

Maharaja Surajmal Institute Law Journal

Vol. 3, No. 1, January–June 2026

Transparency, Accountability, and Legitimacy in Ai-driven Arbitration: A Legal Analysis

Rachit Sharma¹ and Shivpriya k²¹Research Scholar, Amity Law School, Amity University, Gurugram²Assistant Professor, Amity Law School, Amity University, Gurugram<https://doi.org/10.48165/msilj.2026.3.1.4>

ARTICLE INFO

Article history

Received: 03.07.2026

Revised: 23.07.2026

Accepted: 21.08.2026

Keywords

expanding
deployment
algorithmic
arbitration

ABSTRACT

The expanding deployment of algorithmic systems within arbitration signifies a paradigm shift across modern dispute resolution. Moving beyond administrative case handling into predictive tools and automated draft generation, digital mechanisms increasingly execute tasks historically entrusted solely to human adjudicators. Although advocates cite enhanced procedural speed, lower expenditures, and decision-making uniformity, this shift creates significant legal and normative tensions—most notably concerning transparency, liability, and the underlying legitimacy of automated determinations.

This study critically evaluates the structural opacities inherent in automated dispute resolution mechanisms, concentrating on the “black box” dilemma where complex algorithmic operations remain unexplainable to participating litigants. We argue that such procedural obscurity poses a direct challenge to core procedural safeguards and due process guarantees, which form the normative bedrock of private adjudication. Furthermore, the analysis scrutinizes existing responsibility structures under municipal and international frameworks—including the UNCITRAL Model Law alongside key institutional regulations—demonstrating their present insufficiency in mitigating technology-specific risks.

Employing a combined doctrinal and comparative research design, this study examines evolving regulatory paradigms globally, including the European Union’s Artificial Intelligence Act and recent institutional directives, to formulate a rights-based governance framework for algorithmic integration in dispute resolution. Ultimately, the paper demonstrates that transparency and interpretability must function as mandatory procedural prerequisites, proposing compulsory notification regimes,

BACKGROUND

Traditionally, arbitration has enjoyed a unique role in resolving disputes between merchants, and more recently, investors and consumers, because of its flexibility, party autonomy, confidentiality, and the ability to enforce awards in light of instruments like the 1958 New York Convention.¹ The legitimacy of the private adjudication process has historically been based on a

set of procedural protections, known as natural justice or due process, including the right to have equal opportunity of being heard, the right to a reasoned award, and the assurance that the decision-maker is independent and impartial. These procedural rights are no small part of what makes arbitration legitimate; rather, they represent the normative basis of arbitration, the reason the courts can recognize and enforce these awards despite being subject only to very limited judicial review.

In the last ten years, AI has shifted from being a marginal player in the administration of arbitration to playing roles traditionally reserved for human ar-

¹United Nations. (1958). Convention on the Recognition and Enforcement of Foreign Arbitral Awards (New York Convention).

bitrators and counsel. Arbitrators and arbitration institutions have started using AI tools for document review and e-discovery, for making predictions about the likely outcome or quantum of the dispute, for case and scheduling management, for legal research and precedent finding, and in advanced applications, even for drafting portions of procedural or substantive awards. ODR providers who handle high volume, low value consumer and e-commerce disputes have gone yet further, relying on algorithmic systems to screen complaints, suggest ranges for settlement, and make decisions in some cases.²

The motivation for this development results from actual and documented challenges facing the arbitral system, namely, high costs, prolonged delays, inconsistency in the decisions made by different tribunals, and increased numbers of low-value cases that render traditional arbitration unaffordable. Proponents of using AI in this context emphasize the ability of algorithms to reduce costs, accelerate decision-making, increase consistency by reducing idiosyncrasies of individual arbitrators, and expand access to dispute resolution for financially disadvantaged participants. It would be wrong to underestimate the importance of these positive aspects since they do not form the focus of the current discussion.

However, conferring decision-making or quasi-decision-making powers on algorithms raises certain legal and ethical problems which are not addressed by arbitration law. In the first place, one can name what is called the “black box” problem, meaning the inability of many modern AI systems to provide a clear rationale for their decisions since they use opaque algorithms of machine learning or deep neural networks. In cases where such an algorithm impacts the final outcome in one way or another (for instance, by evaluating evidence, ranking arguments, or preparing a decision which the arbitrator later endorses), it becomes impossible for the involved parties and the courts to understand the reasons for this particular decision.³

