. If a legal outcome cannot be justified in the manner legally required, the judge is precluded from rendering this outcome. Second, the judge must approach the decision-making process in a manner that can be reproduced in the judgment's reasoning. Consequently, judges must orient themselves to the same standards during the decision-making process that they may argumentatively invoke when drafting its justification.[139]

A failure to maintain this alignment would result in justifications that qualify as mere pretense, aimed solely at retrospectively lending the appearance of "legal rationality" to a previously arbitrary decision, potentially made for extraneous motives. Such an approach would not comport with neither the concept of the judge's role within a system predicated on the rule of law nor with the corresponding understanding of the application of law.[140] It is, rather, an essential characteristic of the judicial reasoning that it is not formulated in isolation from the decision made. Therefore, the judge who decides a case must also be the one who justifies the decision.

---

[138] This way of understanding the relationship between the judicial decision and its justification is prominent in the scholarship of Ulfrid Neumann. *See, e.g.*, Ulfrid Neumann, *Wahrheit statt Autorität* [*Truth instead of Authority*], *in* Die Sprache des Rechts [The Language of Law] 369, 381 (Kent Lerch ed., 2005) (Ger.); Ulfrid Neumann, *Juristische Argumentationstheorie* [*Theory of Legal Argumentation*], *in* Handbuch Rechtsphilosophie [Handbook of Legal Philosophy] 234 (Eric Hilgendorf & Jan C. Joerden eds., 2017) (Ger.).

[139] For this reason, "giving reasons is seen as a necessary condition of rationality" by many. *See* Schauer, *supra* note 138, at 663.

[140] Even from a legal realist perspective which acknowledges that judicial reasoning may not perfectly align with the idealized narrative of purely rational decision-making, the justification process remains constitutive of the judicial decision-making. The requirement of reasoned explanation fundamentally constrains judicial discretion, despite the legal realistic discovery that "judges were human and therefore were likely in a variety of legal contexts consciously or unconsciously to slip their personal values into their legal reasoning." *See* John Hart Ely, Democracy and Distrust: A Theory of Judicial Review 40 (1980). This is because a legal outcome that cannot be articulated through legally requisite forms of reasoning cannot be legitimately rendered.

C. Conceptual Requirements for Judicial Decision Justification: Reasoning as a Mechanism of Legitimacy

The intrinsic connection between judicial decision-making and reasoning necessitates that the content of judicial opinions, and the process by which judges arrive at them, largely mirror the considerations outlined above in Part I and Part II. Thus, this part of the article focuses on whether the externally communicated nature of judicial reasoning imposes distinct or supplementary requirements beyond those inherent in decision-making as an internal judicial cognitive process.

When examining the specific functions of judicial reasoning, it becomes clear that they are interconnected despite being seemingly diverse: they encompass, e.g., the legitimization of state authority, the rationalization and objectification of decision-making processes, the minimization of arbitrary actions by decision-makers, the assurance of comprehensibility for affected parties, the facilitation of appellate review, the demonstration of judicial accountability to the legal community, the enablement of judicial decision quality evaluation, and the fostering of acceptance and trust in state decisions. At their core, these functions coalesce to provide justification for judicial decision-making.[141]

The structure and content of judicial opinions are not solely determined by their subject matter but are inextricably linked to their intended audience, a direct consequence of the judicial opinion's role as a mediating device. This multifaceted communicative dimension requires judges to navigate the general rules of natural language. As previously elaborated, the framework for judicial communication through opinions is thus shaped by prevailing linguistic conventions.[142]

Legal reasoning in judicial opinions is, however, distinguished from quotidian argumentation by its appeal to what Chaïm Perelman termed a "universal audience" —an idealized audience persuadable only by rational legal arguments.[143] While the universality of these rationality concepts across legal systems remains contentious, the existence of shared rationality standards within a specific legal community is generally acknowledged.[144] These standards are primarily manifested in positive

[141] *See* Schauer, *supra* note 138, at 635 ("When lawyers argue and when judges write opinions, they seek to justify their conclusions, and they do so by offering reasons.").

[142] *See supra* Section II.B and Section II.E.1.

[143] *See generally* Chaïm Perelman, Justice, Law, and Argument 71 (1980).

[144] *See* Post, *supra* note 16, at 47–48.

law, particularly in general legal norms, which both dictate the arguments considered relevant by the legal community and stipulate that judges may base their decisions and reasoning exclusively on these legally relevant factors.[145] To render a judicial opinion, judges must therefore demonstrate an understanding of both factual occurrences and their legal implications, as well as their interrelationships, to recognize the legal relevance underlying actual interests (or conflicts thereof) and draw appropriate conclusions.

