OPENING OF THE INTERNATIONAL ARBITRATION CONFERENCE 2025

"FABRICATION AND DELEGATION: AI IN INTERNATIONAL ARBIRATION"

13 October 2025

Sydney

The Hon A S Bell*

Chief Justice of New South Wales

1 Remarkably, it is less than 3 years since the initial release of ChatGPT which, with allied but competing platforms and products, have revolutionized the world in so many ways. The possibilities, applications, challenges and threats of Gen AI differ between countries, professions and industries.

2 The administration of justice and the rule of law rely in heavy measure on judges and legal practitioners including arbitrators taking *personal* responsibility for their work and being accountable for their decisions and actions in a variety of ways. This tradition does not sit entirely comfortably with the growing use of Gen AI in litigation and arbitration.

3 As many in this audience know, I have adopted a more cautious approach to the use of Gen AI in the courts of New South Wales than most other jurisdictions, both in Australia and internationally.

4 That having been said, I recognise that the advent of Generative AI (**Gen AI**) offers many and varied opportunities to improve the efficiency, and perhaps even the quality, of at least aspects of the judicial and the arbitral process. In this context, Gen AI is currently being used by parties to domestic and international arbitrations to assist with:[1]

* The Chief Justice acknowledges the considerable assistance of his tipstaff, Mr Sebastian Braham, in the preparation of this speech.

[1] See H Eidenmüller and F Varesis, "What is an Arbitration? Artificial Intelligence and the Vanishing Human Arbitrator" (17 June 2020) at 7-14, available at

(1) case management;

(2) legal research and drafting;

(3) document review and production;

(4) selecting suitable experts, counsel and arbitrators;

(5) transcription and translation services;

(6) outcome prediction and settlement proposals.[2]

5 Earlier this year, the Chartered Institute of Arbitrators issued its important Guideline on the Use of AI in Arbitration (2025) ("**CIArb Guideline**") with which many of you will be familiar.

6 The opportunities presented by Gen AI have been enhanced by the emergence of in-house large language models (**LLMs**) within law firms, which purport to provide enhanced privacy and confidentiality protections, and which are connected to the firm's internal data base, allowing the chatbots to be "well versed with the law firm's document corpus".[3] There have reportedly been a "roster of global law firms" competing to be prospective clients of particular technology companies with these products,[4] and one Silicon Valley AI start-up

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3629145; G Wagner and H Eidenmüller, "Digital Dispute Resolution" (22 June 2021), available at https://ssrn.com/abstract=3871612. See also C Morgan and S Chapman KC, "Inside Arbitration: Legally speaking – Are large language models friends or foe?" (Herbert Smith Freehills Kramer, 27 September 2023), available at <https://www.hsfkramer.com/insights/2023-09/inside-arbitration-legally-speaking-%E2%80%93-are-large-language-models-friends-or-foe>.

[2] As far back as 2016, researchers at UCL, the University of Sheffield, and the University of Pennsylvania developed a machine learning algorithm which predicted judicial decisions of the European Court of Human Rights with up to 79% accuracy: see UCL, "AI predicts Outcomes of Human Rights Trials" (24 October 2016), available at https://www.ucl.ac.uk/news/2016/oct/ai-predicts-outcomes-human-rights-trials).

[3] E Chan, K N Gore and E Jiang, "Harnessing Artificial Intelligence in International Arbitration Practice" (2023) 16 (2) *Contemporary Asia Arbitration Journal* 263 at 269.

[4] See S Merken, *Legal AI Race Draws More Investors as Law Firms Line Up*, Reuters (Apr. 27, 2023), available atwww.reuters.com/legal/legal-ai-race-draws-more-investors-law-firms-line-up-2023-04-26/.

is reported to have signed up a third of Australia's "top-tier firms" to an AI technology which "provides instant analysis and data processing for lawyers".[5]

7 The research capacity of Gen AI chatbots, such as ChatGPT, was recently displayed in a 2023 experiment coordinated by the Brazil Branch of the Chartered Institute of Arbitrators (CIArb), whereby ChatGPT-4 was put up against law student advocates in a moot based on the Willem C. Vis International Commercial Arbitration Moot.[6] ChatGPT was required to answer questions from the arbitral tribunal and make rebuttals and surrebuttals to the opposing party's arguments, with humans taking care of the oral presentation.