The other problem connected to transparency is that of accountability. Humans involved in arbitration are held accountable through their duties of disclosure, codes of conduct, challenge mechanisms and, ultimately, potential liability.⁴ By contrast, algorithm-

based systems and their vendors or implementing institutions have not yet faced a comparable accountability framework. In the case when an AI-based tool significantly contributes to the award it is quite difficult to say who bears responsibility for the contribution: the arbitrator, the institution which contracted the service or the developer of the technology. The current legal instruments like the United Nations Commission on International Trade Law (‘UNCITRAL’) Model Law on International Commercial Arbitration and the institutional rules of such major institutional arbitration centres as the International Chamber of Commerce, the London Court of International Arbitration, and the Singapore International Arbitration Centre were written without any consideration of the algorithmic decision support and as such do not provide any provisions in regard to disclosure of the use of AI, testing or auditing of such tools and the standard of human supervision required for the adoption of an AI-supported award.⁵

In general, all three above mentioned issues lead to one further issue of fundamental importance: the legitimacy of the AI-supported award. Legitimacy of the adjudication is not just an ability of enforcement, it is a sociological characteristic of the procedure that rests upon the faith of the parties and society in the fairness of the process and the involvement of reasoned arguments in the dispute as well as in the human authority standing behind the outcome.⁶ In the situation when parties cannot understand or investigate the ground of the decision made with the help of the black box algorithm and there is no accountability mechanism of possible errors or bias contained within this algorithm – the legitimacy of the whole process is genuinely at stake despite the ability to enforce an award.

The paper takes a position in the contemporary discourse of academic scholarship. This discourse builds on the premise that the key question at this stage is not whether artificial intelligence (‘AI’) would be used in arbitral proceedings but how its use could be reconciled with the processes and values of arbitration as an acceptable alternative to litigation. The present research revolves around a detailed analysis of the applicable legal norms and comparative study of regulatory approaches with a specific focus on EU Artificial Intelligence Act and guidance issued by arbitral institutions and professional associations. The aim of the analysis is to determine whether existing frameworks provide

-of-ai-in-arbitration_updated-sept-2025.pdf

⁵Norton Rose Fulbright. (2025). The Chartered Institute of Arbitrators’ new guidelines on use of AI in Arbitration.

⁶ADR Institute of Canada. (2023, March 2). Artificial intelligence and arbitration: A perfect fit? <https://adric.ca/artificial-intel-ligence-and-arbitration-a-perfect-fit/>

²Faegre Drinker Biddle & Reath LLP. (2026, February). Use of AI in arbitral institutions.

³Wolters Kluwer Arbitration Blog. (2024, August 23). Arbitration tech toolbox: AI as an arbitrator: Overcoming the “black box” challenge? <https://legalblogs.wolterskluwer.com/arbitration-blog/arbitration-tech-toolbox-ai-as-an-arbitrator-overcoming-the-black-box-challenge/>

⁴<https://www.ciarb.org/media/bpndtcgu/guideline-on-the-use>

adequate regulation and, in their absence, to develop reform proposals based on a concept of due process in terms of human rights protection.

The importance of the question is enhanced by the variety of arbitral practices emerging through adoption of AI. First of all, there are complex, fact-intensive disputes involving predictive analytics and AI-assisted drafting tools used in international commercial and investment arbitration by resourceful arbitral institutions and law firms. By contrast, high-volume ODR services dealing with consumer, e-commerce and small claims arbitrations are subject to commercial pressure and use automation because of its lower cost compared with human review. In the latter type of cases, the problem of risk discussed herein is especially pronounced because the economic necessity of minimizing the amount of human review is the greatest where the parties have neither the money nor the legal expertise to challenge an incorrect and opaque result. Thus, an AI regulation framework in arbitration should take into account this divide.

OBJECTIVES

There are five interrelated goals which will be pursued through this research. First, this research seeks to identify not only contemporary but emerging uses of artificial intelligence during the arbitral process, distinguishing between administrative or supporting uses of the technology, such as case management and scheduling, on the one hand, and more substantive or quasi-substantive ones, such as predictive analytics to facilitate settlement discussions and tools for the preparation of awards whose outputs the arbitrator can accept with little further editing, on the other. The distinction is considered important since the legal risk of using AI is directly related to the level of its influence on the final result.

The second goal is to analyze the black box problem as a legal, not only technical issue. The paper will analyze how the lack of transparency affects algorithmic decision-making processes and interacts with the legal requirement of natural justice, specifically with the requirement of hearing and reasoned award, and whether an award prepared through an opaque process satisfies those requirements even if the human arbitrator put his signature on it.