When the law, even after comprehensive and methodologically correct interpretation, fails to unambiguously dictate the legal outcome, judges must select among equally valid interpretive options.[146] However, even in these scenarios, judges are not entirely unfettered[147] but are subject to both substantive and procedural limitations aimed at preventing arbitrary decision-making. While the substantive standards were examined in II.C., this section focuses on the procedural dimension—specifically, the comprehensive duty to justify judicial decisions, requiring judges not only to disclose their interpretive process but also to elucidate their ultimate decision-making beyond the interpretive framework.

So, what standards govern judicial reasoning in areas of legal indeterminacy? To deny the existence of any guidelines for this part of judicial reasoning would result in a bifurcation of justification requirements: rational legal arguments on one side (within the interpretative framework) and pure arbitrariness on the other (beyond the interpretative framework). Such a dichotomy, however, is inherently contradictory.[148] The very requirement of reasoning in areas not unambiguously determined by law precludes pure arbitrariness. Thus, even in areas

[145] Consequently, judicial opinions must eschew reliance on "legally irrelevant factors." The interests of litigants are decision-relevant only insofar as the legal system ascribes them legal relevance, thereby transmuting them into legal arguments. On whether judges are, in fact, not influenced by irrelevant factors, see, e.g., My Bergius et al., *Are Judges Influenced by Legally Irrelevant Circumstances?*, 19 Law, Probability and Risk 157 (2020).

[146] *See supra* Section II.C.

[147] As Justice Benjamin N. Cardozo observed, "The judge, even when he is free, is still not wholly free. He is not to innovate at pleasure." Benjamin Nathan Cardozo, The Nature of the Judicial Process 141 (1921).

[148] Even Hans Kelsen himself stressed that the legal system must be understood as a rational whole. *See* Hans Kelsen, *Buchbesprechung von: Spiegel, L.: Gesetz und Recht. Vorträge und Aufsätze zur Rechtsquellentheorie. München und Leipzig, 1913* [*Book review of: L. Spiegel, Statute and Law: Lectures and Essays on the Theory of the Sources of Law. Munich and Leipzig, 1913*], Österreichische Zeitschrift für öffentliches Recht [OZFOR] 766, 770 (1914) (Ger.).

where positive law does not comprehensively provide rationality standards, judges must structurally engage in thoroughly legal reasoning processes. This ensures that judicial opinions fulfill their justificatory function and persuade the ideal audience.[149]

Put differently: When confronted with multiple equally valid interpretative options, mere articulation of motives for the decision-making component involving judicial lawmaking is insufficient. Judges must rationalize their choices through a process that transmutes personal judgment into legal argumentation. This process may draw upon several components, including personal judicial values, recognition of individual judicial priming, the concept of the judge as a functional "substitute legislator," considerations of justice and peace as foundational legal purposes, anticipated decision acceptance, and judicial pragmatism, encompassing considerations of legal certainty and institutional functionality. These components, while not exhaustive, represent categories of normative arguments that judges may employ to justify decisions in areas of legal indeterminacy.[150]

D. The Structure of an Opinion of the Court

Having explicated the fundamental concepts underlying judicial reasoning and its relationship to decision-making, this section examines the core features of court opinions. Against the backdrop of the judicial decision-making process as carved out, court opinions should generally address three facets: statements of facts, evaluation of evidence, and legal assessment. The statement of facts of a judicial opinion could, e.g., encompass the essential submissions and motions of the parties, limited to those retaining relevance at the time of decision. It may further delineate undisputed facts and those treated as uncontested. Ideally, this section should culminate in a clear and unambiguous statement of the established facts that form the basis for legal assessment.

---

[149] Lyons, *supra* note 85, at 178–79 ("[I]t seems implausible to assume that no judicial decision in a hard case can be justified simply because no decision is required by law. Courts are expected to justify their decisions, and this requirement does not dissolve when cases are hard.").

[150] One may also make use in this regard of the principles that Lon L. Fuller famously described as the inner morality of law. *See generally* Lon L. Fuller, The Morality of Law (1964). *See* Robert Binkley, *Review of The Morality of Law*, 1965 Duke L.J. 668, 669 (1965) (elaborating that Fuller understands the "law is a tool useful for many purposes. The inner morality of law is nothing more than the rules for the use of this tool, rules justified by the fact that if they are flouted the purposes for which the tool is designed cannot be achieved, and the tool itself will be damaged or destroyed").

