

RESEARCH ARTICLE

Trustworthy AI in courts: A legitimacy analysis

Andreas Follesdal^{1,*}  and Malcolm Langford² 

¹Norwegian Centre for Human Rights, Faculty of Law, University of Oslo; Norwegian Centre for Trustworthy AI (TRUST), Norway and ²Department of Public and International Law, Faculty of Law, University of Oslo; Norwegian Centre for Trustworthy AI (TRUST), Norway

Corresponding author: Malcolm Langford; Email: malcolm.langford@jus.uio.no

(Received 11 February 2026 UTC; revised 8 May 2026 UTC; accepted 26 May 2026 UTC)

Abstract

The use of AI in courts is highly contested, with positions ranging from outright opposition to unmitigated enthusiasm. However, normative assessments of its deployment often rely on standards or analogies that are unclear in their pedigree, coherence and scope conditions. This article explores how theories of legitimate authority can offer some guidance. Legitimacy theories serve to assess institutions and how they use their authority. The framework allows us to identify central issues and dilemmas, order different normative standards and identify relevant differences and similarities across time and in jurisdictions. To undertake such a legitimacy analysis of AI in courts, we identify six forms of delegation to AI in courts (servant, interlocutor, advisor, partner, agent and trustee) and analyse the legitimacy implications for six aspects of adjudication (legal process, inventorising, problem-solving, reasoned judgment, legitimation and discretion). In some areas, AI may enhance both the efficiency and quality of the judicial process, while we argue that the delegation of judicial power to AI in aspects which require extensive human discretion and human presence is difficult to justify.

Keywords: AI; courts; legitimacy theory; explainable AI; trustworthy AI; delegation theory

1. Introduction

The use of AI systems in the judicial system has been long met with ambivalence, if not contestation (Laux et al., 2024; Sourdin, 2018). Case management and housekeeping tasks are often welcomed but scholarly disagreements grow as AI moves from predictive to generative, with promises or threats to augment or even replace judges' central tasks (Fikfak & Helfer, 2025, p. 73). Thus, some hail how AI can reduce arbitrariness, legal costs and excessive delay (Capasso, 2024; Piccolo et al., 2023; Re & Solow-Niederman, 2019, p. 255; Susskind & Susskind, 2023), while others worry that it will weaken the quality, transparency and authority of adjudication (Committee of Ministers, 2020; European Parliament, 2023, Recital 61; Schäferling, 2023).

How should we assess such criticisms? Which concerns merit attention? One challenge is that the underlying principles or values vary among observers, yielding a plethora of plausible normative standards. Initially, scholarly focus was on how to ensure that judicial use of AI would comply with specific *rule of law* principles, especially the giving of reasons (Deeks, 2019; Tasioulas, 2023) and nondiscrimination (Wachter et al., 2020). More recently, several scholars have appealed to single, broader

*We are grateful for discussions of an earlier version at the Philosophy Department, Malaga University, December 17 2025.

concepts such as *trustworthy AI* (Jiawei & Moreira, 2023; Kaur et al., 2022), *responsible AI* (Fonseca, 2025), *assurance* (Leslie et al., 2022) and *accountability* (Williams et al., 2022), or unordered clusters such as “*equality, justice and democracy*” (Crawford, 2021). This reflects a broader trend within AI scholarship (Li et al., 2023), in which appeals to a singular overarching concept allow simplicity and breadth. However, they sacrifice precision and structure, as more detailed discussions soon reveal numerous desiderata beneath the surface, such as *accuracy, explainability, fairness, alignment, safety or sustainability*.

The result is that many insightful contributions focus on stand-alone norms or overarching values untethered to any broader normative theory – as Laux et al. (2024) find in their overview of 38 normative-oriented papers on trustworthy AI. To be sure: broader, well-structured and theoretically comprehensive underpinnings may often be unnecessary, but the plethora of standards and ideals are bound to conflict among themselves. This conceptual challenge becomes more acute as these standards are brought to bear on institutions, rules and policies and must be “balanced” in practice. In this respect, the floodgates are now open, as courts and other bodies across the world issue a flurry of specific, if not precise, standards on judicial use of AI (Dixon et al., 2025; CEPEJ, 2018). This is part of the broader and rapidly evolving biosphere of regulation (and deregulation) of AI.

A rapidly developing multifunctional and multivalent technology, AI promises both benefits and burdens, known and unknown (Bengio et al., 2024). How to assess AI in courts, as part of its overall governance, is therefore not clear. The vast new opportunities and risks stifle both our immediate normative intuitions and our more established standards developed for adjacent fields such as public policy or constitutions. How might multiple normative standards best be harmonised, and how should policies best seek to “balance” or “trade off” the various concerns? For instance, the UN Special Rapporteur on the independence of judges and lawyers has called on judiciaries to “develop and adopt guidelines on the use of AI” that permit its benefits but tightly regulate its downsides (Satterthwaite, 2025). Alas, the Special Rapporteur offers less counsel on how to perform this balancing act.

This same dilemma arises in the many, often conflicting appeals to *analogies*. For instance, “digital constitutionalism” seeks to adapt the “values of contemporary constitutionalism to the digital society” to enhance fundamental rights while protecting them from digitalisation – and ensure a “balancing of powers in the constitutional ecosystem” (Celeste, 2019). Indeed, the challenge of such analogies is illustrated by the plethora of varieties of digital constitutionalism: liberal, societal and global (de Abreu Duarte et al., 2023, p. 134; Terzis, 2024). These raise questions as to which constitutional values should apply to AI in courts, why and how? Other analogies point to existing delegations of judicial powers to other agents – the “unseen actors” in courts (Baetens, 2019) – such as juries or judicial administration. For instance, Stern (2025) observes that the concerns about the opaque black box of AI may be overdrawn, since we often endorse the jury system, which provides no judicial reasoning for the outputs. Are such analogies and similarities relevant? And if so, which aspects – and why? Assessments require more attention to the often implicit, normative frameworks which may – or may not – lay worries about AI to rest.

In this essay, we explore how theories of *legitimate authority* may offer some guidance for these challenges of ambiguity and help integrate various normative standards and draw lessons from appropriate analogies. We ask: *Why and when is judicial legitimacy at stake by delegating aspects of their power to AI?* Theories about legitimate authority are appropriate not least because their central topic is assessment of *institutions* and their legal powers, rather than norms and values for individuals and their actions. Several such theories are agnostic about whether decisions are made by an individual human or by “nonhuman” actors such as institutions. Thus, applying theories of legitimate authority to nonhuman AI may be a matter of further extrapolation rather than a “category mistake.” A focus on legitimate authority is also *scalable* in that it can apply to particular institutions or to the social and political “basic structure” (Rawls, 1971) – indeed a “global basic structure” (Follesdal, 2011) –

as a whole. We shall also suggest that a concern for justifiable allocation and use of authority allows an *ordering* of different standards, often identified by complaints about illegitimacy grounded in particular standards and features; and we allow that an institution may be *more or less* legitimate. Thus, various standards and values referred to in claims in favour of and against AI in courts provide a helpful departure point for defending or challenging proposed norms and values, as well as reform proposals. Our focus on the *legitimate roles of AI in the judiciary* allows us to see why these different standards and values matter, including why and when accountability, trustworthiness and discretion are of value.

Such a framework may not only bring normative clarity, but it can also help guide the development of ethical and legal standards for the judiciary and assist our shared reflections to develop new legal norms that must adapt to an evolving technology and unexpected human–AI interactions. Moreover, grounding normative analysis in theories of legitimacy also underscores that widely shared assessments about legitimate authority are crucial for sufficiently stable and broad support for institutions and social orders. This is especially so for courts, which enjoy neither the sword nor the purse (Hamilton, 1961). International courts are especially vulnerable as their jurisdiction is based on state consent and must largely rely on states’ willing compliance due to their lack of independent enforcement powers (Hillebrecht, 2014).

To be sure, other theoretical perspectives than “legitimacy” may offer several of these benefits, such as Tasioulas’s Aristotelian “humanist ethics” (Tasioulas, 2022) or Cohen and Liu’s “Rawlsian fairness” (Cohen & Liu, 2025). Our aim here is not to challenge these or to determine where they differ, if at all, but rather to explore this alternative of legitimacy, which focuses especially on assessing allocation of *authority*. We hold this issue central to some of the “high risk” areas of AI in courts.

This article is divided into three parts. We start in Section 2 by asking when and on what grounds court and other adjudicatory bodies [henceforth “courts”] make justified claims to be legitimate authorities – i.e., that their subjects have reasons to defer to their judgments and interpretations. In doing so, we bring Raz’s “service account” of legitimacy to bear on courts. The service they provide is to adjudicate impartially, based on legal rules and the use of legal method and structured processes, and to provide interpretations of those legal rules. We then examine how the delegation of aspects of this judicial service to AI can vary significantly in nature. In this respect, the core of the legitimacy issue in AI-supported adjudication is quite different from that, of say, the classical legitimacy questions about courts, such as judicial review. There, the crux of the issue is the *assumption* of judicial power but, in the case of AI, it is the reverse: it concerns various forms of *delegation* or even possibly the *abdication* of judicial power. The forms of this delegation – just like judicial review – can vary. We *thus* categorise and discuss six forms of delegation to AI – as a servant, interlocutor, advisor, partner, agent and trustee.

In Section 3, we set out the legitimacy framework for assessing such a delegation of judicial power to AI. We explicate how courts’ legitimacy depends on their inputs, processes and outputs, as well as their capacity to generate sociological/social legitimacy. This is complemented by an analysis of how conflicts between these elements can be resolved – for example, when AI might be negative for process legitimacy but positive for output legitimacy.

Section 4 then applies this framework to six aspects of adjudication that may be affected by AI, but in different ways. We differ somewhat from other accounts that analyse different phases of litigation, as we focus on judicial cognitive processes as much as the chronological sweep of litigation. These six aspects are as follows: (i) legal process (e.g., case filing, allocation of judges); (ii) inventorising (identification of issues, facts and legal rules); (iii) problem-solving, including addressing second-order considerations; (iv) justification (explicit reasoning and authoritative interpretation); (v) legitimisation (demonstrating that trust is deserved, both regarding the specific decision and in the judicial system generally); and (vi) iterative use of discretion. We ask under what conditions various uses of AI can strengthen – or weaken – the claims to legitimate authority at the various stages, while noting that the legitimacy of AI in each phase may vary according to not only the nature of the delegation

to AI but also other vectors, such as the area of law (e.g., contract vs criminal law), the hierarchical location of court (first instance vs appeal) and degree of transparency in the AI model (proprietary vs open, private vs public).

Finally, in our concluding discussion, we explore the value added of such a legitimacy framework by expanding on some of our general findings. While AI may be of considerable help when it operates as a delegated judicial decision-support system (Hendrickx, 2025), other uses raise more serious concerns. Among the topics which merit more attention are when AI *de facto* or *de jure* diminishes or even replaces *judicial discretion* to navigate many of these adjudicatory phases, thereby weakening or even replacing human *control*, and the reasons to value *human presence* to justify the use of public power when AI may weaken human *responsibility*.

2. Legitimacy theory and courts

2.1 Principal services of courts

Why should individuals and organisations defer to the judgments of a court, including its general interpretation of legal rules? Raz's account of legitimate authority addresses this as a special case of the more general puzzle of authority: how can we justify that subjects should defer to commands, somewhat independent of the particular content of those commands, so that the commands preempt or exclude some of the subjects' other reasons for action (Raz, 1986, pp. 369–429)? Raz proposes that a body enjoys legitimate authority over a subject insofar as it enables the subject to better act as it has appropriate reasons to do. These reasons may be to promote its own short- or long-term interests and values, as well as to respect and promote the interests of others. The service an authority provides to a subject may, for instance, be due to it being swifter or wiser than the subject, less subject to bias or particular weaknesses, or better able to help subjects avoid coordination problems or include more relevant information about the interests of others. Thus, an authority may be legitimate for some subjects but not for others, and it may be a *more* legitimate authority than alternative arrangements. An authority may even issue some bad decisions while still giving subjects reason to act – as long as the subject is overall able to act somewhat better by deferring to the authority (Raz, 1986, p. 75, 1994, p. 214, 2006, p. 1014).

So, what service do courts provide their subjects? We propose that the core services or tasks of courts are, first, to provide impartial adjudication based on judicial method, including the use of legal reasoning and method, meaningful participation of parties and provision of appropriate remediation (Fuller & Winston, 1978; Sourdin & Zariski, 2018; Sturm, 1990). Indeed, many AI-inflected legitimacy criticisms against the judiciary focus on the right to a legal hearing “before a competent, independent and impartial tribunal” (Satterthwaite, 2025, p. 10). Second, beyond the parties to the hearing, courts' interpretation of legal rules often creates a jurisprudence – precedents which help stabilise expectations about the future behaviour of others (Sabel, 1997). Notably, arbitrators in private adjudication are not understood to provide this broader service, which is one reason their decisions are not made public (Sweet & Grisel, 2017).