8 The law students won the moot, but "those observing considered ChatGPT-4's performance commendable, showcasing its potential to support preparation of a structured argument."[7] The arbitrators were, however, human. One wonders what the result might have been if that were otherwise, and the possibility of fully autonomous AI arbitrators, or AI-assisted arbitrators, is revisited later in this paper.

9 The use of Gen AI in international arbitration, although offering important benefits, also carries with it various risks including of misuse and overuse. I have spoken elsewhere and often about how the unverified use of Gen AI in the preparation of court proceedings has led to self-represented litigants, as well as lawyers, relying on fabricated case references, quotes and pieces of evidence, and the list of examples internationally grows by the day.[8] That observation and risk applies equally to the arbitral context.

---

[5] See https://www.smh.com.au/technology/from-hallucinations-to-high-court-can-ai-deliver-justice-20250909-p5mtpi.html.

[6] A lawyer and a data scientist provided ChatGPT with input over the course of the fortnight preceding the event to train and prepare it. See F R G Pereira et al, "Human vs. Machine?" CIArb News (May 10, 2023). Watch the moot at this link: https://www.youtube.com/watch?v=vnI4bHOkgFM&t=2704s.

[7] See J I Moreira and J Zhang, "ChatGPT as a fourth arbitrator? The ethics and risks of using large language models in arbitration" (2025) 41 *Arbitration International* 71 at 72. See also K Cheung and M A Quiñonero, "The Vis Moot's New AI Rules: Reflecting Current Sentiment & Foreshadowing Issues in Practice" (Kluwer Arbitration Blog, 22 December 2023), available at <https://arbitration-blog.kluwerarbitration.com/2023/12/12/the-vis-moots-new-ai-rules-reflecting-current-sentiment-foreshadowing-issues-in-practice/.

[8] See A S Bell, "Change at the Bar and The Great Challenge of Gen AI" (29 August 2025, Address to the Australian Bar Association). See also the recent publicity in relation to Deloitte relying on numerous

10 My focus in this address, however, is on two other potential categories of misuse, specific to the arbitration context and each with the potential to threaten the legitimacy and international currency of arbitral awards. They relate to the dual themes of fabrication and delegation.

Fabrication

11 One of the features of Gen AI is the ability to present results or to produce a document which appears articulate and authoritative both in terms of its text and look. Judges and arbitrators tend to respond well and positively to such work.

12 Many of you will be familiar with the case of *Contax Partners Inc BVI v Kuwait Finance House* [2024] EWHC 436 (Comm) (**Contax Partners**) where an attempt was made, only last year, to enforce a fabricated arbitration award in England to the tune of £70 million.

13 Substantial parts of the fabricated award had been lifted from a judgment of Justice Picken in *Manoukian v Société Générale de Banque au Liban SAL* [2022] EWHC 669 (QB). The application to register the award passed the first hurdle. Justice Butcher observed at [9] that:

> "This application was put before me, in the ordinary way, on a without notice basis, for consideration on the papers, in early August 2023. Judges of this court have to consider very many paper applications of this type and others. I recall considering this one with some care, in that I did not find it all very easy to understand. I gave, I would say in retrospect, undue allowance for difficulties apparently arising from documents being prepared by people who were not native English speakers and/or whose grasp of English procedure was not perfect. It did not, however, occur to me that any of the documents might be fabrications. I was not on the lookout for fraud, and did not suspect it."

hallucinated citations in a report it was commissioned to produce by the Department of Employment and Workplace Relations: https://www.theguardian.com/australia-news/2025/oct/06/deloitte-to-pay-money-back-to-albanese-government-after-using-ai-in-440000-report.

14 Fortunately, an urgent application to set aside the award was made before it could be enforced. The award was held to be not genuine and a fabrication.