The evaluation of evidence section elucidates the judge's assessment of individual evidentiary elements and the criteria that informed their subjective conviction regarding the probative value of each piece of evidence. The extent and detail of this explanation are, of course, contingent upon the circumstances of the specific case. The reasoning must, however, be at least sufficiently comprehensible to facilitate effective appellate review of the evidentiary evaluation. Consequently, the opinion should not only elucidate why the judge considers certain facts proven but also explicate the rationale for rejecting other potentially relevant factual assertions.[151]

The legal assessment section delineates the judge's application of law to the established facts. It is incumbent upon the judge to explicate which legal norms they deem relevant to the established facts, their interpretation of these norms, and the reasoning that led to the conclusion reflected in the judgment's outcome. This explication is crucial for enabling parties to appeal on grounds of incorrect legal assessment. The judge must demonstrate a methodologically sound application of the relevant legal provisions to the established facts, particularly by disclosing their thought process in interpreting vague legal concepts and engaging with any pertinent debates in case law and legal scholarship.

These three main sections of court opinions directly result from and therefore mirror the different stages of judicial decision-making, as outlined above. Rather than constituting a separate component of the judicial decision-making process, the requirement of the judge justifying themselves by giving reasons for their decisions in a legally sound manner adds a layer of explanation to each of these sections. While the judge's reasoning should render its decision-making process intersubjectively comprehensible, there is no obligation to describe every mental step taken in the course of a court proceeding. The depth and breadth of explanations in each section are primarily contingent upon whether the aspect in question is disputed.[152]

---

[151] *See* Neil MacCormick, Legal Reasoning and Legal Theory 67-72 (1994) (arguing that judicial opinion must encompass not only a justification of the deductive process of legal application but also an explication of why one interpretive option was selected over another. The same holds true, he argues, for the factual point of reference of the legal analysis conducted by the judge).

[152] The European Court of Human Rights has developed a nuanced approach to the extent of judicial reasoning required, considering factors such as the nature of the decision and the specificity of the applicable law. *See, e.g.*, Tyrer v. the United Kingdom, App. No. 5856/72 Eur. Comm'n H.R. Dec. & Rep. 12 (1978);

E. Judicial Reasoning by AI?

This section examines whether AI systems possess the capabilities necessary to replicate the reasoning process required of judges when writing their opinion. Unlike the analyses of AI's potential to replicate the social and legal skills required of a judge,[153] the examination of AI's capability to generate legal opinions necessitates a holistic approach. This is particularly due to the internal process of judicial decision-making and the external process of justification being symbiotically linked, which has profound implications for AI's potential to replicate this aspect of the judicial task.

1. Conceptually Distinguishing Decision-Making from Giving Reasons for the Decision

The first conceptual hurdle in evaluating AI's potential for generating a judicial opinion lies in ascertaining whether AI systems can meaningfully differentiate between the internal process of decision-making and the external process of articulating and justifying the decision. This two-layer approach is crucial for the process of judicial decision-making, characterized by the internal deliberative process of reaching a decision being distinct from the external process of justifying it within the legal community.[154]

Conceptually, AI systems seem indeed capable of replicating this distinction. When an AI system makes a decision, it does not necessarily mean that the decision-making process is simultaneously apparent and accessible beyond the system. Instead, explaining the decision and its reasoning is a separate task for AI, too.[155]

Sakkopoulos v. Greece, App. No. 61828/00, 50-51 (2004). This flexible standard is reminiscent of the U.S. Supreme Court's context-specific approach to procedural due process, as articulated in Mathews v. Eldridge, 424 U.S. 319, 339 (1976).

[153] *See generally supra* Part I and Part II.

[154] *See supra* Section III.B.

[155] In fact, this is precisely why the question of explainable AI is one of the most debated aspects of AI. For an overview, see Rudresh Dwivedi et al., *Explainable AI (XAI): Core Ideas, Techniques, and Solutions*, 55 ACM Comput. Surv. 194:1, 194:5 (2023); Sheikh Rabiul Islam et al., *Explainable Artificial Intelligence Approaches: A Survey*, arXiv (2021), http://arxiv.org/abs/2101.09429 [https://perma.cc/KC49-EZCX].

2. Evaluating AI's Capacity for Giving Reasons—Explaining the Black Box

Analyzing AI's capacity for explaining its judicial decision-making process requires both, assessing the current state of AI technology in terms of its ability to provide transparent and comprehensible explanations for its outcomes in general, and identifying any gaps between AI-generated explanations and the requirements for judicial reasoning specifically.

The opacity of AI decision-making processes and the corresponding notion of AI as an inscrutable "black box"[156] have been persistent concerns in discussions of AI decision-making. The characterization of AI as a "black box" is, however, not a monolithic concept but rather a spectrum of opacity levels dependent on the specific AI architecture employed. ML-based systems, particularly deep neural networks, present the greatest challenges to interpretability.[157] Nevertheless, it is crucial to note that opacity is not an inherent feature of all AI systems but rather a consequence of specific design choices.

