Articles + Publications | May 5, 2026

A Human in the Loop: The Future of AI in Arbitration

WRITTEN BY

Jeremy Heep | Victoria A. Alvarez | Savannah Billingham-Hemminger

Introduction

In today's legal landscape, artificial intelligence (AI) is no longer a distant or speculative concept. Instead, it has morphed into an everyday, practical tool for dispute resolution, especially in arbitration. Although technology-assisted processes have supported arbitral practice for decades, the emergence of generative AI represents a material shift in its application. These systems can generate text, summaries, and analytical outputs by identifying patterns across vast datasets, offering new possibilities for efficiency, cost control, and decision support throughout the arbitral process. As its capabilities continue to expand, the question becomes: how much of our arbitral practice can AI contribute to a more efficient process, and what part of decision-making still requires what commentators increasingly refer to as "a human in the loop"?[1]

The use of AI in arbitration already spans a wide range of applications. At the most limited level, AI tools are employed to assist parties or arbitrators with discrete tasks, such as summarizing submissions, organizing evidence, or reviewing procedural materials. At an intermediate level, AI has been introduced into more consequential procedural functions, including the selection or ranking of arbitrators based on data-driven criteria. At the most expansive end of the spectrum, proposals and pilot programs envision AI playing a direct role in decision-making itself, including the development of AI-assisted or AI-driven arbitral models within institutional frameworks.

As these applications become more sophisticated, they raise concerns that extend beyond efficiency alone. Questions of transparency, explainability, bias, accountability, and human oversight, move from the abstract to the concrete once AI tools influence procedural choices or substantive outcomes. Their deployment implicates essential procedural guarantees, including due process, equality of arms, disclosure obligations, and the parties' confidence in the legitimacy of the arbitral process and award. And when it comes to substantive choices, the implicit bias of inputs and lack of transparency in the process call into question: how many of our systems are irreplaceably human? The future of AI in arbitration will therefore be shaped less by whether AI is used at all and more by how firmly human judgment is kept "in the loop" at each stage of the arbitral lifecycle.

This article examines the use of AI in arbitration across this spectrum of small-, medium-, and large-scale applications. It compares the distinct characteristics, challenges, and implications associated with each approach and considers how AI may be integrated into arbitral practice in a manner that strengthens, rather than erodes, fairness, procedural integrity, transparency, and confidence in arbitral decision-making.

I. Support, Not Substitution: Limited Uses of AI in Arbitration

At the most limited end of the spectrum, AI is deployed as a support tool rather than a decision-maker. In international arbitration, this use most commonly arises in the review and organization of client documents, bundle preparation, and procedural case management. AI tools can classify documents, extract relevant information, suggest targeted search terms, and identify factual or thematic patterns across large datasets almost instantaneously. When integrated into existing workflows, these applications can materially enhance efficiency, reduce cost, and accelerate proceedings, allowing counsel and arbitrators to devote greater attention to legal analysis, procedural strategy, and the merits of the dispute.

Limited-use AI is also increasingly employed as a legal research and drafting aid. Large language models (LLMs) can rapidly identify relevant case law and statutory authorities, summarize complex legal doctrines, synthesize applicable legal standards, and generate primers on unfamiliar subject matter. These capabilities are now embedded in established legal research platforms, including Westlaw, where AI-driven functionality operates within closed, source-verified databases. Used properly, such tools can streamline early-stage research and sharpen issue identification, but they still do not replace legal reasoning or professional judgment.

Practitioners seeking to get started should confine initial use to clearly defined tasks and adopt disciplined prompting techniques. Prompts should narrowly specify the scope of the request, restrict the AI to identified materials or databases, and require the provision of complete, verifiable citations. Instructions to summarize "without adding facts," to quote directly from supplied documents, or to identify authorities "only if traceable to an existing source" can materially reduce the risk of fabricated output. These prompting strategies are, in effect, mechanisms for keeping human control inside the loop — structuring how AI is used, constraining its sources, and signaling that the ultimate responsibility for accuracy remains with the practitioner.