The third objective of the paper is to conduct the analysis of available legal accountability mechanisms established by national and international arbitration laws, as well as arbitral institutions and the UNCITRAL Model Law to see if these mechanisms are able to deal

with the issue of attribution of liability where an arbitral mistake or bias was caused by the AI tool.

Finally, the paper intends to study regulatory trends outside the realm of arbitration in particular the EU Artificial Intelligence Act with its risk-based approach to regulation and guidance provided by arbitral institutions and professional organizations in order to find transferable concepts from this area of law.

Lastly, the paper aims at developing a governance framework concerning the deployment of artificial intelligence technologies in arbitration based on the human rights principle of a right to transparency and explainability. The emphasis is laid on the principle of mandatory disclosure, explainability of algorithms proportional to the importance of the decision, algorithmic audit and testing, as well as clear accountability of the institution and individuals involved. The intention is not to describe but to advance a normative thesis that the transparency and explainability of the processes in which arbitration uses AI tools should be understood as a legal principle rather than best practice. Such view finds its justification in the commitment to the due process principles entrenched in arbitration law.

METHODOLOGY

This research applies a doctrinal and comparative method enriched with interdisciplinary insights provided by literature devoted to explainable artificial intelligence from the technical standpoint where appropriate. The doctrinal method involves an analysis of primary sources of law, including the UNCITRAL Model Law on International Commercial Arbitration and the New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards; the rules of major institutions. Such sources are analyzed for the provisions regulating the use of technology, mandatory disclosure requirements, the standard of reasoning in arbitral awards, and the bases for refusal or opposition to which the recognition and enforcement of an award may be challenged on procedural and public policy grounds.

The comparative analysis looks at regulatory measures and soft law guidance, even if they are not directed at arbitration per se, but present general principles of governance for AI systems, applicable to the arbitral context. The European Union's Artificial Intelligence Act is analyzed with particular attention due to its risk-based categorization of AI systems which imposes differentiated obligations depending on the pos-

sible effects of the use of such a tool on basic rights⁷ thus providing a structurally convenient tool for thinking about different ways of regulating AI tools used in arbitration from purely administrative to substantially decision-making. Also, arbitral institutions' and professional organizations' guidance documents and working papers related to the present use of AI technologies in dispute resolution are analyzed to the degree they are publicly available in order to understand the tendency of practice development.

Additionally, in the course of the research, secondary literature from the sphere of computer science and information systems related to explainable AI, algorithmic auditing, and limits of explainability of complex algorithms is considered. While this literature does not serve as an independent empirical basis for this research, it helps to determine how to analyze the issues of law in terms of the possibilities of technology, namely which kinds of explainability of AI algorithms are technically possible and hence reasonable to impose as a legal obligation.

Due to the confidentiality concerning the risks and uncertainty associated with the use of AI technologies in arbitral cases and lack of public information about how institutions practice, there is a lack of empirical information regarding how institutions utilize the technology. Therefore, it should be noted that this paper does not aim at providing an empirical analysis of the utilization of AI today. In this regard, the information gathered through institutions' public declarations and pilot projects as well as practitioner opinions was used in order to produce a reasonable representation of how this technology is used. This limitation is considered significant and will be discussed further in the conclusion along with suggestions for empirical analysis when institutions become more transparent, which is one of the suggested reforms.

This paper is not based on the juristic approach according to jurisdictions but it is structured in a way to highlight the interconnections among these three concepts by following the logic from transparency to accountability and then to legitimacy.

RESULTS

AI's Black Box Problem and the Right to have a Reasoned Award

Foregoing analysis reveals that the main problem of transparency of AI arbitration is not connected with using artificial intelligence as such but rather with unique features of existing machine learning architectures. Rule-based and strictly statistical methods which were traditionally used to aid litigation allowed for identification of the chain of logic that led to the particular result. However, deep learning machines, including language models currently used for drafting assistance and legal research, operate in a way that involves millions and even billions of weighted parameters which do not correspond to any single human logic chain. Although post hoc explanatory tools may provide for some sort of justification, they usually do not reflect the actual algorithm process. While this distinction between rule-based and deep learning tools is well recognized in technical sources, it still has not found proper reflection in arbitral practices and doctrines.

This point is important because a right to reasoned award, being intimately linked to the right to be heard, is not satisfied by the mere issuance of a document that appears to be a reasoned award. Doctrinal criterion for satisfaction of the right includes real consideration of the arguments and evidence provided by the parties in order for the parties to understand why they won or lost the dispute and, accordingly, be able to challenge the award on certain grounds. If a considerable part of the reasoning behind the award is created through the use of an opaque algorithm that is not understood by the human arbitrator, the award will satisfy the form of reasonable decision-making, but not its substance.