XAI-techniques aim to render AI decision-making processes more transparent and interpretable to human observers, ranging from global interpretability (understanding the entire model)[158] to local interpretability (explaining specific decisions).[159] Some particularly promising XAI approaches include: a) Feature importance methods, which highlight the input features that influenced the AI's decision most;[160] b) Local interpretable model-agnostic

---

[156] Info. Comm'r Off. & Alan Turing Inst., *Explaining Decisions Made with AI*, 69 (2022), https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [https://perma.cc/8PJG-V2JV] (explaining that a black box in the context of AI is "any AI system whose inner workings and rationale are opaque or inaccessible to human understanding").

[157] Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 Big Data & Soc'y 1, 5 (2016); Thomas Wischmeyer, *Artificial Intelligence and Transparency: Opening the Black Box*, *in* Regulating Artificial Intelligence 75, 81 (Thomas Wischmeyer & Timo Rademacher eds., 2020).

[158] Lisa Käde & Stephanie von Maltzan, *Die Erklärbarkeit von Künstlicher Intelligenz (KI): Entmystifizierung der Black Box und Chancen für das Recht* [*The Explainability of Artificial Intelligence (AI): Demystifying the Black Box and Opportunities for the Law*], 36 Computer und Recht [CUR] 66, 68 (2020) (Ger.).

[159] *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 Duke L. & Tech. Rev. 18, 56-57 (2017) (noting that local interpretability uses a single dimension within a complex model).

[160] Zacharias et al., *supra* note 21, at 2161–62.

explanations (LIME), which provide explanations for individual predictions;[161] and c) Counterfactual explanations, which show how changing certain inputs would alter AI's decision.[162]

While these XAI techniques represent significant progress, they often involve trade-offs between accuracy and explainability,[163] and still fall short of providing comprehensive, contextual reasoning.

3. AI Explanations vs. Judicial Reasoning

Most importantly, though, the "ability to reason" attributed to AI systems—even those qualifying as XAI—fundamentally differs from that of humans. Given that judicial decision-making, including the reasoning process as manifested in judicial opinions, is oriented towards how humans think and act,[164] the fact that an AI system is interpretable should not lead to the immediate conclusion that it necessarily meets the requirements for judicial reasoning. Rather, the peculiarities of judicial opinions on one hand and those of explainable AI on the other must be juxtaposed to determine the extent to which the former finds sufficient correspondence in the latter.

3.1. *Active Reasoning vs. Passive Insight: A Structural Incompatibility?*

The reasoning of judicial opinions is an inherently active process, characterized by the judge's "self-explanation" through methodical preparation and articulation of decision-relevant aspects in natural language. This process stands in stark contrast to mere access to court files, which provides passive insight without elucidating the judge's decision-making rationale. The judge's obligation

---

[161] Muhammad Rehman Zafar & Naimul Khan, *Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability*, 3 Mach. Learning and Knowledge Extraction 525, 525 (2021); Xingyu Zhao et al., *BayLIME: Bayesian Local Interpretable Model-Agnostic Explanations*, UAI2021 Proc. Mach. Learning Rsch. 887, 887-88 (2021).

[162] Yingqiang Ge et al., *Counterfactual Evaluation for Explainable AI*, 1-2, (arXiv, Working Paper No. 2109.01962v1, 2021), https://arxiv.org/abs/2109.01962 [https://perma.cc/2UXU-SXTA]; Suffian et al., *supra* note 21.

[163] *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. Davis L. Rev. 653, 692-93 (2017) (asserting that simpler models are generally less accurate but more explainable, while more complex models are more accurate but less explainable).

[164] *See infra* Section III.C.

to provide reasoning is a fundamental aspect of due process, not a burden to be shifted to the parties or the legal community at large.[165]

Many XAI techniques, while enhancing interpretability, fall short of replicating this active reasoning process—an aspect which is, surprisingly, hardly addressed in the scholarship on XAI. The often-proposed solution of "code disclosure" or examination of training data proves insufficient.[166] Similarly, surrogate models, which use an external secondary model to approximate the decision-making AI, fail to meet the standard of judicial reasoning despite their active generation of explanations.[167] These approaches provide, at best, a passive insight into the AI's decision-making process, rather than the active, self-explanatory reasoning required of judges. The characterization of judicial reasoning as an active process therefore significantly narrows the field of potentially suitable AI systems to replicate judicial reasoning.