The risks associated with even these restrained applications are no longer theoretical. A significant number of instances have been reported in state and federal courts in which attorneys were sanctioned for filing briefs containing hallucinated citations. In a recent opinion underscoring the ethical risks of uncritical reliance on generative tools, the U.S. Court of Appeals for the Sixth Circuit removed Kentucky attorney Steven N. Howe from representing appellant John C. Farris. *United States v. Farris,* No. 25?5623, (6th Cir. Apr. 3, 2026) (slip op. at 6) ("by separate order issued on this same date, we remove Howe from further representation of Farris"). The court found that Howe had "committed inexcusable transgressions during the appellate phase of this case" by filing appellate briefs generated with Westlaw's CoCounsel platform. By Howe's own admission, he directed staff "to upload district court documents to Westlaw's CoCounsel program to create a first draft" of both briefs, without "properly verifying the cited legal authorities," which resulted in "multiple misrepresentations of law to this Court," including "three inaccurate quotations" that "do not appear in any legal authorities" — entirely misrepresenting the holdings of the cited cases. *Id*. at 1–4, 6. While acknowledging that "new technologies present significant promise for the legal field," the panel cautioned that "all in the legal profession must be clear eyed about technology's potential pitfalls . . . ." *Id*. at 6. Moreover, they emphasized that "attorneys who rely on artificial intelligence must remain diligent in supervising their work product and carefully examine the accuracy of every citation they present to this Court," criticizing Howe's reliance on "staff" — rather than himself or another attorney — to supervise the AI?generated work product as falling "short of his obligations as attorney of record." *Id*. at 5. The court further emphasized that "attorneys who choose to use artificial?intelligence tools must do so in a manner consistent with their ethical obligations," and that relevant steps "may include reviewing and validating content produced by artificial intelligence; considering whether to disclose the use of artificial intelligence to clients or obtain informed consent; safeguarding confidential client information and preserving attorney?client privilege; implementing

firm?wide policies governing the use of artificial intelligence; adhering to ethical billing practices when using artificial?intelligence tools; and keeping current with jurisdiction?specific guidelines." *Id*. at 5.

In Howe's case, the court found that he "failed to adequately review and verify the draft brief produced by artificial intelligence" and that his "failure to verify the artificial?intelligence output still resulted in the submission of false quotations and misleading legal arguments to this Court," which "necessitated a significant use of judicial resources to investigate the suspected artificial?intelligence improprieties, coordinate a response, and facilitate additional steps of these appellate proceedings." *Id*. at 4–6. The panel concluded that "attorneys have an ethical obligation to verify the citations and propositions they submit to courts; that obligation reflects duties of competence and candor that apply no matter the tools attorneys use," and that "new technologies, moreover, are no substitute for tried?and?true safeguards managed by practicing attorneys," underscoring that technological convenience cannot replace traditional professional judgment and supervision. *Id*. at 5.

In a recent French proceeding, the Tribunal Judiciaire of Périgueux formally identified legal submissions relying on "untraceable or erroneous" precedents, and on December 18, 2025, expressly noted that the claimant's arguments contained fabricated authorities. *See* Tribunal Judiciaire de Périgueux, December 18, 2025, n° 23/00452. Comparable failures have arisen in other common law jurisdictions, too. For example, in a damages action for £89.4 million brought by Hamad Al?Haroun against Qatar National Bank QPSC and QNB Capital LLC in the High Court of Justice, King's Bench Division, the claimant's solicitor placed before the court a schedule of 45 case citations; in 18 instances, the cited case did not exist, and among the authorities that did exist, many "did not contain the quotations that were attributed to them, did not support the propositions for which they were cited, and did not have any relevance to the subject matter of the application . . . ." *R (Ayinde) v. London Borough of Haringey; Al?Haroun v. Qatar Nat'l Bank QPSC & QNB Capital LLC,* [2025] EWHC 1383 (Admin) [73]–[74] (Eng.) In its June 6, 2025, Divisional Court judgment, the court concluded that "[t]he vast majority of the authorities are made up or misunderstood." *Id*. at ¶ 74. The High Court warned that "[t]here are serious implications for the administration of justice and public confidence in the justice system if artificial intelligence is misused" and emphasized that "practical and effective measures must now be taken by those within the legal profession with individual leadership responsibilities (such as heads of chambers and managing partners) and by those with the responsibility for regulating the provision of legal services . . . ." *Id.* at ¶ 9. Moreover, those measures must ensure that "every individual currently providing legal services within this jurisdiction (whenever and wherever they were qualified to do so) understands and complies with their professional and ethical obligations and their duties to the court if using artificial intelligence." *Id*. at ¶ 9.

Concerns extend beyond just hallucinations to confidentiality and data security. Arbitration proceedings routinely involve sensitive commercial information subject to strict confidentiality obligations. Because of the learning process utilized by generative AI, inputting such material into publicly accessible generative AI platforms risks unauthorized disclosure and downstream use beyond the parties' control. For this reason, limited AI use in arbitration should be confined to secure, private environments and accompanied by verification protocols equivalent to those applied to human work product. When deployed within these boundaries, AI can enhance efficiency and speed without undermining accuracy, professional responsibility, or confidence in the arbitral process and the enforceability of the resulting award.