Courts may, thus, be legitimate authorities in this sense insofar as they serve as agents of legislators or trustees of their subjects, in the broad sense, by providing these services the subjects have interests in. Courts can help individuals and organisations with *dispute resolution*. The parties avoid violent determination of conflicts by having the court's decision preempt or replace the losing parties' preferences for another outcome. The courts also help secure subjects' interest in *predictability*. Courts assist each subject to better figure out what the law will be in somewhat similar and yet different situations – and hence what other law-abiding actors will do. Again, the court's decision replaces some of the actors' other reasons for acting otherwise. Courts may also contribute to provide *assurance* that public authorities and private actors are subject to the rule of law. Together, these services give reasons for subjects to defer both to the *general* authority of courts and to the authority of *specific* judgments.

2.2 Legitimacy issues in delegation to AI

Legitimacy concerns with courts can arise with both the assumption and delegation of judicial power, but their nature is different. What is at stake for the current topic is a question of delegation. Does the use of AI take authority hitherto allocated to and exercised by the judge alone, such that the judges modify or relinquish their legitimate role and authority?

In some respects, this concern is akin to the debate on “hidden actors” in courts, e.g., secretariats and clerks. Hidden human actors often provide important judicial services, such as improved efficiency, institutional memory and domain expertise (Baetens, 2019). The roles of these hitherto understudied actors have motivated a stream of scholarship: Computational and qualitative methods have helped identify who has written judgments in different courts (see review in Lie & Langford, 2024). This research is often driven by a concern that such *de facto* delegation or shared authority with these hidden actors raises legitimacy issues, such as whether these hidden services are sufficiently transparent to affected parties, how they influence case selection and the quality of reasoning and whether they are sufficiently grounded in the informed free consent of parties and other branches of government (Pauwelyn & Pelc, 2022; Rosenthal & Yoon, 2010).

In the case of AI, we can think of this technology as yet another hidden actor. Judges draw on the services of AI, or even delegate to AI, to undertake different judicial tasks. How this delegation occurs in practice may vary. Before considering the legitimacy concerns in such an act let us clarify the nature of various forms of services AI may provide, including delegation in various senses.

AI is often discussed in the singular, but it is, in practice, a cornucopia of different technologies with different uses – from highly automated and autonomous goal-driven systems to those which are more informational, dialogical or specific in nature (Bengio et al., 2025; Palmiotti, 2024). A useful way to categorise these uses is through delegation theory. We can discern at least six possible delegated relationships between courts and AI, where AI poses different challenges, some more complex than others. These six should only be regarded as a heuristic typology which is neither mutually exclusive nor exhaustive. To the contrary, we note below added challenges when the relationships are unclear, multiple or contested. Moreover, we note that there is an anthropomorphic quality to some of these labels, raising the common risk that it ascribes them “moral agency” or blurs “ontological boundaries” (Salles et al., 2020, p. 93). However, several terms are common in delegation theory, and our intention is to communicate metaphorically the nature of the relationship, rather than any intrinsic human-like qualities – indeed, our final critique rests on fundamental epistemic and ontological differences. We illustrate these six forms of delegation in Table 1, which run along a spectrum of increasing delegated discretion.

First, AI may be deployed as a *servant*. It carries out tasks which the judicial *Master* could perform, to save the principal’s time, mental acuity, etc. Thus, it is simply automation of existing court processes regarding such things as filing of cases, checking deadlines and assigning cases. In classical AI technologies, such automation of rule-driven processes in courts and legal practice was often a central focus, although machine learning-based AI has been increasingly mobilised for such purposes (Ashley, 2017; Fonseca, 2025). In these AI-as-servant cases, AI has no authority over the judge, its master: the reasons AI provides do not displace the reasons the judge already has. Nonetheless, there may be the standard risks that arise from any form of AI, especially inaccuracy but also a routinisation of tasks that require greater human, legal or domain-specific competence (Pasquale, 2019).

Second, AI may be an *interlocutor* in the sense found in Socratic dialogues, serving either as the questioner or the responder. A judge may question AI in the Socratic process of “eliciting the truth by means of questions aimed at opening out what is already implicitly known, or at exposing the contradictions and muddles of an opponent’s position” (Blackburn, 2006). This dialogical approach leans on the knowledge capacities of large language models (LLMs) and their apparent goal of helping find solutions (Chalmers, 2025) and captures part of the idea of “Scientist AI,” advocated by some of the pioneers of deep learning. “Inspired by a platonic and idealized version of a scientist,”

Table 1. Forms of delegation to AI

Service of AI (ideal type)	Delegation theory (dyad)	Core principle	Delegation function	Risks
Servant	Servant-Master	Execution	Perform a task	Inaccuracy
				Routinisation
Socratic	Facilitator-Interlocutor	Dialogue	Hold informed conversation	Lack of dialogical competence
				Undesirable preference formation
				Manipulation or advice giving
Advice	Advisor-Advisee	Recommendation	Provide suggestions/nudges	Inaccurate information base
				Lack of creativity
				Anchoring/ Manipulation
Partnership	Peer-to-Peer	Equality	Reach an independent decision that is counted amongst others	Inaccuracy
				Lack of explainability
				Manipulation
Agency	Principal-Agent	Autonomy	Reach a goal	Moral hazard
				Information asymmetry
				Inaccuracy
Trusteeship	Beneficiary-Trustee	Interests	Avoid bad decisions	Paternalism/ Manipulation
				Information asymmetry
				Inaccuracy

Scientist AIs are “AI systems designed for understanding rather than pursuing goals” (Bengio et al., 2025, p. 4). A Scientist AI is thus “trained to provide explanations for events along with their estimated probability” as part of the scientific process, e.g., of hypothetico-deductive reasoning, including “first understanding or modeling the world and then making probabilistic inferences based on that knowledge” (Bengio et al., 2025, p. 5). Of course, there can be dangers with such Socratic dialogue, which have been discussed since the time of Plato. As a questioner, a judge may provide, intentionally or not, overly leading or manipulative questions that provide a false sense of confidence in the answers provided by AI. Alternatively, the dialogue may be captured over time by the AI interlocutor who asks leading questions of the judge.

Third, AI may be used as an *advisor*. It assists judges, who are AI’s *advisees*, by helping the judges reflect on their overarching interests/values, and their more specific preferences, especially in applying judicial methods and choosing amongst interpretive strategies to pursue. In this respect, this advice service involves no formal delegation of decision-making authority and may be closer to the

types of conversations judges may have with clerks, secretariats and other judges when they move beyond a Socratic dialogue and request direct advice. There is danger that AI can informally influence the adjudication process to a great extent and skew the deliberation by mechanisms such as framing and nudging (Chong & Druckman, 2007; Thaler & Sunstein, 2008). Thus, Schmauder et al. (2023, p. 799) argue that “blindly outsourcing the development and implementation of nudges to ‘black box’ AI systems means that the ultimate reasons for why such nudges work, including the underlying human cognitive processes that they harness, will often be unknown.” So there is a gray area where AI may serve both as Advisor and as Agent or Trustee (see below), and where it is difficult to determine which roles it serves – and how.

Fourth, AI may be a *partner* in a judicial process. Here, AI reaches a decision on some matter, and this counts formally in the decision-making. However, its decision is not decisive on its own – the decisions of others also count under some procedure of triangulation or deliberation. The ultimate example of such partnership in courts might be AI as a member of a panel, e.g., as one of three judges in a case. More prosaic examples might be in matters of litigation procedure. For example, machine learning models have been shown to perform well in estimating the days needed for a court hearing in the overwhelming majority of cases but struggle with outliers (Bentsen et al., 2026). A partial automated system to determine the number of days might, therefore, take the average of the human judge and AI’s estimation, weighting them equally or even differently for different types of cases. Such partnership approaches will often rest on conscious institutional design and may, in principle, optimise the respective advantages of human and machine decision-making. However, there remain considerable risks, including inaccuracy and lack of explainability, as well as a possibility for manipulation over time. Further important questions concern whether and how AI should acquire such joint authority to make the speech act of legally binding judgments (Austin, 1975).

Fifth, AI may be an *agent*. The delegation of authority by the judicial *principal* furthers the judge’s express and fixed preferences in the face of alternative strategies, some of which may be best uncovered by an agent in a manner that is efficient and enhances quality (Gailmard, 2013). If the principal is sufficiently dissatisfied with the agent, the principal may correct it in the particular case or replace it more permanently. Simple examples of dissatisfaction might be when an AI service hallucinates judicial precedents or translates relevant foreign sources incorrectly.

While AI as the agent for the judge is formally constrained (Stone Sweet & Brunell, 2013, p. 65), such delegation raises a familiar suite of principal-agent challenges. Does the agent have the methods or competence for the particular task (*competent agent problem*)? For instance, Latif argues that “Persistent reliance on AI outputs invites a growing dependence on algorithms that lack genuine comprehension of legal precedents, statutes, and contextual nuances” (Latif, 2025, p. 10). Moreover, will the agent act to further the expressed interests of the principal (*strong agent problem*)? Among several concerns, this raises the challenge of moral hazards. These often arise from a hidden misalignment of the interests of the principal and agent (Gailmard, 2013) and/or asymmetry of information between them, making it difficult for the principal to verify and monitor the actions of the agent. Thus, AI may provide a formal presentation in support of particular outcomes yet hide the particulars of the case that explain the choice of one particular output or decision – which might be appropriate but could also be due to a misalignment of values, principles and rules (Bai et al., 2022). Other strong agent challenges are information problems arising from the AI “black box” and lack of explanation (Miller, 2019) and lack of transparency in the underlying collection and treatment of data (Yu & Barter, 2024).

These challenges may be even more salient in AI models and systems with high levels of agency, such as LLMs and “agentic AI” (Bandi et al., 2025). The latter form of AI is “capable of interacting, reasoning, and adapting in real-time to achieve specific goals autonomously” (Hughes et al., 2025), which is now becoming publicly available (e.g., Claude, Codex) (Caballero, 2026). Thus, there are alignment attempts such as “Constitutional AI,” which seek to deeply embed rules and principles in AI models (Bai et al., 2022).

Finally, AI may assume the role of *trustee* for the judge. In such cases, AI acts to further the judge as *beneficiary*, promoting the judge’s overarching interests or values, such as proper judicial method, even if they are contrary to the judge’s express current preferences. Here, the beneficiary cannot easily revoke the delegated powers of the Trustee, especially not in the particular case (Alter, 2008; Hale, 2024). For such authority to be legitimate, as we shall argue, a necessary condition is that the trustee must somehow be better placed to regularly identify and act on the true interests and needs of the beneficiary – and be seen to reliably do so. There may certainly be scenarios for the justifiable use of AI as Trustee, for instance, to remove persistent bias or to reduce delays (Bentsen et al., 2026; Peer & Gamliel, 2013). Nonetheless, there is a risk of unjustified and unconstrained paternalism, common with all trustee models (Bellamy, 2008), which can amplify the problems of agency models identified above. Indeed, there may be an ultimate danger of “epistemic dominance” by AI in an extensive trusteeship model, which can undermine the “trust in one’s own epistemic capacities and hinder the exercise of those capacities” (Coeckelbergh, 2023, p. 1341).

On its face, AI seems more desirable and legitimacy-enhancing when AI contributes as a servant, interlocutor and advisor, and, to a certain extent, partner and agent, rather than as a trustee. The first forms of use are typically judicial decision-support systems that support judges in their decision-making while formally leaving the final decisions in the hands of the judge (Hendrickx, 2025). The contribution of AI becomes often more controversial and problematic when there is more *de facto* delegated discretion and authority to AI.

Three points merit mention. First, delegation of discretion and authority to an AI system may in practice occur across the spectrum from servant to trustee. Second, there may be a distinction between theory and practice for each of these six relationships, such that their actual deployment blurs the boundaries. This occurs in cases involving “hidden actors” in general and with regard to AI in particular. For example, consider when the secretariat, a human clerk or an LLM suggests a particular legal categorisation or analysis, and the judge agrees and adopts it without much attention to alternatives or details – any public statements and formal responsibility notwithstanding. The role of Advisor then blurs toward being an agent or even a trustee. What should matter for legitimacy analyses is neither the formal allocation of authority nor what the parties claim, but the *de facto* relationships in the ongoing judicial practices. Thus, more research is needed to address further the existence – and contestation – of such blurring of roles. Moreover, it can be important to determine which forms of a particular AI delegation are more likely to lead to such slippage: e.g., is a single nudge as an advisor more likely to lead to anchoring effects compared to a suggestion of different scenarios? For instance, various nudging and framing arrangements create legitimacy challenges also outside AI, in ways that may illuminate the new issues wrought by AI.