NLP systems show promise in actively generating explanations akin to the reasoning in judicial opinions. These systems, designed to "respond" to queries by generating text, bear some resemblance to the process of judicial explanation.[168] Additionally, certain XAI approaches, such as feature importance explanations, demonstrate an active orientation in their ability to elucidate decision-making

[165] The U.S. Supreme Court has consistently emphasized the importance of reasoned judgments, the extent to which reasons are to be provided by the judge, depends, however, on the nature of the case in question. As the Court stressed in *Rita v. United States*, 551 U.S. 338, 357–58 (2007), while the sentencing judge must state their reasons for a sentencing decision, the statement need not be particularly long or detailed.

[166] *See* Rita Matulionyte & Ambreen Hanif, *A Call for More Explainable AI in Law Enforcement* (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3974243 [https://dx.doi.org/10.2139/ssrn.3974243] (noting that trade secret protections prevent unlimited access to models' training data).

[167] *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085, 1113 (2018) (noting that approximating one model with a separate, external model becomes ineffective when replicating complex phenomena).

[168] Yogesh K. Dwivedi et al., *Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research*, practice and policy, 71 Int'l J. Info. Mgmt. 1, 3 (2023), https://www.sciencedirect.com/science/article/pii/S0268401223000233 [https://doi.org/10.1016/j.ijinfomgt.2023.102642]; S. Sean Tu, Amy Cyphert & Samuel J. Perl, *Artificial Intelligence: Legal Reasoning, Legal Research and Legal Writing Practicing Law in the Age of AI*, 25 Minn. J.L. Sci. & Tech. 105, 122–23 (2024).

factors.[169] However, their current capabilities represent only a rudimentary approximation of the complex reasoning required in judicial decision-making.

3.2. *General Form and Substance of Judicial Reasoning*

Apart from the necessity that judicial reasoning be actively generated and communicated by the deciding entity, there are, as elaborated above, further requirements for how the reasoning is to be designed.

Judicial opinions must be articulated in natural language, precluding the use of mere percentages, feature weightings, or complex mathematical formulas as adequate explanations, as they were determined by the AI system in the process of deciding the specific case.[170] This requirement stems from the fundamental principle that judicial decisions must be comprehensible to their addressees, including both the parties of the case and the legal community as a whole.[171]

Beyond using natural language, the content of judicial reasoning must further be comprehensible to the average legal addressee, a standard that precludes explanations requiring advanced technological expertise to interpret.[172] While the specific level of complexity may vary depending on the nature of the case and the parties involved, there are limits as to how technically sophisticated a judicial explanation may be while still fulfilling its purpose.

These two basic requirements regarding the form and substance of judicial reasoning thus further narrow the circle of potentially suitable AI systems for replicating it. However, they do not seem to structurally oppose AI fulfilling the required profile of judicial reasoning. Although many AI systems indeed express their result in the form of a percentage, and XAI techniques sometimes make AI

---

[169] *See* Zachary C. Lipton, *The Mythos of Model Interpretability*, 61 Commc'ns ACM 36, 40–42 (2018) (explaining how the post hoc approach targets factors that influenced decision-making without analyzing the greater intricacies of a given model).

[170] For an example of the implications of AI outputs which can detach themselves from human reasoning in the field of medicine, see Markus Herrmann et al., *Percentages and Reasons: AI Explainability and Ultimate Human Responsibility within the Medical Field*, 26 Ethics Info. Tech. 26 (2024).

[171] *See supra* Section III.C.

[172] XAI approaches are, however, mostly focused on making the AI tool in question explainable to experts in the field, not end users. For an example of how to make sure that the explanation offered by an AI system fits the needs of any type of user, see Vijay Arya et al., *One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques*, arXiv (2019), http://arxiv.org/abs/1909.03012 [https://perma.cc/3HSZ-BL8T].

systems explainable by enabling them to identify and express in mathematical form the weighting of individual decision-relevant features, NLP in particular represents a type of AI system that is structurally able to produce texts corresponding to these basic requirements for reasoning in judicial opinions. As mentioned throughout this article, LLMs have, in fact, become so good at generating text superficially resembling judicial opinions that even legal experts may have trouble telling them apart.[173]

3.3. *The Inseparability of Judicial Decision and Judicial Reasoning?*

Moving beyond these basic requirements for form and content of judicial opinions, I turn to the question of whether and to what extent an AI system is capable of fulfilling the special requirements for structure and content of the reasoning in judicial opinions.