II. Medium-Scale AI Applications: Data-Driven Arbitrator Selection

At the medium range of application, AI moves beyond administrative support and into a procedurally consequential role: arbitrator selection. Trust in the arbitrator is central to the legitimacy of arbitration. Parties expect decision-makers who are competent, independent, diligent, and attuned to the issues that matter most in a given dispute. Traditionally, the selection process has relied on professional networks, institutional rosters, published awards, and informal reputational knowledge. Increasingly, however, that process is being reshaped by data-driven and AI-enabled tools that aggregate, analyze, and systematize information about arbitrators' backgrounds, experience, procedural tendencies, and conflicts.

This development fits naturally within existing arbitral frameworks. Because arbitration is a creature of contract, the method of selecting arbitrators is dictated by party agreement. In *ad hoc* proceedings, parties must identify mutually acceptable candidates on their own. In administered arbitrations, institutions play a central role in this determination. Under the rules of the International Chamber of Commerce, the Court appoints arbitrators when parties cannot agree. *See*, 2021 Arbitration Rules – ICC – International Chamber of Commerce. By contrast, the American Arbitration Association and its international division, the International Centre for Dispute Resolution, rely on a list method. Based on party input, the institution generates a list of potential candidates drawn from its roster. It then provides résumés, conducts conflict checks, and ultimately appoints arbitrators based on party rankings. AI tools are now being introduced to assist with these functions by scanning résumés, identifying relevant expertise through keyword mapping, assessing availability and caseload, and flagging potential conflicts more efficiently than traditional manual review. *See*, 2026_ICDR_Arbitrator_Selection_Services.pdf. In this configuration, AI acts as a sophisticated filter and organizer, but human case administrators and parties remain responsible for interpreting the results and making the final selection.

For parties and counsel, AI-enabled arbitrator selection tools promise several advantages. Platforms that aggregate publicly available awards, enforcement decisions, professional writings, and institutional data can cast a wider net of potential candidates, including lesser-known or more diverse arbitrators who may not surface through informal networks. Tools developed by providers such as Jus Mundi, including its professional directory and conflict-mapping features, use extractive AI to organize arbitrators' experience by industry, issue area, language, and procedural role*. See,* Jus Connect | The professional network for arbitration. Institutions have followed suit. In 2024, the AAA-ICDR announced the beta launch of a generative AI–powered panelist search tool designed to broaden and deepen roster searches when preparing arbitrator lists. *See*, AAA-ICDR Launches AI Tool for Panelist Selection.

At this intermediate level, AI does not make the appointment decision. Humans remain in the loop and discern between varying options. But AI meaningfully shapes the pool from which choices are made. At the same time, medium-scale use of AI in arbitrator selection raises distinct concerns. International commercial arbitration is largely confidential. Awards are often unpublished, heavily redacted, or unreasoned. Even where institutions publish awards, such as the ICC's collaboration to publish selected awards through Jus Mundi, the available data is partial and curated. In three-member tribunals, outcomes may reflect compromise rather than the views of any single arbitrator, and dissenting opinions remain rare. As a result, the datasets on which AI tools rely are incomplete and, in some cases, misleading. Thus, predictive insights, drawn from such data, risk overstating precision and attributing philosophies or tendencies that cannot be reliably inferred.

Bias presents an additional challenge. AI systems trained on historical arbitrator data may reinforce existing patterns of repeat appointments and underrepresentation. International arbitration has long been criticized for a

lack of diversity, particularly along gender, racial, and regional lines. Because historical data disproportionately reflects appointments of senior, Western, male arbitrators, AI models trained on that data risk perpetuating the same imbalance. Efforts to promote diversity, such as the ICDR's stated diversity targets, may be undermined if AI outputs are treated as neutral or objective without scrutiny. *See*, Diversity | International Centre for Dispute Resolution. Moreover, many diversity-related attributes are not captured in professional biographies or résumés, limiting the ability of AI systems to account for them in any meaningful way.

Accordingly, medium-scale AI applications in arbitrator selection demand careful calibration. AI can improve efficiency, expand visibility, and systematize conflict analysis. It can assist both institutions and parties in managing increasingly complex datasets. But it cannot replace judgment, contextual understanding, or ethical responsibility. Decisions about who should decide a dispute remain normative, not purely technical. At this level of use, the promise of AI lies in informed assistance, not automation, and its legitimacy depends on transparency, critical interrogation of outputs, and continued human responsibility for the ultimate appointment decision. In other words, the technology may curate and analyze, but humans must still choose, thereby remaining in the loop.

III. Large-Scale Application: Can AI Be the Decision Maker?

At the far end of the spectrum, AI is no longer confined to support or select functions, but is embedded directly into the adjudicative process itself.