Third, and importantly for our purposes, if AI-assisted human judges, or AI acting alone, demonstrably come to outperform human judges acting on their own on some such tasks, such a service may come to have a stronger claim to authority than the human judges acting without AI. Such a scenario may seem unthinkable in many areas of judging; the rapid advances in AI counsel us to consider multiple scenarios. We regard it as strengths of this Razian account that, given that these scenarios are possible, the account indicates a fruitful way to address the issues and some of the conditions for such conclusions. It also allows for dynamic changes in the legitimate authority of such roles of AI over time, as AI and its tasks and services vis-à-vis human judges change.

Conclusions regarding each type of delegated role, and their various combinations, must take account of the particular types of legitimacy factors that emerge with its specific use and the phase of the adjudication cycle, to which we now turn in Sections 3 and 4. These factors may legitimate trustee approaches in certain contexts and delegitimate servant, interlocutor, advisor and agent approaches in others.

3. Undertaking legitimacy assessments

3.1 Legitimacy clusters

Criticisms of institutions couched in terms of illegitimacy often indicate the reason for the criticism. In the case of theories about legitimacy, the reasons are often clustered around the *input*, *process* or *output* aspects of exercises of authority (Follesdal, 2020, pp. 481–482; Langvatn & Squatrito, 2017; Peter, 2010; Scharpf, 1999). What unifies this tripartite list is that these elements are the main ways that institutions can fail in their tasks and hence fail to be legitimate authorities.

To put it more formally, those subject to institutional power may challenge:

- (1) Input legitimacy – the framework which establishes the relevant rule-making power: for instance, many criticisms of claims to authority are of the form that the authority is, in essence, undemocratic – those subject to the authority had little or no influence on the establishment, shape or selection of the relevant institutions or of the particular office holders.
- (2) Process legitimacy – how the outcomes were arrived at: for example, that agreed procedures, hearings, etc. were not followed, that the existing practices were biased or that judges used their powers inappropriately, even if originally grounded in consent.
- (3) Output legitimacy – the nature of the benefits from those authorities, be it that the results are meagre or unfairly distributed.

While these three aspects have a chronological character, tracking the origins and exercise of power, they should be understood as interacting yet distinct. Regardless of the field of application, these three types of legitimacy concerns may arise independently of each other, and – as we shall see – their different nature often results in questions over whether they are sufficiently commensurable for purposes of comparison and balancing.

Finally, normative legitimacy analysis must often include a fourth element: (4) *sociological assurance*. When an institution increases its power (e.g., judicial review to judges) or delegates it to another (e.g., a judge to a jury or AI), one added issue regarding legitimacy is whether there is public knowledge that others are likely to comply with that institution's decisions – e.g., for our topic, whether most subjects respect the rule of law and will obey the orders of a court. Widespread noncompliance with an institution's rulings may challenge the normative legitimacy of a body (Follesdal, 2020). This may be because some desired outcomes require general, if not universal, compliance, or because the unchecked presence of free riders lessens the incentive of others to comply or accept an institution's power, making the collaboration highly unstable. Assurance about others' compliance partly depends on sociological legitimacy (Gibson, 2007). The latter concerns whether "the governed" believe and accept that an institution has, or maintains, legitimate authority over them. In this Weberian sense, we may ask whether individuals affirm "a system of authority" and lend it "prestige," such that obedience may follow (Weber, 1964, p. 382).

This sociological assurance elucidates and justifies one sense of *trustworthiness* for the normative legitimacy of the judiciary. The social functions of peaceful adjudication and development of law require that the judiciary both (a) actually secures these core tasks regularly and most of the time – that it is worth the subjects' trust; and (b) that the general public can *determine* whether this is indeed the case (Laux et al., 2024). This combination is often crucial to induce general compliance among "contingent compliers" who are prepared to do their share in social settings as long as they believe that the objectives and rules are fair and they have reason to believe that others think likewise and comply (Follesdal, 2006; Goodin, 1992). Such complex public trust in the judiciary – including assurance that public authorities comply with the rule of law – is important to secure the collective good of nonviolent dispute settlement by legal methods.

We submit that these four elements often form the basis for legitimacy assessments, which in our case concerns the delegation of judicial power to AI models and systems. It is useful to parse each of these elements to identify the types of legitimacy issues that AI tends to expose.

Regarding *input legitimacy*, many criticisms of institutions question, for example, whether the institution and its mandate are set up in the proper way. In the case of courts, it might be whether adjudicators are appointed according to appropriate procedures and sufficiently impartial among the parties to a dispute, or whether the laws they interpret and apply are sufficiently fair – or rather perpetuate domination of particular genders, social classes, ideologies or states. Such requirements are often phrased as resting on the free and informed consent of subjects (Buchanan & Keohane, 2008). Such consent might be given directly in a contractual sense – e.g., consent to arbitration in a private law agreement. But often any such consent is at best hypothetical or implicit, e.g., contained in essence in constitutions or democratic based-law making. Input legitimacy can also be interpreted more broadly and integrate broader rule of law principles (Fuller, 1964). For international institutions, the central question is often whether states rather than non-state actors have consented to the delegation of power (Buchanan & Keohane, 2008).

In the case of AI, there may be concerns that the use of AI models – including data for them – contravenes existing laws that, for the sake of the argument, express and constitute, from a normative perspective, the best interests of the subjects. Indeed, administrative law in many countries requires some form of human control in decision-making, while instruments like the EU's GDPR, and to a large extent the AI Act, heavily constrain the use of automated decision-making in courts as well as reuse of data for new purposes. Use of AI by institutions surrounding the court can also raise broader input legitimacy issues, such as internal authorisation procedures. In the same way that there is an expectation that appointments of court secretariats and judicial clerks must meet some basic procedural and competence requirements (e.g., interviews for administrative officials: law degrees for clerks), there can be an expectation that delegation of judicial power to AI meets some basic administrative requirements.

A further set of legitimacy challenges for an institutional court relate to the *process* behind an act of authority. For instance, a central concern in the debate over the use of powers of judicial review to invalidate or direct legislation is whether courts lack the appropriate form of democratic legitimacy. They may be engaging in questions that should be determined by elected *representatives* (Waldron, 2006), broader *participatory* processes (Bellamy, 2008) or structured *deliberation* (Gargarella, 2022). In the case of AI, including its use in courts and decision-making in public administration, we tend to see these concerns molded together in broader concepts like “democratic governance” of the use of AI to ensure that it is trustworthy (EU, 2019). In practice, there has been a strong focus on participatory democracy, in which critics lament the lack of citizens’ and users’ (judges/caseworkers) ability to influence when and how AI is mobilised (Broomfield & Reutter, 2022; Holten Møller et al., 2020). This is now complemented by a growing emphasis on representative and deliberative democracy, given the potential far-reaching consequences of AI and the complexity of decisions concerning its deployment. Buhmann and Fieseler (2023) argue that responsible AI innovation “calls for public deliberation: well-informed ‘deep democratic’ debate that involves actors from the public, private, and civil society sectors in joint efforts to critically address the goals and means of AI” (Buhmann & Fieseler, 2023).

Besides the decisions to use (and continue to use) AI, there are a range of other process concerns. Typical aspects of adjudication that trigger normative inquiry often include issues such as transparency, impartiality and independence – which are necessary for the social functions of courts – but also broader AI issues such as cybersecurity and sustainability of energy use (Jegham et al., 2025; Kayode et al., 2025). For instance, transparency demands openness in methods – e.g., court hearings should be public, and arrangements should ensure that other actors than judges do not make the final decision (Baste et al., 2025).¹ On our view, this is one aspect of legitimacy which is at stake in

¹The European Court of Human Rights (ECtHR) has stated that it “protects litigants against the administration of justice in secret” and is “one of the means whereby confidence in the courts, superior and inferior, can be maintained.” *Fazliyski v. Bulgaria*, no. 40908/05, 16 April 2013, para. 64.

the “black box” problem in AI-supported decision-making, with inductive machine learning methods which have occupied much attention. Likewise, potential bias in datasets and misalignment in AI algorithms raise concern about the impartiality of judges, which could be seen as a process legitimacy flaw.

In relation to the giving of reasons, we do not see it as a central part of the process which leads to the delivering of judgments. We rather see this as an essential output of the decision itself – perhaps unlike some other authors (Tasioulas, 2022, p. 13). Reason-giving by courts both enables better predictions about the legal rules in future situations and provides the requisite assurance that justice is not only done but seen to be done. The use of an AI model as part of judicial reflection and consideration – e.g., AI in advisor modus – that does not provide sufficient explainability could be seen as raising a process concern. Such a situation can be analogised to a judge accepting information provided by a clerk who is unable to explain its source. Whether the information is reflected in the final judgment as reasons is irrelevant – the concern here is processes leading to a judgment that might influence a judge in a problematic manner.

A third cluster of legitimacy objections concerns the *output* of the institution: that the judges fail to deliver reasoned judgments, for instance, due to a failure to identify all relevant legal norms or precedents, because their interpretations are incorrect or inconsistent, or that the mill of justice grinds too slow. The justification for AI models is often such output legitimacy, with arguments that it can improve effectiveness as well as fairness, reducing bias and arbitrariness (Tasioulas, 2023). However, three concerns tend to arise.

The first is *accuracy* proper. While predictive and generative models are improving considerably in many fields (Deuschel et al., 2024), their performance in the field of law remains jagged and somewhat uneven – especially when there is insufficient data or doctrinal uncertainty and complexity (Siino et al., 2025). Moreover, it is challenging to quantify responsibly the resulting level of uncertainty and then qualify it in such a way that is actionable for use in legal decision-making (Marusich et al., 2023). The second is *fairness*. Underlying bias in the learning data for models (historical but also through reinforcement learning) may be reflected in the resulting algorithms. The empirical debate over whether COMPAS software for criminal sentencing in US courts incorporates biases against African American defendants is a case in point (Donohue, 2018; Kehl & Kessler, 2017). The third is *explainability*. While there is now an array of “explainable AI” (XAI) approaches, for more advanced AI models it remains difficult to know in what sense, if any, the reasons given were in fact those “used” in a “decision.” This is a longstanding challenge in understanding human judging (as the legal realism debate attests), where AI-supported judging based on machine learning rather than pre-programmed expert rules creates new challenges, for instance, regarding what the AI input is “about” (Cappelen & Dever, 2021a). Moreover, there is uncertainty over whether AI can take into account both past and future contexts, which are relevant for such reason-giving, and potential and second-order effects – a concern we address below.

Finally, the need for *sociological assurance* underscores the importance of communicating signals to citizens that the institution is legitimate and worthy of compliance – and is regarded sufficiently as such by many others. For courts, this can include forms of *transparency* about procedures, etc., to show that the judiciary is indeed a “competent, independent and impartial tribunal.”² It can also require that the judiciary is *accountable* in certain ways even though it is independent. It cannot be directly subject to control by the executive or the legislature – that would prevent its ability to provide the services. But it must be possible for outsiders to determine, for instance, whether the court actually applies judicial method to the appropriate legal sources in the particular case, or whether existing forms of automation exclude meaningful participation. This arguably contributes to foster *general compliance* with the judgments and consistent interpretations by the public and public knowledge

² Article 14 of the International Covenant on Civil and Political Rights (ICCPR) and Article 10 of the Universal Declaration of Human Rights (UDHR).

about both. Without stable expectations about such future compliance, the judiciary cannot provide the services of dispute resolution and authoritative interpretations.

Empirical research on citizen attitudes toward AI-based adjudication confirms the need for a sufficient focus on assurance. A consistent finding is that citizens' trust in judicial decision-making is lowered if there is no "human in the loop" (Laux et al., 2024). Moreover, we are prepared to accept some "human failures" more than "AI failures" (Longoni et al., 2022). However, there are several important nuances to this growing body of research. It appears that for public decision-making, we trust AI less than a human, but we can often trust a human who uses AI more than a human who doesn't use AI at all (Laux et al., 2024). Note that there are important variations. Sometimes, minority groups are more likely to trust machine-based decisions given previous patterns of discrimination, while citizens in countries with weaker rule of law are more likely to trust AI compared to human decision-making. Recent research in Europe also indicates that citizens are less concerned about AI if it speeds up decision-making and more concerned about the transparency of algorithms rather than whether it is publicly or privately owned (Bentsen & Johannesson, 2026). Moreover, the importance of explanation diminishes somewhat if there is an opportunity to appeal. This all suggests that the precise forms of delegation and the jurisdictional context will matter a lot, and that the perceived importance of a "human in the loop" may well change over time.