15 A similar scheme was reportedly engaged in by a Kuwaiti figure by the name of Sheikh Ahman Al-Sabah (**Al-Sabah**), a former government minister, and Member of the International Olympic Committee (and so-called "king-maker" within the football world)[9]. The dispute arose of out videos, suspected to be deepfakes, which had been circulated online in 2013 by Al-Sabah and submitted to Kuwaiti authorities and which purported to show two former politicians discussing a plot for a coup.[10]

16 Seeking to confirm the veracity of the videos, Al-Sabah brandished an arbitral award purporting to have been rendered by a Swiss arbitrator. That arbitral award was reportedly presented to the High Court in London.[11]

17 In September 2021, the first instance criminal court of Geneva found Al-Sabah, his aid, and three lawyers guilty of charges of forgery of documents. The Court found the arbitration was a sham orchestrated to "create the impression that the video footage was genuine by relying on an alleged arbitral award establishing the veracity of the videos".[12]

18 The fabricated arbitral award was said to have been drafted by a British lawyer, and a Geneva-based lawyer had reportedly "[taken] on the role of arbitrator and signed a ruling stating that the videos were authentic and received a 10,000-

---

[9] https://www.abc.net.au/news/2017-05-02/aoc-olympic-merit-world-football-kingmaker-fifa-bribery-probe/8490214

[10] See https://www.acerislaw.com/fraudulent-arbitrations-a-few-bad-apples/. See also https://www.dw.com/en/kuwaits-sheikh-ahmad-found-guilty-in-case-involving-fake-videos/a-59146792; https://www.nytimes.com/2018/11/19/sports/indicted-kuwaiti-sheikh-steps-aside-from-ioc.html

[11] See https://www.swissinfo.ch/eng/business/geneva-court-convicts-kuwaiti-olympic-official-of-fraud/46938212; https://www.nine.com.au/sport/olympics/news-2024-international-olympic-committee-15-year-ban-olympic-power-broker-sheikh-ahmad-kuwait-20240504-p5jb4p.html.

[12] Ibid. See also M Burgstaller and S Macpherson, "Deepfakes in International Arbitration: How Should Tribunals Treat Video Evidence and Allegations of Technological Tampering?" (2021) 22 *Journal of World Investment & Trade* 860 at 880.

Swiss-franc payment in return."[13] In December 2023, the convictions were upheld on appeal by the appeal chamber of the Geneva Court of Justice.

19 Whether artificial intelligence was used to fabricate the awards sought to be enforced either in *Contax Partners* or by Al-Sabah is unclear. That Gen AI *is capable of* constructing such a fabrication, however, is unquestionable. The creation of a compelling fabrication would require both the language and the presentation of the document to resemble an authentic arbitral award. Neither criterion presents any significant obstacle. That may be explained technologically by reference to two prominent types of Gen AI.

20 You will likely be familiar with the basic architecture behind Gen AI; it works by "training models to understand and replicate patterns in data to create new content", typically in response to a prompt or request.[14] The development of a given model includes numerous relatively uniform steps, including: data collection, processing, model building, training, deployment, fine tuning & iteration, and evaluation.[15] But different models are trained in different ways.

21 Perhaps the most well-known deep learning architecture is the "transformer" technology behind Chat GPT and other GPT (literally Generative Pre-Trained Transformer) models, the key innovation of which is their self-attention mechanisms which allow them to process vast amounts of data from the internet in an unsupervised manner, and to weigh the importance of different parts of the data, thereby arming the chatbots with "a deep understanding of syntax, semantics, and context."[16] These transformer models are renowned for their ability to spit out "human-like text that exhibits fluency, coherence and creativity."[17] It is not difficult to imagine Chat GPT sweeping publicly available arbitral awards (as well as judgments of the court in common law jurisdictions),

---

[13] See https://www.france24.com/en/live-news/20210910-powerful-kuwaiti-royal-convicted-in-swiss-forgery-case.
[14] Z Huang, "Chapter 2: Unleashing Creative Potential" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 12.
[15] Ibid.
[16] Ibid at 15.
[17] Ibid at 14.

replicating their form, syntax and semantics, and spitting out a replica applied to the facts of a particular dispute.

22 Then there are Generative Artificial Networks (**GANs**), which were introduced in 2014 by American computer scientist and engineer, Ian Goodfellow, and which are known for their ability to generate highly realistic images (as well as text and audio). They have been used to create images of "human faces with astonishing accuracy", as well as in medical imaging, art, fashion and to assist in architectural modellings.[18] They have also been applied to generate "realistic video sequences, transforming images into different artistic styles and even in producing realistic human speech and music compositions".[19]

23 GANs are fundamentally engineered towards producing indistinguishable replicas of real data – the phenomenon of the deep-fake. Everyday in areas beyond the law, we learn of the creation of ever more realistic deepfakes. The scope for confusion and fraud is alarming.