The most prominent example to date is the AI-led arbitration initiative launched by the American Arbitration Association, International Centre for Dispute Resolution (AAA-ICDR). *See,* AI Arbitrator, Fast and Fair Dispute Resolution by AAA. An "AI arbitrator," in this context, does not mean a system that independently decides disputes. Rather, it refers to an AI-enabled adjudicative framework designed to assist human arbitrators in delivering faster, more cost-effective, and transparent outcomes in narrowly defined cases, while preserving human judgment at every decisive stage. The AAA-ICDR AI arbitrator was developed for two-party, document-only construction disputes — a category of cases where claims, evidence, and legal issues are often structured and repeatable. *Id.* Its development relied on a curated legal and factual knowledge base created by AAA-ICDR attorneys and dispute-resolution experts, who annotated more than 1,000 prior construction cases. These materials include arguments, evidence, outcomes, and structured reasoning frameworks, forming a high-quality dataset intended to reflect how experienced arbitrators analyze claims rather than how a generic language model predicts text. *Id.* To operationalize this expertise, AAA-ICDR partnered with QuantumBlack, an advanced analytics and AI firm, which worked alongside arbitrators and industry specialists to codify arbitral reasoning into prompts and agents. *Id.* AAA Construction Panel arbitrators and AAA-ICDR attorneys were involved from the outset, ensuring that the system follows real arbitration logic step by step, rather than abstract or purely statistical inference. *Id.* Under the AAA-ICDR AI-Led Arbitration Rules, the process is structured and transparent; parties submit their claims and supporting evidence through the platform; the AI system produces summaries of the submissions, which the parties must validate for accuracy; the AI then analyzes the claims, reviews the evidence, applies the relevant law, and drafts a proposed award. *Id.*

Crucially, that draft never becomes final on its own. A human arbitrator, trained specifically for AI-led cases and vetted through the same institutional processes as any other AAA appointment, reviews the AI's analysis, has access to the full record, makes any necessary revisions, and issues the final, binding award. *Id.* Every AI-led case therefore involves human adjudication at the decisive moment. The system is entirely opt-in, requiring the

express consent of both parties; absent agreement, the dispute proceeds under traditional AAA arbitration. *Id.* Accordingly, early use, indicates cost savings in the range of approximately 35-45% compared to traditional documents-only arbitration, with time savings of roughly 20-25%. *Id.* Despite these safeguards, large-scale AI use raises concerns distinct from those associated with smaller applications. Practitioners and clients may question whether reliance on AI, however supervised, affects perceptions of neutrality, due process, or legitimacy. Others may worry about transparency in reasoning, the scope of disclosure obligations, or whether the use of AI should be affirmatively disclosed to preserve enforceability.

These concerns are not merely theoretical. They intersect directly with how courts and enforcement bodies may view awards produced through AI-assisted processes, particularly across jurisdictions with differing regulatory approaches to automated decision-making. In this respect, the EU's approach under the EU Artificial Intelligence Act is instructive. *See,* The AI Act Explorer | EU Artificial Intelligence Act. The regulation emphasizes that individuals should not be subject to decisions producing legal effects, or similarly significant consequences, where those decisions are based solely on automated processing. Article 14 requires meaningful human oversight, ensuring that automated systems do not operate autonomously in legally binding decision-making. *See,* Article 14: Human Oversight | EU Artificial Intelligence Act. Automated decision-making is permitted only in limited circumstances, such as necessity for contract performance, express legal authorization, or explicit consent, and even then, must be accompanied by safeguards, including transparency about the logic involved, the right to human intervention, and the ability to contest the outcome. *Id.*

Although arbitration is contractual and party-driven, these principles underscore a broader regulatory sensitivity to systems that might be perceived as replacing human judgment in legally consequential decisions. Against that backdrop, the AAA-ICDR model is designed to align with emerging regulatory and enforcement expectations. Human legal judgment remains central, and the arbitrator issuing the award is identifiable, accountable, and subject to the same ethical obligations, disclosure duties, and institutional oversight as in any other AAA proceeding. A dedicated AI Governance Committee oversees compliance, ethics, and model performance.

Whether national courts will treat AI-assisted awards differently at the enforcement or vacatur stage remains an open question. What is clear, however, is that large-scale AI use in arbitration will be scrutinized through the lens of consent, transparency, human oversight, and procedural fairness. In that sense, AI-led arbitration does not eliminate traditional concerns about legitimacy, it reframes them, placing renewed emphasis on how technology is integrated, supervised, and disclosed within the arbitral process. Future models that move closer to fully automated adjudication are likely to collide with these requirements unless they retain a clearly identified human in the loop capable of explaining, revising, and ultimately owning the decision.