3.2 Cluster conflicts

Even though many agree on these elements, several questions are unanswered: is this just an eclectic collection of considerations or is there some unifying perspective that might also help in cases where the various elements need to be "balanced" or traded off against each other (Buchanan & Keohane, 2008)? For instance, can "input" legitimacy be sacrificed completely if "output" is excellent? For example, can more reliable and swifter AI-assisted decisions be more legitimate than AI-free judicial processes – even though they are less transparent? Or must each element meet a threshold for the institution to be legitimate?

Recall that we regard these clusters as the principal ways by which institutions can fail in their tasks and hence fail to be legitimate authorities (Follesdal, 2020). At least four issues deserve attention. The first is to be aware of the *fallacy of accumulation*. While these elements are interconnected, they are not simply cumulative: A complete lack of input legitimacy cannot always be compensated by better output legitimacy (Follesdal, 2017). In the case of AI, Li et al. (2023) make the same point: "AI trustworthiness cannot be achieved through hillclimbing on a set of disjoint criteria." To avoid this fallacy of accumulation or "hillclimbing," many argue that at least some threshold of each element must be met, and we would agree with this (Goodin, 2013; Follesdal, 2006; Taylor, 1987).

There is likewise disagreement over the "weighting" of each element. For some, the minimum or "gold" standard is process legitimacy, and in particular democracy (Peter, 2010, pp. 15–18) – meaning that, at the very least, the threshold for this aspect may be higher than for the others. In our context, this seems to be the assumption for many normative scholarly analyses of the use of AI in courts which emphasise respect for democratic principles and process, including inclusivity in design (Laux et al., 2024); and has been affirmed in some recent empirical studies (e.g., Magalhães et al., 2026). In contrast, the EU's (2019) definition of trustworthy AI is silent on this issue, although it does pay more attention to the rule of law. To be sure, it may be argued that a rigid process threshold, requiring fully fledged legitimate processes, is not needed in all circumstances. When normatively urgent substantive interests are at stake, we may also have reason to protect them against certain procedural concerns. Nonetheless, such sacrifices can have problematic long-term consequences, and, under these conditions, other standards of process *might* be imposed: e.g., transparency, due process, reason giving or consultation, or outputs that help enforce democracy (Bodansky, 1999).

The second is the challenge of *ontological commensurability*. The intrinsic qualities of the different aspects of legitimacy make it difficult to undertake proportional "balancing" for trade-offs. Even

when we can agree on the relative importance of each legitimacy cluster, it is challenging to create a single metric for a balancing exercise, a common challenge in the mixing of any method (Beach & Kaas, 2020). To illustrate: it is often not possible – or indeed meaningful – to say that an institutional design with 10 percent more output legitimacy is as legitimate as an institution with 10 percent more process legitimacy. In the case of AI, there have been attempts to develop metrics that permit these calculations for selected elements of trustworthiness. For example, Giudici et al. (2024) provide commensurable ranking metrics for robustness, accuracy, fairness and explainability in AI models which permit analysis of trade-off. However, the model excludes other important aspects of trustworthiness, such as democratic governance, privacy, safety and environmental sustainability. And issues of how to “weigh” each of these factors remain to be addressed. In practice, commensurability may be achieved through rough triangulation (e.g., through comparing simple ordinal-like units like “none,” “none+,” “small-,” “small,” “small+,” “some-,” “some,” “some+” and “large-” (Kern, 2018)), or some other incomplete quasi orderings (Sen, 2017); pragmatism (each cluster is considered on its “own terms” and put subsequently into a conversation with the other (Tait & Taylor, 2022)); or some form of democratic process based on representation, participation and/or deliberation. We here note that our legitimacy framework leaves these issues as open questions for further attention, including for which purposes any such single metric would be helpful.

The third is *tensions* between the clusters. Even if we could measure in a commensurable way each cluster, a feature of legitimacy analysis has been the inescapable dilemmas that arise in institutional design. There may be internal conflicts between and within a cluster that are difficult to navigate and require inescapable trade-offs. To illustrate, Dunoff and Pollack (2017) argue that there are trilemmas of institutional design for international courts which prevent them from being fully independent, transparent and accountable to states. In the case of AI, such trustworthy dilemmas have garnered increasing attention in the past 5–10 years (Li et al., 2023). For instance, improvements in explainability often come at the expense of accuracy, as deep learning models are controlled to ensure greater transparency and interpretability. Or improvements in privacy can come at the cost of nondiscrimination and fairness, as the masking of personal data removes the possibility to correct for bias. Thus, not all elements or clusters of legitimacy seem required in all cases. Rather, close scrutiny of the development of AI combined with systematic normative arguments is required to understand how well the various clusters of rules ensure justifiable institutions.

This makes it necessary to consider whether we can and should speak of more or less legitimacy in such contexts. The aim may be to improve on, satisfy or optimise legitimacy in a certain context without attempting to maximise legitimacy or meet a high threshold for each element. In our view, it is possible to speak of and discuss whether delegation of judicial power to AI enhances or diminishes legitimacy of adjudication, especially when a legitimacy assessment is carried out in a particular context. Given that a singular policy, rule or action can enhance one aspect of legitimacy yet disturb others, determining whether legitimacy is increased or decreased requires often a full assessment of the various aspects and how they fit together in the particular case. AI may thus often improve efficiency, reduce human bias/fallibility and improve access to public goods, but it encounters questions over a potential weakening of input, process and output legitimacy together with increased sociological uncertainty due to the potential for free riders. We should especially determine how much such conflicts matter from a normative legitimacy perspective. For instance, it may be less problematic if the various standards for input and process legitimacy must be satisfied to some threshold level rather than maximised. Thus, more attention may be appropriately directed to these inescapable trade-offs which require optimisation solutions (subject to thresholds) or hard consequence-informed choices.

The final challenge concerns *lack of output legitimacy*. For instance, a classical question is whether an institution can be legitimate if it does not generate sufficient benefits. A state may have impeccable democratic credentials but is it legitimate if its citizens are significantly and avoidably deprived? In the case of AI, one is often faced with the question as to whether AI *must* be deployed if it provides clear benefits, including for adjudication. It is reasonable to presume that some output standard of

justice and/or the common good must be met. One important question is what standard. In the context of the debate on judicial review, the debate on this question is complex. Some argue that court must at least ensure avoidance of “extreme injustice” (e.g., a small core of civil rights and minimum subsistence) (Buchanan & Keohane, 2008, p. 44), others claim that court must ensure more adequate outcomes (good governance, social trust, significant protection of rights and sufficient education) (Kane & Patapan, 2010, p. 605), yet others make thick demands for optimal outcomes or full rights realisation (Peter, 2010, pp. 13–14). In the case of AI, the technology is still too nascent to gain a sense of its full potential, but it is likely that the duty to use AI in courts could be advocated as it becomes ever more powerful and effective.

Two preliminary conclusions from this discussion are that the legitimacy of institutions depends on these four factors (input, process, output, assurance), but the assessment is not simply monotonic: neither more democratic control, more transparency and due process, nor more beneficial outputs will always increase the legitimacy of the institution. And legitimacy assessments often require a detailed understanding of the intricate details of rules and actual practices within the institution and its complex interconnection within a web of other institutions.

4. Adjudication phases and the roles of AI

We now apply this framework to different aspects of adjudication, assessing how particular uses of AI may affect the authority of courts, generally and in specific judgments. Scholarship on the topic often divides litigation in a chronological and tripartite manner, with a focus on process, decision-making and remediation (Fikfak & Helfer, 2025, p. 92), while others break down the decision-making elements into a larger number of phases (Kolkman et al., 2024). Our six-part taxonomy partly follows a chronological approach, while seeking to better capture the cognitive and sociological dynamics of judging: legal process, inventorising, problem-solving, reasoned judgment, legitimation and discretionary decision-making. Regardless of the taxonomy, it is important to recognise that these aspects interact in several ways – for instance, legal interpretation can affect both legal procedures and remedies, while procedural decisions frame and shape the legal and factual issues at stake.

4.1 Legal process

Legal process can include all aspects of a case that have a strong administrative dimension, whereby the steps taken have an indirect or direct bearing on the content of the substantive decision. This can include case filing, allocation of judges, holding of interim hearings, ordering of injunctions based on clear predetermined criteria,³ transcribing proceedings, interpretation and translation of foreign languages, translating evidence and other sources, and summarising and publishing judgments. In the legal process, AI can be deployed, for example, to extract important case details at various stages (Kolkman et al., 2024, p. 7) and cluster and/or analyse cases when they first enter the court system, suggesting an order of importance or urgency or recommending an assignment to the appropriate court, or assist with similar and repetitive cases that often overwhelm courts (Fikfak & Helfer, 2025, pp. 85, 99; Fonseca, 2025).

In several of these legal process areas, the use of AI tends to be regarded as largely “uncontroversial” for AI (Fikfak & Helfer, 2025, p. 93). During the last decade, AI has been adopted for such functions and, with the arrival of generative AI, its usage has accelerated (Langford, 2026a). In some instances, it can be argued that deployment of AI in legal processes contributes to significantly enhance process and output legitimacy. For example, in Norway, if AI were used to transcribe proceedings in courts (which currently does not already occur), it could improve the quality of judgment writing and public accountability. Likewise, the use of AI to de-identify personal data more quickly could allow the publication of most, rather than a small proportion of, judgments (Baste et al., 2024).

³Here the line between procedure and substantive is often difficult to discern (see, e.g., Capper, 1991).

However, there is also some dissent. Satterthwaite (2025) notes that LLMs are often poorly trained on small – so-called “low resource” – languages, which raise risks when AI is used even for seemingly simple tasks, from fact summaries to translation and transcription. Given that most litigation is fact-intensive, in that facts are often more determinative of the outcome than a legal interpretation, these risks should not be underemphasised. Small errors can be determinative, suggesting that highly automated trustee models of delegation without any human control – whether in innovation processes or in actual practice – can be problematic. Further challenges arise in highly automated systems that are not adaptive to changes in procedural law and the broader ecosystem of litigation. Procedural uses of AI tied to legal rather than administrative aspects of process require greater vigilance; and abrupt changes in nature and forms of claims also require reflexivity. Indeed, such sudden changes can or will be wrought by AI itself. For instance, several public administration agencies report that in certain areas of law there has been an exponential increase in the volume and length of complaints since generative AI was made publicly available, while LLM-generated deepfakes for court evidence are a growing problem (McConvey et al., 2026).

4.2 Inventorising

Ahsmann (2011) describes an intermediate process of decision-making as “inventorizing.” In this phase, the judge “reads the case file to determine what it is about, what has been claimed, and what legal questions are involved” and “determine the applicable rules and case law, list the relevant facts, and consider the propositions put forward by the claimant and the defence” (as summarised by Kolkman et al., 2024, p. 3). The extent to which this phase of issue and rule identification and factual presentation is driven respectively by parties/lawyers and judges varies between jurisdictions and different areas of law. Ultimately though, judges will need to decide on the inventory that they *take as relevant* in any oral hearing and eventual decision/s.

Early AI models purported to assist legal research in specific domains (Hill, 2025), while domain-specific predictive models, such as Legal-BERT and RoBERT, performed well in identifying arguments within the European Court of Human Rights corpus (Al Zubaer et al., 2023). However, the “democratisation” of generative AI with LLMs and their steadily increased power has seen their greater use by parties and judges. This has fostered rapid inventorising of relevant legal rules, which can help judges identify and cluster similar cases/applicable laws, and potentially enable a judge to see new patterns and subsume the case within new legal categories. However, the performance of such models “in the wild” can vary significantly given their dependence on underlying judicial data. The challenge of hallucinations is well known, with a new global database documenting the incidence of hallucinated citations in judgments (Curran et al., 2023; Charlotin, 2026). Even when models are trained on domain-relevant data with routine augmentation, Magesh et al. (2025, p. 216) find that AI research tools made by LexisNexis and Thomson Reuters each hallucinate between 17 percent and 33 percent of the time. This suggests that there should be significant caution in using current AI models to legally inventorise in an agent or trustee role. AI models are arguably most useful for identifying potential legal rules and arguments in a servant, interlocutor or – with some caveats – an advisor manner. We might assume that “false positives” will diminish over time – that phantom rules and precedents will not be offered – and that courts and judges will build up peer-reviewed and experiential knowledge on when AI can operate as an agent in such contexts.