This brings us back to the intrinsic link between the two layers of the judicial decision: Judicial reasoning as expressed in a judicial opinion is not merely an ex-post explanation of the decision, but an integral dimension of the decision-making process itself. Therefore, judicial reasoning cannot be detached from the decision it purports to explain, it rather builds on it.[174]

The analysis in Part I and Part II has revealed significant technological deficiencies in AI's ability to fully replicate the judicial decision-making process, both in terms of fact-finding and legal analysis. These limitations stem from the inherent "otherness" of AI compared to human cognition and the anthropocentric nature of judicial decision-making.[175]

This intrinsic difference raises doubts about whether AI's approach to judicial decision-making and its results can be properly characterized as legal decision-making at all. As I conclude above, current AI technology cannot engage in legal decision-making, as it is understood to be the core judicial task. Therefore, it

[173] For a recent review of what LLMs are capable of generating when it comes to legal texts, including judicial opinions, see Joshua Krook et al., *Large Language Models (LLMs) for Legal Advice: A Scoping Review* (2024), https://papers.ssrn.com/abstract=4976189 [https://perma.cc/7RUH-PAH3].

[174] *See supra* Section III.B. and Section III.C.

[175] On the core manifestation of this "otherness" of AI, see, e.g., Kerr and Mathen, *supra* note 9, at 23 ("[T]he very activity of judging requires following rules", and AI cannot be "a true participant in [...] rule-following"), 25 (arguing that AI lacks a normative commitment to the rules, or to the lived experience that makes the rules significant).

necessarily follows that AI cannot reason legally, either. In other words: no judicial reasoning and no judicial opinion without a judicial decision.

The fact that a judicial opinion is not merely a presentational format but is intrinsically linked to the cognitive process of judicial decision-making becomes particularly clear when zooming in on the three main categories of reasoning in judicial decisions, namely fact-finding, evaluation of evidence, and legal assessment.[176] As carved out in this article, each component requires specific cognitive skills and legal knowledge that current AI systems struggle to replicate. Any attempt to examine AI's capability to merely replicate these specialized aspects of judicial reasoning by producing a text *sounding like* an element of a judicial opinion quickly reveals the artificiality of separating the production of a judicial decision from its presentation. The essence of judicial opinions, characterized by their unique composition of fact-finding, evidentiary evaluation, and legal analysis, cannot be reproduced by focusing solely on the presentational aspect while ignoring the underlying cognitive processes of production.

When reducing these three components of judicial opinions to the presentational dimension, they at once become indistinguishable from one another; they are "mere" AI-generated text pieces, a chain of words built on the basis of statistical analyses. This homogeneity starkly contrasts with the cognitively distinct processes that a judge employs in arriving at each component. The qualification of a textual segment as fact-finding, evidentiary evaluation, or legal assessment stems not from its mere presentation, even in natural language, but from the judge's application of specific methodological guidelines when arriving at these conclusions.

If we entertain the counterfactual premise that AI could master both the processes of finding and establishing the facts of a case as well as their legal assessment, the presentational aspect as such—facilitated by NLP applications—would indeed be feasible. This hypothetical scenario would merely involve translating pre-determined decision steps and valuations into textual form.

The realization of AI-based judicial reasoning in judicial opinions, however, does not only require linguistic proficiency but also the fulfillment of myriad social and legal competencies currently beyond AI's reach. The replication of

[176] *See supra* Section III.D.

judicial reasoning by AI ultimately falters due to the absence of those cognitive capabilities in AI systems that are essential to the judicial decision-making process, as elucidated in the preceding parts of this article.[177]

F. Conclusion

This analysis concludes that the comprehensive transfer of judicial reasoning to AI systems is currently infeasible. While AI, particularly NLP systems, may generate text that superficially resembles the reasoning of judicial opinions, it fundamentally lacks the cognitive capabilities required for the judicial decision-making that judicial opinions are meant to justify. These deficiencies of AI identified in Part I and Part II regarding the social and legal skills intrinsic to the judicial role, including the ability to weigh complex social factors, apply legal principles contextually, and exercise discretion in a manner consistent with the rule of law, therefore have direct implications for AI's ability to write legal opinions.

Any attempt to separate the "production" and the "presentation" of judicial decision-making reveals a fundamental misconception about this very process. The essence of judicial reasoning lies not in its presentational form but in the underlying cognitive and evaluative processes the judge must undertake. A mere textual output that resembles a judicial opinion does not constitute judicial reasoning if it lacks the foundational judicial decision-making process.

IV
Concluding Thoughts: Moving Beyond the Human-Centric Concept of Judging?

This article has undertaken a comprehensive analysis of the prevailing conception of the judicial role, specifically examining the extent to which AI can fulfill the core requirements for serving as a judge given the current state of technology. The analysis revealed that while AI and human decision-making processes may show superficial parallels, fundamental differences persist upon closer examination.

Although current legal frameworks, while implicitly assuming a human judge, do not explicitly categorically preclude non-human entities from judicial roles, the

[177] *See generally supra* Part I and Part II.

article found significant deficiencies in AI's ability to replicate judicial capacities resulting from these fundamental differences in how it makes decisions.