24 GANS operate through what is described as an "adversarial" technological framework. That technology was explained by Jérôme De Cooman, Research Professor (Premier Assistant) at the University of Liege, as follows:[20]

> "It a nutshell, a GAN is composed of two competing sub-models that are simultaneously trained. A generative Model G generates new data (output) based on simple random variable (input). A distinctive model D then estimates the probability that output belongs to the training dataset or was generated by G. These models are called adversarial because they play a zero-sum game: either the discriminator correctly classifies the example as real or generated, or alternatively fails and the generator successfully fools the discriminator. Eventually, the discriminator is unable to classify the examples it has to label and reaches a 50 per cent success rate (i.e., with the same accuracy as predicting heads or tails when flipping a coin*). Indistinguishability has never been better defined.*" (emphasis added).

---

[18] Ibid.
[19] Ibid.
[20] J De Cooman, "Chapter 5: Unnatural Selection? A Darwinian Reading of the Economic Consequences of Generative AI on the Art Market" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 70.

25 This same risk has manifested in other fields, including creative ones, such as creative writing, poetry, and art. [21] For example, in April 2024, German photographer Boris Eldagsen won the award for the Creative Open Competition Category of the 2023 Sony World Photography Awards, with the following photo, entitled *Pseudomnesia: The Electrician*:[22]

26 The photograph was created by AI. Eldagsen refused to accept the award and wrote in an open letter on his website, [23] alongside the slogan "#promptography is not #photograhy":

> "Thank you for selecting my image and making this a historic moment, as it is the first AI generated image to win in a prestigous (sic) PHOTOGRAPHY competition. How many of you knew or suspected that it was AI generated? Something about this doesn't feel right, does it? AI images and photography should not compete with each other in an

[21] Ibid at 65-85. See also K Hseih, *Transformer Poetry: Poetry Classics Reimagined by Artificial Intelligence* (Paper Gains Publishing, 2019); H Aboutalebi et al, "DeepFakeArt Challenge: A Benchmark Dataset for Generative AI Art Forgery and Data Poisoning Detection" (https://doi.org/10.48550/arXiv.2306.01272).
[22] Ibid at 65, citing M Novak, "Artist Reveals His Award-Winning 'Photo' Was Created Using AI", *Forbes*, 17 April 2023 at https://www.forbes.com/sites/mattnovak/2023/04/17/artist-reveals-his-award-winning-photo-was-created-using-ai/.
[23] https://www.eldagsen.com/sony-world-photography-awards-2023/

award like this. They are different entities. AI is not photography. Therefore I will not accept the award."[24]

27 Similarly, although in this case not a forgery, there is the partnership between ING, Microsoft, TU Delft, Maurithuis, and Rembrandthuis, who in 2016 developed a painting described as "*in the style of* Rembrandt" by "using a deep learning facial recognition algorithm which first identified and then replicated Rembrandt's patterns."[25]

28 There are, regrettably, far more sinister examples. You may be aware of the case of a missing four-year-old boy, Gus Lamont, who was last seen on the 27th of September playing at his family's homestead in the South Australian outback. Late last week, a post circulated on Facebook of a boy who bore a striking resemblance to Gus (with long blonde curly hair) appearing to have been kidnapped by a man in a four-wheel drive vehicle, accompanied by the

[24] The World Photography Organisation claimed that the judges were always aware this was an image created using elements of A.I.

[25] J De Cooman, "Chapter 5: Unnatural Selection? A Darwinian Reading of the Economic Consequences of Generative AI on the Art Market" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 70. See also https://news.microsoft.com/europe/features/next-rembrandt/.

text: “Is this a kidnapping case?”.[26] The image was a fabrication generated by AI. South Australia’s Commissioner for Victims’ Rights described the post as an “extremely cruel act that has very real consequences for Gus’s family, the community police and those involved in the search for Gus.” [27]

29 That example brings me to a related but distinct risk: that, within a genuine arbitration, certain evidence might be fabricated. There have always been skilful forgers of documents, but I am particularly concerned with ability of Gen AI to facilitate the technological tampering of documentary, audio and video evidence, including deep fakes.[28] The risk posed by deep fake evidence is real and is exacerbated by the increased democratisation of the technology; it has been observed that “[t]he capacity to generate persuasive deep fakes will not stay in the hands of either technologically sophisticated or responsible actors. For better or worse, deepfake technology will diffuse and democratize rapidly.”[29]