The same conclusion on the dangers of trustee/agent uses of AI arguably applies to the identification and collation of the relevant issues and facts. Issue identification concerns the relevant facts in a case – i.e., what questions are legally relevant in the case? It partly precedes the identification of legal rules but is also an iterative process that occurs in dialogue with legal rules. Solid legal training to “see like a lawyer” is crucial. Issue identification is thus a skill that is often tested later rather than earlier in legal education, especially in more complex fact and legal scenarios. Nonetheless, as an advisor, AI may assist in identifying unseen but applicable patterns in the case law due to human blind spots.

However, human control is often needed given AI's blind spots, especially its reliance on past cases – the problem of *status quo* bias (Fikfak & Helfer, 2025, p. 102).

In relation to the listing of relevant facts, current AI technologies can certainly help in organising and summarising facts. Beyond such time-saving uses, its most powerful role might be in helping judges understand facts, especially given that legal training involves little focus on evidence and that courts often must cover every societal domain. Nonetheless, the familiar challenges of accuracy (hallucination) and selection bias (absence of training data) should be taken into account. The greatest danger may be cognitive offloading or automation bias (unreflexive deference to AI-generated responses) in cases with voluminous facts or many cases with similar facts (Gerlich, 2025). High levels of delegation can diminish judicial understanding in a particular case or cases over time as judicial learning and development is restricted.

4.3 Problem-solving

The next phase of adjudication concerns what we may think of as problem-solving. It involves attention to the facts – both their “selection” (establishing them and listing the points of dispute) and their “assessment” (determining whether relevant facts have been stated and if these facts are sufficiently substantiated) – as well as the decision, including remedies and sentencing (Ahsmann, 2011; Kolkman et al., 2024). Problem-solving can also include determinations of the law when its content is not clear. Such cases are often “hard cases,” arising “when no settled rule dictates a decision either way” (Dworkin, 1974).

In our view, problem-solving also includes the necessity to harmonise second-order considerations. Consideration of the systemic effects of a general legal interpretation, subsumption (application of law to facts) or a particular remedy is often a key part of judicial reasoning (Fuller & Winston, 1978; Ruhl et al., 2017). Legal solutions can have significant consequences for public policy, individuals and corporate entities (Horowitz, 1977). To be sure, some admonish courts to largely abstain from these sorts of considerations (Vermeule, 2006), especially the “creative” part, as attested by descriptions such as “legislation by judges,” “judicial activism,” etc.

However, harmonisation is often unavoidable for courts, even if it is not always acknowledged explicitly. Such instances particularly arise when new challenges arise that disturb the current legal equilibrium, be it due to social change, new technologies or reforms in other parts of the law. Most thus agree that some harmonisation by a court is sometimes necessary if it is going to be able to deliver legally sound judgments. The important disagreements concerning the harmonisation by courts thus should instead be about how often and how much it occurs.

This process of harmonisation – analogous to “reflective equilibrium” (Rawls, 1971) – often involves coming to see new patterns in perceptions, which seem better justified than those of the past. As MacCormick (1990, pp. 557–558) states:

The work of legal dogmatics is sometimes dismissed as mere transcription, with the scholar simply acting as no more than scribe to the judge or legislator. But this is a total misunderstanding. For it calls for the exercise of creative intelligence and disciplined imagination to master the large and always changing bodies of material involved, to grasp them all together, and to reconstruct them altogether into systematized and coherent wholes articulated out of complex and internally articulated parts ... the maintenance of doctrinal integrity through the good practice of legal dogmatics is what makes possible intelligent legislation and well-grounded judicial decision-making ... The juristic task has always been to establish intelligibility, not merely to discover it.

The key challenge is then how to govern or guide this judicial role to reduce risks of domination by unaccountable adjudicators. Who should have the authority to decide when adjudicators should not

stretch interpretations, or how to “rebalance” adjudicators’ priorities among legal norms? From the perspective of legitimacy, three points appear central to such discussions: *who performs these processes* of discovery and creation; *how* their quality can be *monitored*; and *how the processes or their outcomes can be changed*, and *by whom*.

Using AI in problem-solving can bring many advantages to judicial problem-solving, whether as advisor or agent. AI can be used to identify previous bias (Piccolo et al., 2023), correct for human heuristics in the assessment of law and facts (Piperides et al., 2006) and provide better access to legal information (Satterthwaite, 2025; Sunde, 2023). Indeed, some studies of “legal reasoning” in LLM show a tendency in the models toward a legal formalism that is less affected by sympathy for one of the parties (Posner & Saran, 2025), and AI may thus reduce mistrust that adjudicators are biased. AI systems may also assist in enhancing judicial quantitative skills, such as calculation of damages and penalties (Ruohonen & Hjerpe, 2020). Moreover, they may help navigate complex cases, in which levels of epistemic uncertainty are high due to the complexity of the problem caused by multiple, interconnected and dynamic elements (Funke, 2010). AI together with data science methods like network analysis and agent-based models can enable “systems thinking” by enabling formal or informal modeling of consequences of different legal approaches (Ruhl et al., 2017; Schwartz, 2020).

However, the limits of AI in judicial problem-solving tend to be visible as the levels of complexity increase in any central aspect of a legal problem. This is generally a challenge for LLM models addressing any complex problem-solving (Shojaee et al., 2025). In the case of judging, it is doubtful whether AI models alone can navigate multiple legal sources, address multiple and incommensurable factors in subsumption (e.g., proportionality analysis) or consider the distribution or magnitude of second-order effects. Several issues must be addressed by further research. First, it is unclear whether AI *can* or could perform the creative work – the sorts of conceptual/perceptual shifts and paradigm changes – involved in the reflective equilibrium or harmonisation processes. For example, in the UK, judges suddenly overturned established law that a husband cannot be guilty of raping his wife (UK, 1991). Second, in some complex cases, it is arguable – contrary to Dworkin (1986) – that no unique right answer can be found by proper judicial method, despite best efforts, be it with human or artificial Herculean cognition. For example, should the right to strike or the right to move goods across state borders in the EU have priority (De Vries, 2013)? In such cases, it is arguably human presence that is *currently* important in arriving at a legitimate decision. AI models – even as they improve – should be most likely used in interlocutor or advisor modus in addressing complex cases. This is because it is *adjudicators* who are currently thus authorised and made publicly accountable. To be sure, other branches of government may adjust such legal norms over time, but it is currently unclear what would be justifiable allocation of such authority with AI.

Moreover, when it comes to subsumption, even if LLMs may demonstrate greater legal formalism and are potentially less biased, they are limited in their sensory inputs, and it may be that the more visceral judicial experience of the courtroom provides additional information that may guide a fairer outcome (Posner & Saran, 2025). Bias may arise in other aspects of problem-solving, such as the determination of sentencing as the Compas experience shows (Neil & Zanger-Tishler, 2025). This suggests that AI models that seek to highly automate just one aspect of problem-solving or many aspects in very simple cases – from factual and legal assessment to subsumption and conclusions – should be subject to significant testing and strong and affordable rights of review. There is a risk that important information, including outliers and exceptions, is lost in the process.

4.4 Reasoned judgment

Much critical attention on the judicial use of AI has focused on explanation (Deeks, 2019), which is necessary for authoritative interpretation and justifying findings of fact, subsumption and remedies/sentencing. Richmond et al. (2024, p. 6) argue that AI must meet at least three specific explanatory functions to count as legal reasoning: it must be *comprehensive* (“supply evidence, support or

reasoning for each output”), *comprehensible* (“provide explanations that its users can understand”) and *correct* (an explanation should accurately reflect the process the AI system used to arrive at the output). However, the particular legal demands for explanation can vary significantly across different jurisdictions and fields of law.

Earlier classical rule-based AI models fulfilled these criteria to a certain or large extent, while struggling to ensure accuracy (Ashley, 2017). The same could be said for simpler quantitative models such as linear and logistic regression, general additive models, Bayesian models and decision trees, in which the relevant factors are specified in advance (Burrell, 2016). The “black box” problem has, however, bedeviled more recent and powerful AI models – whether predictive or generative AI. Reason-giving is not an intrinsic part of their design, as they extrapolate the most likely text based on past data.

Although it seems that this kind of algorithm performs some sort of case-based reasoning based on precedents—looking at the (factual, legal) aspects of previous cases to come to a decision in a new case—all they do is learn to identify statistical, possibly spurious, correlations between words in the text and the case decision. Hence, these algorithms cannot provide the legally relevant reasons for the outcome of a case. (Kolkman et al., 2024, p. 3)

This challenge is not unique to AI. There has been a similar and long-standing critique of juries as they rarely give reasons (Stern, 2025). However, not unsurprisingly, many government and national court systems have been under pressure – including from international human rights courts – to justify this delegation of judicial power for determining factual questions (Roberts, 2011). This has led some countries to “circumscribe or abolish their systems of lay participation” (Hans et al., 2024, p. 141).

The broader field of “XAI” has sought to enhance reason-giving through different “black box” methods (Hassija et al., 2024; Linardatos et al., 2020; Roscher et al., 2020). These may be *post hoc* – e.g., through Shapley values and LIME (Local Interpretable Model-agnostic Explanations) approximations – or counterfactual examples that interpret through selecting instances of the dataset, although challenges remain in identifying causal effects in complex AI systems (e.g., deep learning), accounting for inferences across domains and sub-domains, and communicating meaningful explainability (Jullum et al., 2021; Miller, 2019; Richmond et al., 2024). Addressing these challenges has been enhanced by the recent wave of “AI reasoning models” that are designed to “think” through problems using step-by-step logic, using methods such as chain of thought pretraining, backtracking and “self-reflection,” and greater use of reinforcement learning (Chen et al., 2025; Wang et al., 2026). However, questions still remain over the extent to which these methods sufficiently meet high legal explainability standards in decision-making, raising questions over the degree of output legitimacy and sociological assurance they create (Barry, 2024; Richmond et al., 2024).

From a systemic perspective, explanation also plays an additional role. If the concern is to avoid favouring one party to the dispute, AI may be a helpful mechanism, though other concerns must also be considered, such as attention to the societal impact of the interpretation, whether such choices can be monitored, so that legislatures may correct what they regard as misinterpretations by judges. In particular, as a judge has normative power to determine what is valid interpretation, they will sometimes set a new precedent. As Tasioulas (2023, p. 34) states, “the explanation of its decision-making obviously does not consist in the application of existing law to the case at hand.” Thus, explanation is not only important for a particular case (Richmond et al., 2024) or ensuring judicial accountability and assurance (see below), but it also provides the service of prediction, giving reasons for a departure from existing precedent so that it can be extended to other similar cases. Are there reasons why subjects should grant AI similar authority?

4.5 Legitimation

The fifth aspect of judging concerns assurance that the task is well done. Obviously, transparency about premises and explicit legal reasoning is an important aspect of providing assurance that the various and necessary elements of a judgment are sufficiently in place, such that it deserves compliance and that the court deserves general trust (Voeten, 2012). Thus, courts have increasingly devoted greater attention to public communication (Spencer, 2017). In addition, scholars observe that most courts are concerned with their sociological legitimacy – if not, survival – meaning that, in making judgments or imposing remedies, they tend to lag changes in public opinion or seek to triangulate between opposing societal interests to maintain their judicial capital (Roux, 2009; Tushnet, 2008).

At the same time, there is a strong public preference for effective decision-making in courts and public administration, such that weaknesses in process legitimacy may be *partly* traded off against gains in output legitimacy. This may be the case with AI. Empirical evidence shows that speedy adjudication is valued highly by citizens (Laux et al., 2024), with effectiveness sometimes even preferred over fairness (Bansak & Paulson, 2024). Thus, if AI can improve the efficiency and effectiveness of courts, then this might lower the demands on process legitimacy. However, some process legitimacy thresholds must be arguably met. Even in this respect, AI may help by providing more transparency about whether a process and decision are within the “normal” range, whether sources and precedents are correctly applied, etc. Thus, it can provide more “second opinions” and enhance the task of peer control by other jurists. Moreover, AI may make it easier for affected parties to decide whether to appeal. Over time, AI systems could develop in a multi-agent direction, so that different AI systems monitor each other and could signal to observers – whether there is consensus about the outputs or not – as certain observatory organisations do today.

Even assuming that phantoms and hallucinations are somehow eliminated in later versions, it is less clear which contributions AI can provide where transparency is required concerning AI models’ training and testing data, and their underlying algorithms and evidence on reliability, validity and statistical regularities. Managing legitimacy in complex cases with high levels of epistemic uncertainty, or in controversial cases with high levels of legal discretion and moral uncertainty, may remain challenges for AI in the modes of delegation. These cases require high levels of reflexivity and discretion in negotiating the legal, political and communicative space that judges operate within (Langford, 2026b). Attaining this type of competence is already a challenge for embodied human cognition (Dixon, 2023), and it is less clear how artificial cognition can master the nuances of such contextual multifaceted thinking and decision-making. Thus, it is notable that most empirical surveys tend to find little support for purely AI-based decision-making in courts (Laux et al., 2024), showing that the trustee models currently lack sociological assurance or legitimacy.