Accountability Gaps in Existing Arbitration Law

A comparison of the UNCITRAL Model Law, the New York Convention, and other institutional rules reveals that none of these texts includes provisions dealing specifically with AI-related accountability.⁸ The provisions of the UNCITRAL Model Law and the grounds for refusal of enforcement of awards under the New York Convention, including grounds related to the inability to present one's case and the public policy grounds, can be understood broadly.⁹ In theory, such provisions may be used against an award affected by AI; however,

⁷artificialintelligenceact.eu. (n.d.). High Level Summary of AI Act. Future of Life Institute. Retrieved July 2, 2026, from <https://artificialintelligenceact.eu/high-level-summary/>

⁸United Nations. (1958). Convention on the Recognition and Enforcement of Foreign Arbitral Awards (New York Convention). https://uncitral.un.org/en/texts/arbitration/conventions/foreign_arbitral_awards

⁹Wolters Kluwer Arbitration Blog. (2026, April 11). Can AI assisted arbitral awards survive enforcements under the NY Convention? <https://legalblogs.wolterskluwer.com/arbitration-blog>.

their application to AI-related issues is not yet clear and requires interpretation. Additionally, there is no institutional rule reviewed herein which provides for any separate duty on arbitrators and arbitral institutions to reveal whether AI technologies have been used in the preparation of procedural decisions or awards, nor are there any institutional rules that provide for any standard of independent review which an arbitrator would have to meet before using AI-generated material.

Such a discontinuity results in a diffusion of responsibility identified by the paper as the main problem of accountability in current practice. Four sources of accountability are discussed: the arbitrator, the arbitral institution, the AI provider, and the parties who may have agreed to use AI assistance or asked for such assistance. According to the law, the formal responsibility for the award lies with the arbitrator who signs it. However, formal accountability is inconsistent with a substantive accountability mechanism if the process is either not disclosed or not disclosable — for example, because the arbitrator was unaware of the extent to which AI tools procured by the institution were used. As for the institution, it is usually not accountable for the content of awards because such a responsibility lies with the tribunal while the institution controls the procurement and use of the AI tool. Providers of the tools work under commercial agreements with the institutions but not under public law duties or the responsibilities of the arbitral law towards the disputing parties. The disputing parties do not have the means to demand disclosure or assessment of the tools used by the institution or the tribunal.

Regulatory Developments: The EU AI Act and Institutional Guidance

As such, the European Union's Artificial Intelligence Act may serve as the most sophisticated regulatory framework presently existing regarding the regulation of AI systems affecting legal rights. The risk categorization scheme of the Act, which distinguishes between four types of AI systems, including unacceptable-risk, high-risk, limited-risk, and minimal-risk systems, each of which is associated with respective obligations, may be regarded as structurally transferable to arbitral context,¹⁰ even though falls within the scope of application of the act and the systems involved in it pose serious unresolved jurisdictional issues not falling within the purview of this paper. An AI system used to affect the decision of a dispute might be classified as a high-risk system, which would impose certain specific

requirements relating to risk assessment, data governance, technical documentation, human supervision, and post-market monitoring. These obligations would coincide with the identified gaps of transparency and accountability discussed above.

As for the guidance provided by arbitral institutions and organizations, it should be noted that they are at a relatively early stage of development and do not possess binding character. They reflect a general consensus on certain soft-law principles, according to which parties should be notified about the use of AI to assist in creating submissions and procedural documents;; arbitrators will remain responsible for the awards regardless of their reliance on AI assistance; confidentiality and data protection obligations apply to data handled by AI. Nevertheless, it can be seen that the principles mentioned have a declarative character since no major organization has adopted any mandatory disclosure obligations, algorithmic audits, and human supervision standards enforceable through the set-aside procedure.¹¹

Toward a Human Rights-Centered Governance Framework

In contrast to prior analysis, in this paper, we propose a framework of governance with four key pillars. First, it is mandatory disclosure. Parties should know prior to the start of proceedings about any AI tool used for managing case, assessing evidence, or drafting procedural orders or award. We measure materiality by its impact on the result of proceedings, not administrative convenience. Second, we propose a system of graduated explainability where the level of explanation of the AI tool output correlates with its decisional weight. For example, there is no need in explanation of purely administrative tools, however, for tools that have material impact on finding of facts, assessment of credibility, or legal conclusions, we would like to require a level of explainability that will make it possible for the arbitrator to independently verify output, and in case of challenge, for courts to be able to reconstruct it. Third, there is a necessity of algorithmic audit and testing where AI tools performing tasks with high decisional weight are audited and tested on a periodic basis for accuracy, consistency, and absence of bias, and then summarized information about these audits is shared both with institutions and in some form, with the parties based on audit and post-market surveillance provisions for high-risk category from the EU AI Act. Fourth, we define institutional and individual accountability

¹⁰artificialintelligenceact.eu. (n.d.). Retrieved July 2, 2026, from <https://artificialintelligenceact.eu/high-level-summary/>

¹¹Stockholm Chamber of Commerce. <https://sccarbitrationinstitute.se/en>

where arbitral institutions will be responsible for procurement and monitoring of AI tools used in their proceedings, while arbitrators will remain accountable for their independent verification of any material AI output.

