Artificial Intelligence Hallucinations in Arbitration: Analyzing the Landmark Award Annulment in ARIHQ c. Santé Québec

May 4, 2026

The advent of generative artificial intelligence (AI) has brought unprecedented capabilities and efficiencies to the legal profession. However, alongside its benefits come profound risks, particularly when AI is employed in adjudicative decision-making. The recent landmark ruling by the Quebec Superior Court in *Association des ressources intermédiaires d'hébergement du Québec (ARIHQ) c. Santé Québec* (2026 QCCS 1360) serves as a stark warning to the international and domestic arbitration communities. In an unprecedented move, the Court annulled an arbitral award after finding that the sole arbitrator had implicitly delegated his decision-making authority to a generative AI tool, which resulted in the inclusion of "hallucinated" jurisprudence and legal doctrine.

This article unpacks the *ARIHQ* judgment, explores the comparative international grounds for setting aside such awards under the New York Convention and the UNCITRAL Model Law, surveys the growing global judicial consensus on AI, and proposes practical safeguards for inclusion in arbitration agreements and Terms of Reference.

The Judgment: *ARIHQ c. Santé Québec*

The dispute originated from an unpaid retroactive remuneration

claim between a healthcare intermediary resource (Osman) and a Montreal health authority (CCSMTL, now Santé Québec). An arbitrator dismissed the claim on preliminary grounds, ruling that the contractual notice periods had expired. The claimants sought to annul the award before the Quebec Superior Court, relying on two grounds: first, that the award violated public policy by modifying statutory prescription periods; and second, that the arbitral procedure was not respected because the arbitrator relied on non-existent legal authorities, strongly suggesting the use of generative AI.

The Honourable Martin F. Sheehan swiftly dismissed the public policy argument, affirming the well-established principle that a mere error of law—even concerning public policy provisions—does not justify setting aside an arbitral award unless the outcome itself is fundamentally repugnant to public order.

However, the Court decisively upheld the procedural challenge. Reviewing the award, the Court verified that the central doctrinal and jurisprudential authorities cited by the arbitrator to support the validity of the forfeiture clause simply did not exist. The arbitrator had referenced a fabricated doctrinal article by a real author and hallucinated judgments ostensibly from the Quebec Court of Appeal and the Superior Court (e.g., fictitious citations for *Ville de Montréal c. Syndicat des cols bleus* and *Tremblay c. Commission scolaire de la Jonquière*).

The Court emphasized the *intuitu personae* nature of arbitration—the principle that an arbitrator is chosen for their personal expertise, judgment, and authority. Relying on an AI tool to formulate the substantive reasoning without verifying the output violated the fundamental maxim *delegatus non potest delegare* (a delegate cannot delegate) and breached the secrecy of deliberations.

Translating the crucial analytical and operative parts of the

judgment, Justice Sheehan concluded:

"The preponderance of the evidence therefore leads to the conclusion that the Arbitrator's authority was delegated and that he abdicated his role to review the result. This conclusion is inevitable given that all of the doctrinal and jurisprudential references on which the Arbitrator relies are non-existent and 'hallucinated'. For this reason, the Award must be annulled."

The Court, however, added a critical caveat: not every use of AI will automatically lead to annulment. If the AI usage is minimal, or if the hallucinated references do not form the crux of the reasoning, the award might survive if the breach did not fundamentally taint the proceedings. But in this case, the breach struck at the very heart of the arbitrator's reasoning and affected the confidence of the parties in the arbitration regime.

The operative part (*dispositif*) reads:

"FOR THESE REASONS, THE COURT: ANNULS the arbitral award rendered on August 8, 2025, by the impleaded party, Mr. Michel A. Jeanniot; ORDERS the parties to choose a new arbitrator within 60 days of this judgment; THE WHOLE, with legal costs."

Domestic Set-Aside Grounds and Comparative International Frameworks

The Quebec Code of Civil Procedure (C.C.P.) is deeply influenced by the UNCITRAL Model Law, making this judgment highly instructive for international practitioners. The Court annulled the award under Article 646(3) of the C.C.P., which allows set-aside where "the applicable arbitral procedure was not respected."

The UNCITRAL Model Law and the New York Convention

Under Article 34(2)(a)(iv) of the UNCITRAL Model Law, and

identically under Article V(1)(d) of the New York Convention, an award may be set aside or refused enforcement if "the arbitral procedure was not in accordance with the agreement of the parties."

When parties agree to arbitrate, they contract for a bespoke dispute resolution process wherein a specific human tribunal exercises independent intellectual judgment. The use of an unverified Large Language Model (LLM) to draft the award's substantive reasoning is a structural deviation from this agreed procedure. Furthermore, feeding confidential submissions into a public AI tool violates the confidentiality and secrecy of deliberations inherent in arbitration.