4.6 The legitimacy challenge of discretion

One aspect of the judicial function which poses particular challenges for legitimacy assessments of AI is the normative power of *discretion*. Judges exercise discretion in several of the stages of cases discussed above. This includes inventorising (how to (re)classify or subsume a case under some category), problem-solving (how to “balance” and “harmonise” legal norms within a range consistent with proper judicial method and rule, specifying a remedy or punishment) and giving reasoned judgment (based on some decision about which explanations are the most relevant to articulate).

From a legitimacy perspective, it is crucial to ask questions concerning input (are judges now authorised to share or even delegate such discretion to AI?) and process (*should* they do so?). Notably, until a few years ago, output legitimacy would have been an equally central question in most cases of the use of AI, as there was considerable doubt (including with first-generation LLMs) over whether AI could engage in even standard legal interpretation and application which involved aspects of discretion. The improved predictive and reasoning abilities of AI, especially where there is sufficient training data and less need for creative problem-solving, mean that the discussion has shifted to focus

more on the input and process dimensions – although certain output aspects remain central, such as explanation.

It may be useful to analyse the question of discretion by beginning with the initial delegation of adjudicative power to courts. Courts typically act as trustees for the parties to a dispute. As subjects of the courts, the parties have a shared overarching and second-order preference for an (sufficiently) impartial arbiter to settle their first-order dispute by particular means – legal sources – according to processes of judicial methods. An important part of these methods is that judges are authorised to exercise discretion in interpreting and applying the law and imposing remedies. To be sure, there are not only methodological limits to their discretion: subjects may also revoke the authority enjoyed by the domestic or international court through legislative or constitutional changes or treaty amendments, respectively. Consider what happens when judges delegate parts of these trustee tasks to AI. Two challenges arise, even in the best of cases in which AI *does* promote the appropriate judicial values. First, how can the judges using AI be sure that AI aligns dynamically with the underlying judicial values and preferences *over time* for this purpose (an issue of output legitimacy) (Abiri, 2024)? Second, it is not at all clear whether judges can or should be able to delegate their discretion regarding interpretations and remedies without loss to the authority of resulting decisions (an issue of input/process legitimacy). What reasons are there for subjects to replace their other reasons for action and defer to the discretionary decisions not of a judge but of AI? That is, do subjects have reason to accept such shifts in who exercises discretion and how, so that these modified authorities remain legitimate – or gain or lose legitimacy?

We underscore that whether the judges’ normative power requires humans in the loop is not simply a conceptual issue. Instead, it seems to be a received, often shared assumption – though with exceptions. Our point here is limited to suggesting that the (tacit) agreement to defer to judges’ discretion must be rethought to determine whether those or other arguments also apply to nonhuman agents, against a backdrop of which are the best alternatives. For instance, one reason to rely on the discretion of impartial judges in interpretation is that the parties may prefer that sort of bounded arbitrariness to being subject to decisions ultimately controlled by one of the parties to the conflict. This may, for instance, be the reasoning behind states’ decision to agree to “incomplete contracts,” which sometimes explicitly and implicitly authorise arbitrators to exercise discretion (Pauwelyn & Pelc, 2022). Note that this concern may be equally well secured by the mechanical solution of flipping a coin in cases where legal considerations “run out.” However, perceptions also matter (Keren & Teigen, 2010). Similar avoidance of bias by the bounded arbitrary exercise of discretion may be secured by relying on AI – *if* AI can be trusted to be impartial. Indeed, AI may well come to be seen as more reliably impartial, and that may in turn support attitudinal changes. But again, AI might not gain such a reputation.

A second reason to accept judges’ discretion may be the recognition that categorisation, interpretation, etc. must heed several somewhat conflicting considerations – simplicity, consistency, predictability, legislators’ intent, etc. – in some sort of (reflective) equilibrium. And the vetting process of judges combined with deliberative processes will arrive at some *non-objectionable* judgment which incorporates all such considerations. As Fuller (1964, pp. 41, 45) argued, the simultaneous pursuit of all aspects of the rule of law is unlikely to be feasible, making it “necessary to pursue a middle course which involves some impairment” of the relevant values or “desiderata.” Similar signaling, proxies and mechanisms of legal reasoning are hitherto absent from AI based on text prediction, as discussed above, even if AI is able to propose optimisation scenarios and – perhaps – increasingly explain them. Again, such features and requisite transparency may change in the near future. What may be lacking is a sufficiently overt authorisation by the subjects of the law to AI to exercise discretion alone or in conjunction with human judges, in ways that carry normative power.

5. Conclusion and discussion

In investigating the growing and potential use of AI in courts, we have sought to invoke Raz's legitimacy theory to highlight the underlying tensions and potential solutions in making normative assessments. We have distinguished different forms of delegation of judicial power to AI and examined the implications across six aspects of judging – from legal process and inventorising to problem-solving, reasoned judgment and legitimation. We have also flagged puzzles regarding judges' use of discretion. We conclude with three general topics for further discussion – leaving aside the issue of which core elements of input legitimacy must be in place regarding AI. These are process legitimacy under strong delegation, the roles of human discretion and the significance of human presence.

First, it seems clear AI *can* contribute to the output legitimacy of courts – both in specific judgments and generally, especially by enhancing effectiveness, reducing arbitrariness and providing new perspectives. However, these benefits are conditional and context-based. Even for seemingly low-risk uses of AI for translation and summarisation in legal processes, challenges of inaccuracy or selection bias can create challenges in particular cases. Uses are likely to be *more* legitimate when they improve the existing situation in courts (which may be suffering severe resource constraints), and when AI uses have been subjected to some form of evidence-based or experimental testing for their reliability and validity, and when there is some form of control over their use.

Likewise, the type of delegation to AI appears to affect its legitimacy, due to how and the extent to which it supplements and otherwise interacts with adjudicators, whether formally or informally through unintended cognitive offloading. We have sought to reflect AI's role through a delegation spectrum that ranges from AI acting as servant, interlocutor and advisor to partner, agent and trustee. The practice of “unseen actors” – such as secretariats and clerks – *de facto* exerting some aspects of judicial authority is not unknown. Debates more unique to AI arise when the benefits it brings to output legitimacy through its institutional memory and legal competence may be outweighed by other concerns. These include a lack of formally delegated authority to AI, its judicial competence, biases and blind spots and over-reliance on its contributions by overworked judges who cannot second-guess or quality control the preparations. The analysis tended to show that the deployment of AI as a servant for specific clear tasks, or as an interlocutor or advisor for dialogue, knowledge enhancement or nudging, raises less legitimacy concerns than the use of it as an autonomous agent, and especially as a trustee – promoting the judge's overarching interests or values according to AI, even if these assessments are contrary to the judge's express current preferences. If AI is granted some of the authority hitherto enjoyed by judges as trustees, we run into problems with the opaque and otherwise uncheckable discretion of AI to exercise normative power to issue judgments and interpretations, absent any public justifications based on legal norms.

Second, AI faces a particular epistemic challenge: that of judicial discretion. Ever since the rise of legal technology in the 1980s, the dilemma of how to model and mimic discretion in legal reasoning and method has been central (Ashley, 2017). Recent advances in machine learning have demonstrated that while aspects of that dilemma can be overcome by learning from past decisions, discretion is a multivalent concept in judicial reasoning, with a need for its exercise in different ways in both legal and cognitive processes. Judicial discretion is needed to navigate complex cases or to ensure that courts retain sufficient sociological legitimacy. However, it remains unclear to what extent AI can fill this role. Implications for the normative status of judicial discretion interacting with AI across these services remain a challenge for future research. When judges need to radically reclassify or re-cluster cases and harmonise conflicting legal norms – and second-order considerations – through reflective deliberation and articulation of reasons, it is not clear that AI can be more than a somewhat distant advisor or coach that assists the deliberative process.

The need for judicial discretion is not only triggered by complexity. Even in simpler cases, subsumption often requires bespoke application of law to specific cases, attention to the granularity of each individual case: “the autonomy to decide what should be done for each individual case”

(Christin, 2017, p. 2). In the case of criminal law, this “aligns with the principle of individualised justice that is responsive to the individual offender, their background and the offense” (Kolkman et al., 2024, p. 3). While some other areas of law may be strongly routinised or provide sufficient training data to cover all the factors to enable such responsiveness, this should not be automatically assumed.

Finally, AI currently faces an ontological challenge – that process legitimacy for courts currently appears to be partly built on *human presence*. Several important issues must be addressed elsewhere, such as the intelligibility of AI-generated output – what they “mean” (Cappelen & Dever, 2021b); and whether AI “believes” that their conclusions are correctly based on legal sources and methods (Chalmers, 2025). We have instead explored how human involvement might be needed for our acceptance of the exercise of judicial discretion in the face of high levels of epistemic uncertainty. Indeed, a fascinating aspect of adjudication is that judges have specific and inherent “normative powers” in this regard. When judges exercise their discretion to resolve vagueness in law or to select one among a set of underdetermined conclusions (e.g., regarding sentencing, compensation), their decision has legal significance. It is comparable to democratic legislative processes, which also carry normative power: a duly constituted legislature following certain procedures within some guardrails can create new binding law which carries legitimate authority over subjects. Unless the process is actually carried out, there is no such binding law. So, it seems impossible to simply *replace* the democratic deliberative decision about whose interests to give more weight in a particular policy domain. Similarly with judges, even though AI might do a “better” job of balancing, it lacks the normative power to establish new valid law – at least for now.

The reasons behind such an intuitive need for human presence are not always clear. It may be partly domain-specific. An important motivation for parties in litigating in courts, or accepting an adverse judgment, is the sense that one has been heard (Grieshofer, 2022). The conditions for this feeling are likely to be less met by AI, although it depends on the extent to which standard forms of adjudication can demonstrate, or perform, human presence – especially in fully written proceedings. This intuitive need may be more generic. It may be because there is a relational aspect to legitimacy (and trust) that forms an inherent accountability. Interestingly, in legitimacy theory, reciprocity is both inherent and foregrounded: service benefits are traded for subject compliance. It may be that this underlying or implicit social contract requires ultimately a fundamental and elemental human expression in which the corporeality and lived experience of the consequences of judicial authority requires that human judges stand publicly for that decision. This relational intuition finds support in arguments for “role-reversible” judgment, under which those who judge must themselves be exposed to the kind of normative assessment – and consequences – that they impose on others: AI cannot currently meet this condition (Brennan-Marquez & Henderson, 2019). We may thus want that particular persons or professionals to exercise the particular kinds of discretion involved in discovery and creation, rather than some other person or mechanism, even though the result in many ways might be better for us by that other mechanism.

This, of course, suggests that appeal courts should be particularly important for AI-assisted judging as they remain a last bastion for litigants. However, it should be remembered that for most citizens a first instance court is their only court: they usually have neither the capacity nor legal possibility to appeal. At the same time, discretion is often exercised not by an individual but by a collective of judges or by an institution. One implication may be that in some sense the collective – or AI – is not only a cognitive tool for the individual judge but also an integral part of the judges’ cognitive processes – as a case of “active externalism” of mind (Clark & Chalmers, 1998). Or, just as we may come to understand intentions as plans for action held not only by individuals but also by institutions, we may ascribe plans – and intentions – to AI (Bratman, 2022). Or there may be sufficiently trustworthy accountability systems that correct for errors before and after decisions. Thus, upon greater scrutiny, the current scepticism against AI’s role in judicial discretion may be less foundational and may turn out to change over time – and the AI as partner model may become particularly relevant or attractive here. We suggest that Raz’s approach may still be of help during such processes of change.

In assessing the legitimacy of AI in courts, what potentially matters is thus not only better outcomes, but who has the authority to carry out these tasks. While debates often concern the quality of outputs and degree of assurance, they are essentially also about process – about how judgments are made. Even if AI can or could reliably provide better outcomes, ontological and process-based challenges to its several contributions to legitimate authority remain.

Acknowledgements. We acknowledge comments received at the *Conference on Towards Trustworthy and Responsible AI in Courts*, Erasmus University, 17 October 2025, SMU Centre for Digital Law Seminar on 24 November 2025, UNSW *Comparative Constitutional Studies Conference*, 3–5 December 2025, and *Trustworthy AI Pop-Up Workshop*, University of Oslo, 16 April 2026.