In this regard, this paper makes an attempt to offer a human rights-based argument in favor of the proposed approach rather than an exclusively efficiency-oriented justification based on the idea that these rights are not means but rather constitutive elements of legitimate decision-making processes. Under the proposed approach, obligations of transparency and explainability are not regulatory constraints to be reduced to a minimum but important legal requirements comparable with already established principles of due process. Hence, this paper argues that obligations of transparency and explainability are things to be regulated by institutional rules and further interpretation or amendment of instruments such as the UNCITRAL Model Law.

JURISDICTIONAL DIVERSITY AND JUDICIAL REVIEW

The paper suggests a number of differences in the way in which courts are likely to deal with challenges to arbitral awards concerning the use of artificial intelligence even though there is a small number of reported cases in this area. The jurisdictions that adhere to the more liberal approach to the ground of public policy used to refuse to enforce awards are in a much better position to review AI-related awards as are jurisdictions that conduct thorough investigation into whether the parties really had an opportunity to present their case. In turn, jurisdictions that restrict the use of these grounds because of the finality of arbitration are less likely to investigate AI-assisted awards.

Such a difference confirms the thesis of the paper according to which reform must be achieved via modification of the institutional rules and even via collective interpretation or amendment of the Model Law, but not via piecemeal case law development, which is often reactive and inconsistent. Institutional rules tie down arbitrators, and thus arbitral awards made by the latter are bound from the start with their duty to disclose, explain, and audit information. In this way, reliance on ex post judicial review, which is unpredictable and delayed and thus may not save a party who did not get a hearing, is avoided.

CONCLUSION

Despite all the positive effects that the incorporation of AI into arbitration processes may have, including efficiency gains and enhanced access to justice, structural problems of lack of transparency and accountability exist, which are not properly addressed by contemporary arbitration law. The “black box” problem inherent to the modern machine learning systems is not a technological problem alone, but rather the doctrinal one that threatens fundamental arbitral principles of rights to be heard and a reasoned award.

From the above comparative review, it becomes clear that both the UNCITRAL Model Law and the New York Convention, along with the rules of major arbitral institutions, do not specify any AI-specific duties. This gap results in a diffusion of accountability between arbitrators, arbitral institutions, and vendors, thereby preventing an identification of a party responsible for errors or biases that arise because of the use of algorithms in an award. At the same time, the evolution of regulation outside of arbitration field, especially the European Union Artificial Intelligence Act, along with evolving non-binding guidance governing the use of artificial intelligence by arbitral institutions, demonstrates that concepts of practical governance such as risk-based categorization, mandatory disclosures, and algorithmic auditing are easily applicable to arbitration practice.

Based on the above discussion, this paper offers a governance model based on mandatory disclosure, graduated explainability, algorithmic auditing and testing, and clear individual and institutional accountability. This model assumes that transparency and explainability are not only best practices but mandatory requirements that correspond to the due process principles upon which arbitration was traditionally regarded as a valid way to resolve disputes. Without such governance reforms, it is possible to expect an emergence of awards that, while being formally compatible with the current conventions on enforcement of awards, would be substantively unrelated to reasonable human decisions.

The limitations of the paper include the limits imposed by carrying out doctrinal and comparative analysis without empirical evidence concerning the practical implementation of AI within private arbitrations. Future studies should consider carrying out empirical analysis of the use of AI within different institutions and types of disputes, and examine whether the implementation of any disclosure, explainability, and auditing obligations is functioning properly at all. Comparisons between different jurisdictions, other than those covered in this paper, particularly the examination of

the role of national courts when setting aside or enforcing arbitration awards due to procedural flaws connected to the use of AI, will contribute to developing the framework presented in this paper even more. Despite these limitations, the main argument made in this

paper aims to withstand in the face of future technological development: for arbitration to be legitimate in the age of artificial intelligence, transparency and accountability must be treated as mandatory rather than optional elements of the arbitral process.