A. The Implications of the Findings

Given these grave deficiencies of AI, which became apparent rather early on in this article, one might question the necessity of such an extensive and detailed examination of the different facets of the judicial role. However, this chosen comprehensive approach serves several crucial purposes which are often overlooked in the debate surrounding AI judges:

Firstly, by adopting a holistic approach, this analysis not only assesses the current feasibility of AI judges but also lays the normative groundwork for future discussions on integrating AI in the judiciary, the adaptation of legal frameworks to technological advancements, and the fundamental nature of law and adjudication in an increasingly digital world. While the analysis concludes that AI currently cannot replace human judges entirely, it provides a nuanced foundation for considering how AI may, after all, supplement or support judicial activities in specific areas. By breaking down the judicial role into its constituent parts and examining AI's capabilities in relation to each of them, this article opens avenues for targeted integration of specific AI applications in areas where it can enhance judicial efficiency without compromising the essential human elements of judging.[178]

By thoroughly examining the existing legal concept of the judge and judicial decision-making in light of technological advancements, this article also aims to discourage hasty creation of overly specific regulations for emerging technologies. The risk of premature legislation is particularly acute in the rapidly evolving field of AI, where today's cutting-edge technology may be outdated tomorrow.[179] The analysis suggests through its capability-based rather than entity-based focus that existing legal principles constituting our current understanding of judges may, in fact, be more adaptable to technological change than initially apparent. Against

[178] It goes without saying, though, that AI tools being used by judges in the course of their decision-making process also raise many questions. As a matter of fact, keeping a human (judge) in the loop may add additional challenges beyond the scope of this article. For a comprehensive assessment of human in the loop-scenarios, see Crootof, Kaminski, and Price, *supra* note 24.

[179] *See, e.g.*, Ramak Molavi Vasse'i, *The AI Act—The Epitome of Outdated Tech Governance—Exploring the Need for Innovative Regulation and Pathways to Modern Tech Governance*, 25 Comput. L. Rev. Int'l 72 (2024).

this backdrop, this article argues for a measured, foundational approach to legal adaptation rather than reactive, technology-specific laws.

Furthermore, the detailed comparison between human and AI capabilities does not only highlight AI's deficiencies but also points to the limits of human performance in judicial roles. Adopting this balanced perspective, the article underscores that while humans currently outperform AI in meeting the requirements for judges, human decision-making is not without flaws, either. By explicitly acknowledging these human limitations, it showcases the importance of revisiting old debates on the nature of judgment and utilizing them when assessing the potential of AI in judicial decision-making.[180]

Throughout the article, the topic of having "understanding" for the process of judicial decision-making as well as all its facets has proven to be one of the fundamental issues when it comes to determining whether AI may serve as a judge. The core question which has popped up in various shapes and forms is whether it is sufficient for AI to produce results that humans deem correct or explanations that seem comprehensible to humans, or whether we require "more" of AI. AI's lack of a deeper understanding as is ascribed to humans pushes us to dig deep and debate

[180] One example for such an old debate revisited in light of AI concerns the implications of subjective experiences, perceptions, values, and cultural conditioning of human judges on their decision-making process and whether AI could help to shield judges and their decisions from them. Proponents argue that AI could serve as a panacea for human weaknesses. They contend that AI, lacking human-like self-interests, physiological needs, emotional biases, or political leanings, could ensure more uniform and arbitrary-free application of general norms. *See, e.g.*, Robert Buckland, *AI, Judges and Judgment: Setting the Scene* 6–7 (Harvard Kennedy Sch. M-RCBG Assoc., Working Paper, Series No. 220, 2023), https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/working.papers/Final_AWP_220.pdf [https://perma.cc/WDB6-5S5U]. This way of arguing is, of course, closely connected to the premise of legal realistic concepts of judicial decision-making. *See, e.g.*, Dan Priel, *Law Is What the Judge Had for Breakfast: A Brief History of an Unpalatable Idea*, 68 Buff. L. Rev. 899 (2020); Manotar Tampubolon, Tomson Situmeang & Paltiada Saragih, *Judicial Breakfast as an External Factor in Judicial Decision Making in Courts*, 12 F1000Rsch. 9 (2023), https://doi.org/10.12688/f1000research.126482.1. Critics, in contrast, emphasize AI's fundamental dependency on human input. They argue that AI systems, being human-created products, are intrinsically dependent on human decisions regarding hardware, software, types of algorithms, and training data. This dependency, they claim, precludes AI from making truly independent decisions. For a critical view with regard to using these types of arguments to justify using AI instead of a human judge, see Konstantin Chatziathanasiou, *Beware the Lure of Narratives: "Hungry Judges" Should Not Motivate the Use of "Artificial Intelligence" in Law*, 23 Ger. L.J. 452 (2022); Shai Dothan & Gregor Maucec, *It Is Our Flaws That Make Us Humane: How Technology Ruined Judicial Craft*, SMU Sci. & Tech. L. Rev. (forthcoming 2025), https://ssrn.com/abstract=5191049 [https://perma.cc/YHF5-GQS9].