Author Contributions. Credit Taxonomy Andreas Follesdal Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Supporting, Investigation-Equal, Methodology-Equal, Resources-Equal, Validation-Equal, Writing - original draft-Equal, Writing - review & editing-Equal Malcolm Langford Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Lead, Investigation-Equal, Methodology-Equal, Project administration-Equal, Resources-Equal, Validation-Equal, Writing - original draft-Equal, Writing - review & editing-Equal.

Funding statement. We acknowledge support from the Norwegian Research Council (TRUST: The Norwegian Centre for Trustworthy AI, Grant No. 357425) and Nordforsk (NordAid: Trustworthy AI in Public Decision Making project, Grant No. 233342).

Competing interests. The authors declare none.

References

- Abiri, G. (2024). Public constitutional AI. *arXiv preprint arXiv:2406.16696*.
- Ahsmann, M. J. A. M. (2011). *De weg naar het civiele vonnis*. Boom juridisch.
- Al Zubaer, A., Granitzer, M., & Mitrović, J. (2023). Performance analysis of large language models in the domain of legal argument mining. *Frontiers in Artificial Intelligence*, 6, 1278796.
- Alter, K. (2008). Agents or trustees? International courts in their political context. *European Journal of International Relations*, 14(1), 33–63. <https://doi.org/10.1177/13540661070877>.
- Ashley, K. D. (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge University Press.
- Austin, J. (1975). *How to do things with words* (2nd ed.). Harvard University Press.
- Baetens, F. (2019). *Legitimacy of unseen actors in international adjudication*. Cambridge University Press.
- Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., & McKinnon, C. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint arXiv:2212.08073*.
- Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet*, 17(9), 404.
- Bansak, K., & Paulson, E. (2024). Public attitudes on performance for algorithmic and human decision-makers. *PNAS nexus*, 3(12), 520.
- Barry, B. M. (2024). AI for assisting judicial decision-making: Implications for the future of open justice. *Australian Law Journal*, 98(9), 656–669.
- Baste, Ø., Cyndecka, M., Esayas, S., Langford, M., Lison, P., & Weitzenboeck, E. (2025). *Open justice data in Europe: A patchwork* Paper presented at the SSRN Working Paper, CLEANUP Project, University of Oslo.
- Baste, Ø. F., Esayas, S. Y., & Langford, M. (2024). *Licensing of Reuse of judgments: Analysis of selected European jurisdictions*. Strasbourg: Council of Europe.
- Beach, D., & Kaas, J. G. (2020). The great divides: Incommensurability, the impossibility of mixed-methodology, and what to do about it. *International Studies Review*, 22(2), 214–235.
- Bellamy, R. (2008). The democratic constitution: Why Europeans should avoid American Style constitutional judicial review. *European Political Science*, 7, 9–20.
- Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., & Richardson, O. (2025). Superintelligent agents pose catastrophic risks: Can scientist AI offer a safer path? *arXiv preprint arXiv:2502.15657*.
- Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Shalev-Shwartz, S., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., ... Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. *Science*, 384(6698), 842–845.
- Bentsen, H. L., & Johannesson, M. (2026). Would people reject AI judgments as being procedurally unfair? SSRN.
- Bentsen, H. L., Skiple, J. K., Langford, M., Birkeland Deichman-sørensen, M., & Grendstad, G. (2026). Predicting case processing times in Norwegian Courts using tree-based machine learning. Available at SSRN 6672618.

- Blackburn, S.** (2006). *Plato's Republic: A biography*. Atlantic Press. 1151
- Bodansky, D.** (1999). The legitimacy of international governance: A coming challenge for international environmental law? *The American Journal of International Law*, 93(3), 596–624. 1152
- Bratman, M.** (2022). *Shared and institutional agency: Toward a planning theory of human practical organization*. Oxford University Press. 1153
- Brennan-Marquez, K., & Henderson, S. E.** (2019). Artificial intelligence and role-reversible judgment. *Journal of Criminal Law and Criminology*, 109(2), 137–164. 1154
- Broomfield, H., & Reutter, L.** (2022). In search of the citizen in the datafication of public administration. *Big Data and Society*, 9(1), 20539517221089302. 1155
- Buchanan, A., & Keohane, R.** (2008). The legitimacy of global governance institutions. In R. Wolfrum & V. Röben (Eds.), *Legitimacy in International Law* (pp. 25–62). Springer. 1156
- Buhmann, A., & Fieseler, C.** (2023). Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly*, 33(1), 146–179. 1157
- Burrell, J.** (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1), 2053951715622512. 1158
- Caballero, L.** (2026). Moltbook: AI bots use social network to create religions and deal digital drugs – But are some really humans in disguise? *The Conversation*, 5 February. 1159
- Capasso, V.** (2024). For an...«Artificially Intelligent» Process: When access to justice and efficient court management go hand in hand. *International Journal of Procedural Law*, 14(2), 245–272. 1160
- Cappelen, H., & Dever, J.** (2021a). *Making AI intelligible: Philosophical foundations*. Oxford University Press. 1161
- Cappelen, H., & Dever, J.** (2021b). *Making AI Intelligible: Philosophical foundations*. Oxford University Press: Oxford University Press. 1162
- Capper, D.** (1991). Worldwide Mareva injunctions. *The Modern Law Review*, 54(3), 329–348. 1163
- Celeste, E.** (2019). Digital constitutionalism: A new systematic theorisation. *International Review of Law, Computers & Technology*, 33(1), 76–99. 1164
- CEPEJ.** (2018). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. Council of Europe. 1165
- Chalmers, D. J.** (2025). What we talk to when we talk to language models. *Philosophy Papers*. 1166
- Charlotin, D.** (2026). Dealing with hallucinations: a guide for international judges and arbitrators. *Journal of International Dispute Settlement*, 17(3), idag020. 1167
- Chen, Q., Qin, L., Liu, J., Peng, D., Guan, J., Wang, P., & Che, W.** (2025). Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*. 1168
- Chong, D., & Druckman, J. N.** (2007). Framing theory. *Annual Review of Political Science*, 10(1), 103–126. 1169
- Christin, A.** (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2). 1170
- Clark, A., & Chalmers, D.** (1998). The extended mind. *Analysis*, 58(1), 7–19. 1171
- Coeckelbergh, M.** (2023). Democracy, epistemic agency, and AI: Political epistemology in times of artificial intelligence. *AI and Ethics*, 3(4), 1341–1350. 1172
- Cohen, J., & Liu, L. T.** (2025). The reach of fairness. *ACM Journal on Responsible Computing*. <https://doi.org/10.1145/3718989> 1173
- Committee of Ministers.** (2020). On the human rights impacts of algorithmic systems, Appendix 1. *Recommendation CM/Rec(2020)(Apr 8)*. <https://rm.coe.int/09000016809e1154>. 1174
- Crawford, K.** (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. 1175
- Curran, S., Lansley, S., & Bethell, O.** (2023). Hallucination is the last thing you need. *arXiv preprint arXiv:2306.11520*. 1176
- de Abreu Duarte, F., De Gregorio, G., & Golia, A.** (2023). Perspectives on digital constitutionalism. In B. Brożek, O. Kanevskaia, & P. Palka (Eds.), *Research handbook on law and technology* (pp. 315–329). Edward Elgar Publishing. 1177
- De Vries, S. A.** (2013). Balancing fundamental rights with economic freedoms according to the European Court of Justice. *Utrecht Law Review*, 9(1), 169–192. 1178
- Deeks, A.** (2019). The judicial demand for explainable artificial intelligence. *Columbia Law Review*, 119(7), 1829–1850. 1179
- Deuschel, J., Foltyn, A., Roscher, K., & Scheele, S.** (2024). The role of uncertainty quantification for trustworthy AI. In C. Mutschler, C. Münzenmayer, N. Uhlmann, & A. Martin (Eds.), *Unlocking Artificial Intelligence: From theory to applications* (pp. 95–115). Springer. 1180
- Dixon, H. B. Jr, Goddard, A. H., Grossman, M. R., Rodriguez, X., Schlegel, S. U., & Thumma, S. A.** (2025). *Navigating AI in the judiciary: New guidelines for judges and their chambers*. Paper presented at the Sedona Conf. J. 1181
- Dixon, R.** (2023). *Responsive judicial review: Democracy and dysfunction in the modern age*. Oxford University Press. 1182
- Donohue, M.** (2018). A replacement for justitia's scales: Machine learning's role in sentencing. *Harvard Journal of Law and Technology*, 32, 657–678. 1183
- Dunoff, J. L., & Pollack, M. A.** (2017). The judicial trilemma. *American Journal of International Law*, 111(2), 225–276. 1184
- Dworkin, R.** (1974). Hard cases. *Harvard Law Review*, 88(6), 1057. 1185
- Dworkin, R.** (1986). *Law's Empire*. Belknap Press. 1186
- EU.** (2019). *Building Trust in Human-Centric Artificial Intelligence*. 1187

- European Parliament.** (2023). AI Act. (P9_TA(2023)236).
- Fikfak, V., & Helfer, L.** (2025). Automating international human rights adjudication. *Michigan Journal of International Law*, 46(1), 69–120.
- Follesdal, A.** (2006). The legitimacy deficits of the European Union. *Journal of Political Philosophy*, 14(4), 441–468.
- Follesdal, A.** (2011). The distributive justice of a global basic structure: A category mistake? *Politics, Philosophy & Economics*, 10(1), 46–65.
- Follesdal, A.** (2017). Constitutionalization, not democratization: How to assess the legitimacy of international courts. In N. Grossman, H. Cohen, A. Follesdal, & G. Ulfstein (Eds.), *The legitimacy of international courts* (pp. 307–337). Cambridge University Press, Forthcoming.
- Follesdal, A.** (2020). Survey article: The legitimacy of international courts. *Journal of Political Philosophy*, 28(4), 476–499.
- Fonseca, B.** (2025). *Towards Responsible AI Governance in the Brazilian Judiciary*. Paper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.
- Fuller, L.** (1964). *The morality of law*. Yale University Press.
- Fuller, L., & Winston, K. I.** (1978). The form and limits of adjudication. *Harvard Law Review*, 92(2), 353–409.
- Funke, J.** (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133–142.
- Gailmard, S.** (2013). Accountability and Principal-Agent Models. In M. Bovens, T. Schillemans & R. E. Goodin (Eds.), *Oxford Handbook of Public Accountability* (pp. 90–105). Oxford University Press.
- Gargarella, R.** (2022). *The law as a conversation among equals*. Cambridge University Press.
- Gerlich, M.** (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6.
- Gibson, J.** (2007). The legitimacy of the U.S. Supreme court in a polarized polity. *Journal of Empirical Legal Studies*, 4(3), 507–538.
- Giudici, P., Centurelli, M., & Turchetta, S.** (2024). Artificial Intelligence risk measurement. *Expert Systems with Applications*, 235, 121220.
- Goodin, R.** (1992). *Motivating political morality*. Blackwell.
- Goodin, R. E.** (2013). *Green political theory*. John Wiley & Sons.
- Grieshofer, T.** (2022). The importance of being heard: Stories of unrepresented litigants in small claims cases and private family proceedings. *Language and Law/Linguagem e Direito*.
- Hale, T.** (2024). *Long problems: Climate change and the challenge of governing across time*. Princeton University Press.
- Hamilton, A.** (1961). The Federalist no. 78.
- Hans, V. P., Diamond, S. S., Kutnjak Ivković, S., & Marder, N. S.** (2024). Judgment by peers: Lay participation in legal decision making. *Annual Review of Law and Social Science*, 20(1), 141–161.
- Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Hussain, A., Spinelli, I., Mahmud, M., & Hussain, A.** (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, 16(1), 45–74.
- Hendrickx, V.** (2025). *Rethinking the judicial duty to state reasons in the age of automation?* Paper presented at the Cambridge Forum on AI: Law and Governance.
- Hill, C.** (2025). Judge rules that ross Intelligence infringed Westlaw's Copyright in Landmark AI Opinion. *Legal IT Insider*, 12, February.
- Hillebrecht, C.** (2014). *Domestic politics and international human rights tribunals: The problem of compliance*. Cambridge University Press.
- Holten Møller, N., Shklovski, I., & Hildebrandt, T. T.** (2020). Shifting concepts of value: Designing algorithmic decision-support systems for public services. *Paper presented at the Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*.
- Horowitz, D.** (1977). *The Courts and Social Policy*. The Brookings Institution.
- Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., & De, R.** (2025). AI agents and agentic systems: a multi-expert analysis. *Journal of Computer Information Systems*, 65(4), 1–29.
- Jegham, N., Abdelatti, M., Koh, C. Y., Elmoubarki, L., & Hendawi, A.** (2025). How hungry is AI? Benchmarking energy, water, and carbon footprint of LLM inference. *arXiv preprint arXiv:2505.09598*.
- Jiawei, Z., & Moreira, J. I.** (2023). Promoting trustworthiness in the application of artificial intelligence in the judiciary: the intersection of media communication, court decisions, and public trust. *International Journal of Criminal Justice Sciences*, 18(2), 481–496.
- Jullum, M., Redelmeier, A., & Aas, K.** (2021). groupShapley: Efficient prediction explanation with Shapley values for feature groups. *arXiv preprint arXiv:2106.12228*.
- Kane, J., & Patapan, H.** (2010). Recovering Justice: Political legitimacy reconsidered. *Politics & Policy*, 38(3), 589–610.
- Kaur, D., Uslu, S., Rittichier, K. J., & Durresti, A.** (2022). Trustworthy Artificial Intelligence: A review. *ACM Computing Surveys*, 55(2), Article 39. <https://doi.org/10.1145/3491209>
- Kayode, B., Adebola, N., Akerle, S., Fagbohun, O., Agbo, C., Bantale, O., & Nwokocha, L. C.** (2025). The state of AI-driven cybersecurity: Trends, challenges, and opportunities. *Journal of Artificial Intelligence, Machine Learning and Data Science*, 3(2), 2731–2739.