The Public Policy Exception

While the Quebec Court dismissed the substantive public policy argument, an award heavily reliant on AI hallucinations could trigger the procedural public policy exception under Article 34(2)(b)(ii) of the Model Law and Article V(2)(b) of the New York Convention. An award anchored in fabricated law deprives the losing party of the right to a fair hearing and the opportunity to present its case (*audi alteram partem*). The parties cannot possibly address or distinguish phantom precedents during the proceedings. Enforcement courts in major hubs like London, New York, or Singapore would view the abdication of the adjudicative function to a "black box" algorithm as a severe affront to fundamental notions of justice, procedural fairness, and due process.

Global Court Directions on AI Usage in Proceedings

The *ARIHQ* judgment resonates with an emerging global judicial consensus that demands strict guardrails around the use of generative AI in dispute resolution. Courts and tribunals worldwide have recently confronted the consequences of "AI hallucinations" and have responded with firm directives.

United States

The U.S. has been ground zero for AI in litigation, notably following the infamous *Mata v. Avianca* case, where lawyers were sanctioned for submitting AI-generated fictitious case law. Consequently, numerous federal and state judges have issued standing orders. These orders typically mandate that counsel file a certificate attesting that no generative AI was used in drafting, or that if it was, a human attorney rigorously verified the accuracy of all citations and legal propositions. The focus is squarely on human accountability and preventing the waste of judicial resources on fabricated jurisprudence.

United Kingdom

In late 2023, the UK Courts and Tribunals Judiciary issued official guidance for judicial office holders. While acknowledging the utility of AI for administrative tasks, the guidance explicitly warns judges about the severe risks of hallucinations and deepfakes. The UK Master of the Rolls has emphasized that judges remain personally responsible for their judgments and must not delegate legal analysis or substantive drafting to AI tools, safeguarding the integrity of the "judicial mind." Furthermore, it firmly warns against entering any confidential case information into public AI chatbots.

Middle East (QFC and ADGM)

Progressive commercial courts in the Middle East have taken decisive action against AI misuse. In the Qatar Financial Centre (QFC), the Civil and Commercial Court's late-2025 judgment in *Sheppard v Jillion LLC* found a lawyer in contempt for repeatedly citing AI-generated fake cases. This prompted the QICDRC to issue Practice Direction No. 1 of 2026, which established a comprehensive framework requiring lawyers to independently verify AI-generated submissions. Similarly, in the Abu Dhabi Global Market (ADGM), the Court in *Arabyads v Gulrez Alam* [2025] ADGMCFI 0032 ordered severe wasted costs against a legal team that submitted a defense riddled with

false, AI-generated legal authorities, concluding that unverified AI use amounts to reckless professional conduct.

Proposed Directions for Terms of Reference and Arbitration Agreements

The *ARIHQ* decision underscores a critical vulnerability: waiting for a court to set aside a tainted award means the parties have already suffered the immense costs and delays of a compromised arbitration. To proactively mitigate these risks, parties should expressly regulate the use of AI in their Arbitration Agreements, or more practically, in the Terms of Reference (ToR) and Procedural Order No. 1 (PO1).

Practitioners should consider integrating the following AI protocols:

- **Explicit Prohibition of Delegation:** The ToR should clearly state that the arbitral tribunal must not delegate its adjudicative, analytical, or substantive decision-making functions to any generative AI tool. The arbitrator must assume full, personal responsibility for the reasoning, legal research, and drafting of the award.
- **Permitted Uses and Mandatory Verification (Human-in-the-Loop):** Parties can define acceptable AI use, such as utilizing AI for translation, formatting, or document summarization, provided it does not replace substantive analysis. PO1 should include an affirmative obligation for both the tribunal and counsel to strictly verify any AI-assisted output against primary legal sources. The tribunal could be required to issue a "Statement of Human Authorship" within the final award confirming that all citations and factual references have been manually verified.
- **Confidentiality and Data Security Guardrails:** To protect the secrecy of deliberations and party confidentiality, the ToR must prohibit the uploading of confidential case

information, pleadings, or evidence into open-source or publicly accessible generative AI models (e.g., public versions of ChatGPT). Only secure, closed-loop enterprise AI systems that do not retain data for model training should be permitted.

- **Transparency and Disclosure:** If a party intends to rely on materials or legal submissions substantially generated by AI, they should be required to disclose this fact to the tribunal and opposing counsel. Likewise, the tribunal should disclose if AI tools were used for any significant administrative or drafting assistance.

Conclusion

The Quebec Superior Court's annulment of the award in *ARIHQ c. Santé Québec* serves as a defining precedent in the modern era of dispute resolution. It establishes a clear boundary: while technology can enhance the efficiency of the arbitral process, it cannot usurp the human judgment that parties contractually bargained for. As generative AI tools become ubiquitous, the international arbitration community must proactively regulate their use. By embedding robust AI guidelines into Terms of Reference, parties can protect the integrity of their proceedings, safeguard confidentiality, and insulate their awards from the devastating consequences of digital hallucinations.

Author: Mahmoud Abuwasel
Title: Partner – Disputes
Email: mabuwasel@waselandwasel.com
Profile: https://waselandwasel.com/about/mahmoud-abuwasel/

Lawyers and consultants.
Tier-1 services since 1799.
www.waselandwasel.com
business@waselandwasel.com