what it really is that we value the most in judicial decision-making: Is it merely the outcome, or do we place intrinsic value on the underlying process of reasoning, informed by a deep understanding of the world of an entity which knows what it is like to *be (human)*? This article has argued for the latter.

Finally, a key insight from the analysis of this article is that human weaknesses and their potential manifestation in the course of judicial decision-making are seemingly perceived as more manageable with regard to our current concept of judging than AI's deficiencies.[181] This is not because humans are inherently "better" than AI, but because human errors are more predictable, understandable, and thus less alarming to us (i.e., other humans) due to our shared human nature.[182] The familiarity, however, is a double-edged sword. While it makes human errors more palatable, it may also blind us to systematic biases or flaws in human reasoning. AI, while not being the appropriate tool to compensate for these human shortcomings with regard to judicial decision-making, offers a novel way of critically reflecting on them.

B. Reconceptualizing Law and the Judicial Function in Light of AI? A Paradigm Shift

I want to conclude by framing what I hope to be the core take-away from this article in a seemingly paradoxical way: On the one hand, it has become clear that AI is currently not capable of replicating any of the core facets, let alone the process of judicial decision-making as a whole, in any meaningful way, and there are good reasons to assume that this will not change in the foreseeable future. On the other hand, this fundamental structural incompatibility of judicial decision-making and

[181] This became apparent with regard to multiple aspects of judicial decision-making throughout the paper. Human judges struggle, for example, when it comes to accurately determining which emotions a witness is feeling during the hearing and to correctly derive from their demeanor whether they are telling the truth. *See supra* Section I.D, note 72. Furthermore, human judges are "black boxes", too. What happens inside a human judge, how they actually make decisions and for which reasons, is not accessible for anyone besides themselves. *See, e.g.*, Aziz Z. Huq, *A Right to a Human Decision*, 106 Va. L. Rev. 611, 640–646 (2020). Nevertheless, we have found ways in our legal system to handle these human weaknesses and safeguard judicial decision-making against any negative impacts, particularly through the requirement of giving reasons. Some of these human weaknesses may even be of particular value in the context of judicial decision-making. *See* Dothan & Maucec, *supra* note 182, at 21–30.

[182] *See, e.g.*, Crootof, Kaminski, and Price, *supra* note 24, at 467 ("Whatever one's thoughts about the opacity of the human mind, society has developed ways of querying decisionmakers and identifying reasoning errors.").

AI is not the result of humans being "better" than AI in absolute terms or legal systems putting human actors on some sort of pedestal as an end in itself. Rather, it "merely" follows from our current conception of the judge and its decision-making process.

Against this backdrop, future research could therefore explore reconceptualizing the judicial role based on AI's capabilities instead of assessing AI's ability to meet current requirements for judges, as this article did. This alternative approach offers two advantages: The first one is alignment with technological realities. Inverting the analytical framework avoids forcing AI into a role designed around human capabilities, instead acknowledging the fundamental differences between human and AI functioning. Such a perspective could, in turn, lead to novel judicial structures that leverage the strengths of both human and AI.

The second one builds upon the contrast between legal flexibility and technological constraints: While AI may face insurmountable technological limitations in replicating certain aspects of the judicial function, as conceptualized by current legal frameworks, these legal provisions are inherently flexible and can be modified to accommodate technological advancements. This realization opens up new possibilities for reimagining the role of judges and the nature of legal reasoning in an AI-augmented future.

Ultimately, the question of what function law and judges should serve in society remains a decision to be made by humans. Any potential reconceptualization of the judicial role to accommodate AI must therefore be grounded in societal consensus about the purpose of law and adjudication in our evolving technological landscape. This necessitates an ongoing dialogue between legal scholars, technologists, policymakers, and the public to ensure that our legal institutions evolve in a manner that preserves core democratic values while harnessing the potential of new technologies. As we stand at the cusp of a potential paradigm shift in legal reasoning and the judicial function, it is imperative that we approach these changes with a nuanced understanding of both the capabilities and limitations of AI as well as humans, always keeping in mind the fundamental principles that underpin our legal system—even if we do so only to make the decision to change it.