IV. A Human in the loop, the future of AI in Arbitration

The phrase "human in the loop" began as a technical way to describe a person inside a system's chain of action and response.[2] The "loop" is the cycle: information comes in, a decision is made, something happens, and the result feeds back into the next decision. The term stuck because it captures a powerful idea: the human is not merely watching the machine, but sensing, judging, correcting, and sometimes overruling it. Today, in automation and AI, it points to a larger question: where should human judgment still enter when machines can act on their own? The question is no longer whether artificial intelligence will play a role in arbitration, but how far that role will extend.

At the limited end of the spectrum, AI tools that assist with document organization, bundle preparation, and legal research are already embedded in arbitral practice. At an intermediate level, data-driven tools are reshaping arbitrator selection by aggregating experience, mapping conflicts, and expanding visibility beyond familiar networks. At the most ambitious end, initiatives such as the AAA-ICDR's AI-led arbitration framework demonstrate that institutions are prepared to experiment with AI at the adjudicative core of the process. Taken together, these developments suggest that AI-enabled arbitrator tools are likely to become a permanent feature of institutional arbitration. If that is the direction of travel, the arbitration community must be prepared. Preparation, in this context, means not only learning how to use AI tools, but also articulating where human control, responsibility, and explanation must remain non-negotiable.

Those preparations begin with restraint. The efficiency gains offered by AI, speed, cost reduction, and analytical support, are real, but they do not diminish the need for diligence, verification, and professional judgment. The documented incidents of hallucinated authorities, and sanctioned filings, underscore that even limited AI use can undermine legal credibility, if outputs are treated as authoritative rather than provisional. In arbitration, where confidentiality and accuracy are paramount, AI must be approached as an assistive technology, not a substitute for human responsibility. As AI assumes a more structural role, particularly in arbitrator selection, its influence becomes normative as well as practical. Data-driven tools can broaden candidate pools and support more informed appointments, but they also operate on incomplete and historically skewed datasets. Without careful governance, they risk reinforcing repeat appointments and existing disparities rather than mitigating them. Transparency about data sources, critical interrogation of outputs, and continued human discretion, remain essential to preserving party autonomy, and confidence in the process. These are all dimensions of a human-in-the-loop architecture, in which people remain answerable for how AI is used and for the choices it ultimately informs.

Large-scale applications, such as AI-led arbitration, bring these issues into sharp relief. Models that embed AI in the reasoning process, while preserving human oversight, party consent, and institutional accountability, reflect an effort to reconcile innovation with due process. Yet they also invite scrutiny at the enforcement stage, particularly in light of emerging regulatory frameworks like the EU Artificial Intelligence Act that emphasize meaningful human oversight and limit legally binding decisions based solely on automated processing. The enforceability of AI-assisted awards will ultimately turn not on the presence of AI, but on whether human judgment, transparency, and procedural fairness remain demonstrably intact. If major arbitral institutions do move toward widespread use of AI arbitrator tools, the legitimacy of arbitration will depend less on technological sophistication than on governance.

Clear procedural rules, informed consent, disclosure where appropriate, robust oversight, and professional education will determine whether AI strengthens arbitration's core values or strains them. Arbitration has always evolved through adaptation. The challenge is to close the loop now to ensure that, as AI becomes part of the arbitral architecture, it is integrated in a way that enhances efficiency without compromising fairness, enforceability, or trust in the arbitral process. Put differently, the sustainable future of AI in arbitration is one in which humans remain visibly, accountably, and meaningfully in the loop.

---

[1] *See, e.g.*, Cole Stryker, *What is human-in-the-loop?*, IBM.com, https://www.ibm.com/think/topics/human-in-the-loop.

[2] *See* Cole Stryker, *What is human-in-the-loop?*, IBM.com, https://www.ibm.com/think/topics/human-in-the-loop;

Ge Wang, *Humans in the Loop: The Design of Interactive AI Systems*, hai.standford.edu (October 19, 2019), https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems; Kim Herrington, *Be THE Human In The Loop: Data And AI Literacy Is Your Edge*, Forrester.com (March 27, 2025), https://www.forrester.com/blogs/be-the-human-in-the-loop-data-ai-literacy-is-your-edge/.

** Matteo Rubolino, visiting foreign law clerk, and Alex Maria Conway, previous summer associate, who is not licensed to practice law in any jurisdiction, contributed to this article.*

RELATED INDUSTRIES + PRACTICES

- Artificial Intelligence
- Business Litigation
- Construction
- International Arbitration