1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250

- Kehl, D. L., & Kessler, S. A. (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*. Retrieved from. Responsive Communities Initiative, Berkman Klein Center for Internet & Society.
- Keren, G., & Teigen, K. H. (2010). Decisions by coin toss: Inappropriate but fair. *Judgment and Decision Making*, 5(2), 83–101.
- Kern, F. G. (2018). The trials and tribulations of applied triangulation: Weighing different data sources. *Journal of Mixed Methods Research*, 12(2), 166–181.
- Kolkman, D., Bex, F., Narayan, N., & van der Put, M. (2024). Justitia ex machina: The impact of an AI system on legal decision-making and discretionary authority. *Big Data & Society*, 11(2). <https://doi.org/10.1177/20539517241255101>.
- Langford, M. (2026a). Anticipatory Governance and AI in Courts. In M. Zalnieriute & A. Limante (Eds.), *The Cambridge Handbook of AI and technology in courts* (pp. 102–122). Cambridge University Press.
- Langford, M. (2026b). *Judging complex cases*: Forthcoming.
- Langvatn, S. A., & Squatrito, T. J. (2017). Conceptualising and measuring the legitimacy of international criminal tribunals. In N. Hayashi & C. Bailliet (Eds.), *The legitimacy of international criminal tribunals* (pp. 41–65). Cambridge University Press.
- Latif, Y. A. (2025). Hallucinations in large language models and their influence on legal reasoning: Examining the risks of AI-generated factual inaccuracies in judicial processes. *Journal of Computational Intelligence, Machine Reasoning, and Decision-Making*, 10(2), 10–20.
- Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy Artificial Intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1), 3–32.
- Leslie, D., Burr, C., Aitken, M., Katell, M., Briggs, M., & Rincon, C. (2022). Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal. *arXiv preprint arXiv:2202.02776*.
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1–46.
- Lie, R. H., & Langford, M. (2024). The computational turn in international law. *Nordic Journal of International Law*, 93(1), 38–67.
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18:11–45.
- Longoni, C., Cian, L., & Kyung, E. (2022). AI in the government: Responses to failures. *PsyArXiv*, June 1.
- MacCormick, N. (1990). Reconstruction after deconstruction: A response to CLS. *Oxford Journal of Legal Studies*, 10(4), 539–558.
- Magalhães, P. C., Arnesen, S., Kern, C., Koenig, P. D., Schiff, D. S., & Tyler, T. R. (2026). Transparency, neutrality, voice, and respect: How procedural fairness considerations affect AI acceptability in algorithmic societies. *Technology in Society*, 103398.
- Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-Free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies*, 22(2), 216–242.
- Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2023). Using AI uncertainty quantification to improve human decision-making. *arXiv preprint arXiv:2309.10852*.
- McConvey, K., Burkell, J., Deng, Y., Ebrahimi, S., Grossman, M., Mahdaviomghaddam, J., & Bagheri, E. (2026). Datasets, detection methods, and evaluation strategies for managing synthetic evidence in the judicial system. *Paper presented at the Governing Agentic AI Symposium*, Lisbon.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.
- Neil, R., & Zanger-Tishler, M. (2025). Algorithmic bias in criminal risk assessment: The consequences of racial differences in arrest as a measure of crime. *Annual Review of Criminology*, 8(1), 97–119.
- Palmiotto, F. (2024). When is a decision automated? A taxonomy for a fundamental rights analysis. *German Law Journal*, 25(2), 210–236.
- Pasquale, F. (2019). A rule of persons, not machines: The limits of legal automation. *George Washington Law Review*, 87, 1.
- Pauwelyn, J., & Pelc, K. (2022). Who guards the “Guardians of the System”? The role of the secretariat in WTO dispute settlement. *American Journal of International Law*, 116(3), 534–566.
- Peer, E., & Gamliel, E. (2013). Heuristics and biases in judicial decisions. *Court Review*, 49, 114.
- Peter, F. (2010). Political Legitimacy. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*.
- Piccolo, S. A., Katsikouli, P., Gammeltoft-hansen, T., & Slaats, T. (2023). On predicting and explaining asylum adjudication. *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, 217–226.
- Piperides, C., Allen, R. J., Dhami, M. K., Flessner, A., Hastie, R., Koehler, J. J., & Wagner, G. (2006). Group Report: What is the role of heuristics in litigation. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the Law* (pp. 343–377). MIT Press.
- Posner, E. A., & Saran, S. (2025). Judge AI: Assessing large language models in judicial decision-making. *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper*(2503).
- Rawls, J. (1971). *A theory of justice*. Columbia University Press.
- Raz, J. (1986). *The morality of freedom*. Oxford University Press.
- Raz, J. (1994). Law, authority and morality. In J. Raz (Ed.), *Ethics in the Public Domain* (pp. 210–237). Oxford University Press.
- Raz, J. (2006). The problem of authority: Revisiting the service conception. *Minnesota Law Review*, 90(4), 1003–1044. http://www.philosophy.ox.ac.uk/_data/assets/pdf_file/0018/1467/Authority.pdf

- Re, R.M., & Solow-Niederman, Alicia. (2019). Developing artificially intelligent justice. *Stanford Technology Law Review*, 22(2), 242–289.
- Richmond, K. M., Muddamsetty, S. M., Gammeltoft-Hansen, T., Olsen, H. P., & Moeslund, T. B. (2024). Explainable AI and law: An evidential survey. *Digital Society*, 3(1), 1–33.
- Roberts, P. (2011). Does Article 6 of the European convention on human rights require reasoned verdicts in criminal trials? *Human Rights Law Review*, 11(2), 213–235.
- Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8, 42200–42216.
- Rosenthal, J. S., & Yoon, A. H. (2010). Judicial ghostwriting: Authorship on the Supreme Court. *Cornell Law Review*, 96, 1307.
- Roux, T. (2009). Principle and pragmatism on the Constitutional Court of South Africa. *International Journal of Constitutional Law*, 71(1), 106–138.
- Ruhl, J., Katz, D. M., & Bommarito, M. J. (2017). Harnessing legal complexity. *Science*, 355(6332), 1377–1378.
- Ruohonen, J., & Hjerpe, K. (2020). Predicting the amount of GDPR fines. *arXiv preprint arXiv:2003.05151*.
- Sabel, C. F. (1997). Constitutional orders: Trust building and response to change. In J. R. Hollingsworth & R. Boyer (Eds.), *Contemporary Capitalism: The Embeddedness of Institutions* (pp. 154–188). Cambridge University Press.
- Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, 11(2), 88–95.
- Satterthwaite, M. (2025). *Report of the Special Rapporteur on the Independence of judges and lawyers: Artificial intelligence in judicial systems: Promises and pitfalls*: UN doc. A/80/169, 16 July 2025.
- Schäferling, S. (2023). *Governmental automated decision-making and human rights: Reconciling law and intelligent systems*. Springer.
- Schärf, F. (1999). *Governing in Europe: Effective and democratic?* Oxford University Press.
- Schmauder, C., Karpus, J., Moll, M., Bahrami, B., & Deroy, O. (2023). Algorithmic nudging: The need for an interdisciplinary oversight. *Topoi*, 42(3), 799–807.
- Schwartz, A. (2020). Agent-based modeling for legal studies. In R. Whalen (Ed.), *Computational legal studies* (pp. 312–327). Edward Elgar Publishing.
- Sen, A. (2017). *Collective choice and social welfare* (An Expanded Edition ed.). Harvard University Press.
- Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S., & Farajtabar, M. (2025). The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *arXiv preprint arXiv:2506.06941*.
- Sino, M., Falco, M., Croce, D., & Rosso, P. (2025). Exploring LLMs applications in law: A literature review on current legal NLP approaches. *IEEE Access*, 13, 18253–18276.
- Sourdin, T. (2018). Judge v Robot? Artificial Intelligence and judicial decision-making. *University of New South Wales Law Journal*, 41(4), 1114–1133.
- Sourdin, T., & Zariski, A. (2018). What is responsive judging? In T. Sourdin & A. Zariski (Eds.), *The Responsive Judge: International Perspectives* (pp. 1–38). Springer.
- Spencer, R. (2017). Communication beyond the judgments: The Australian High Court, speaking for itself, but not tweeting. In R. Davis & D. Taras (Eds.), *Justices and Journalists: The Global Perspective* (pp. 39–57). Cambridge University Press.
- Stern, K. (2025). Commentary on John Tasioulas, The rule of algorithm and the rule of law. *Court of Appeal, Supreme Court of NSW*. https://supremecourt.nsw.gov.au/documents/Publications/Speeches/2025-speeches/Stern_20250902.pdf
- Sturm, S. P. (1990). A normative theory of public law remedies. *Georgetown Law Journal*, 79, 1355.
- Sunde, J. (2023). *1000 år med Norsk Rettshistorie: Ei Annleis Noregshistorie om Rett, Kommunikasjonsteknologi, Historisk Endring og Rettstat*. Dreyer Forlag.
- Susskind, R., & Susskind, R. E. (2023). *Tomorrow's lawyers: An introduction to your future*. Oxford University Press.
- Sweet, A. S., & Grisel, F. (2017). *The evolution of international arbitration: Judicialization, governance, legitimacy*. Oxford University Press.
- Stone Sweet, A., & Brunell, T. (2013). Trustee courts and the judicialization of international regimes. *Journal of Law and Courts*, 1(1), 61–88.
- Tait, K., & Taylor, W. K. (2022). The possibility of rights claims-making in court: looking back on twenty-five years of social rights constitutionalism in South Africa. *Law & Social Inquiry*, 48(3), 1–30.
- Tasioulas, J. (2022). Artificial Intelligence, humanistic ethics. *Daedalus*, 151(2), 232–243.
- Tasioulas, J. (2023). The rule of algorithm and the rule of law. *Vienna Lectures on Legal Philosophy* (2023). Sydney: Bloomsbury.
- Taylor, M. (1987). *The possibility of cooperation*. Cambridge University Press.
- Terzis, P. (2024). Against digital constitutionalism. *European Law Open*, 3(2), 336–352.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Tushnet, M. (2008). *Weak courts, strong rights: Judicial review and social welfare rights in comparative constitutional law*. Princeton University Press.
- UK, L. L. (1991). *R v R*. [1991] 4 All ER 481 UKHL 12.
- Vermeule, A. (2006). *Judging under uncertainty: An institutional theory of legal interpretation*. Harvard University Press.

Voeten, E. (2012). *Does a professional judiciary induce more compliance? Evidence from the European Court of Human Rights*. SSRN Working Paper: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2029786.

Wachter, S., Mittelstadt, B., & Russell, C. (2020). Bias preservation in machine learning: The legality of fairness metrics under EU non-discrimination law. *West Virginia Law Review*, 123, 735.

Waldron, J. (2006). The core of the case against judicial review. *The Yale Law Journal*, 115(6), 1346–1406.

Wang, P.-Y., Liu, T.-S., Wang, C., Li, Z., Wang, Y., Yan, S., & Xu, J. (2026). A survey on large language models for mathematical reasoning. *ACM Computing Surveys*, 58(8), 1–35.

Weber, M. (1964). *The Theory of Social and Economic Organization*. Free Press.

Williams, R., Cloete, R., Cobbe, J., Cottrill, C., Edwards, P., Markovic, M., Pang, W., Ryan, F., Singh, J., & Pang, W. (2022). From transparency to accountability of intelligent systems: Moving beyond aspirations. *Data & Policy*, 4, e7.

Yu, B., & Barter, R. L. (2024). *Veridical data science: The practice of responsible data analysis and decision making*. MIT Press.