30 Fabricated evidence or evidence containing hallucinations has already been relied upon in court proceedings,[30] including expert evidence, such as one quite remarkable case heard in the US District Court for the District of Minnesota concerning a law which purported to prohibit deep fakes “with the intent to injure a political candidate or influence the result of an election.”[31] An expert witness provided a report providing background about “the dangers of deepfakes to free speech and democracy”, but used GPT-4o to draft it, and ultimately relied on two non-existent academic articles.[32] The irony was not lost on the Court.

---

[26] See https://www.abc.net.au/news/2025-10-09/sa-ai-image-of-missing-child-appears-online/105868548.
[27] See https://www.abc.net.au/news/2025-10-09/sa-ai-image-of-missing-child-appears-online/105868548.
[28] Burgstaller and Macpherson (n 12). Article 3(60) of the EU AI Act defines a deepfake as an “AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.”
[29] R Chesney and D K Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security” (2019) 107 CLR 1759.
[30] See, eg, *Murray on behalf of the Wamba Wemba Native Title Claim Group v State of Victoria* [2025] FCA 731.
[31] *Kohls v Ellison* No 24-cv-3754. The relevant law was Minn. Stat. § 609.771.
[32] Ibid.

31 In 2020, a UK family lawyer reported that "deepfake evidence was used in a custody battle to try and portray a father as threatening" by using freely available systems on the internet.[33] Further, on 9 September of this year, the Supreme Court of California issued a terminating sanction after finding that the plaintiffs had submitted exhibits which were deep fakes, including video "testimonials" purporting to be the statements of a key witness, as well as technologically altered photographs.[34] The Court reasoned, in relation to the testimonials:[35]

> "But while exhibit 36 appears to capture a real-life interaction, certain characteristics of exhibits 6A and 6C, such as the lack of facial expressions, the looping video feed, among other things, suggested that these exhibits were products of GenAI – i.e., "deepfakes".

32 That decision was handed down less than a fortnight after I expressed the following concerns about deepfakes being relied upon in court proceedings:[36]

> "the scope for mischief and, still worse, fraud as a result of the increasing sophistication of technology able to generate "deepfake" evidence, both audio and video as well as documentary, is great and will present significant forensic challenges to courts and the Bar in the years ahead. The threat lies in the ease of creation of deepfake digital media through widespread and free or relatively inexpensive apps, on the one hand, and difficulties in detection, on the other." (footnotes omitted).

33 The same threat applies to arbitrations. Investor-State Dispute Settlements may be particularly at risk, since they often concern or involve high-profile political figures whose likeness is more vulnerable to manipulation by deepfake technology.[37] In the last eight years, for example, prominent deepfakes have been made of President Obama making uncharacteristic statements,[38] Tom

---

[33] K Jones and B Jones, "How Robust is the United Kingdom Justice System Against the Advance of Deepfake Audio and Video" available at chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://epluse.ceec.bg/wp-content/uploads/2022/12/20220912-04.pdf. See also https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/.

[34] *Ariel Mendones,* et al v. Cushman and Wakefield, Inc, et al (Case No. 23CV028772).

[35] Ibid at 2. One of the deep fakes can be accessed here: https://drive.google.com/file/d/1xI3xCC6Xdq94PZvz7QCl6M_XDsg8q55k/view.

[36] A S Bell (n 8) at [36].

[37] Burgstaller and Macpherson (n 12) at 861.

[38] C Silverman, "How to Spot a Deepfake Like the Barack Obama-Jordan Peele Video"

Cruise playing golf and performing magic tricks,[39] Ali Bongo (former president of Gabon) delivering a New Year's address, as well as different sex tapes allegedly featuring a Malaysian Minister of Economic Affairs, and the Governor of São Paulo, Brazil.[40] There is also in circulation a YouTube video which contains footage of two NSW Supreme Court judges speaking, but the words each of them is saying are not their words but words literally put into their mouths by the creator of the data file.[41]

34 The risk of deepfakes also carries with it the phenomenon of the so-called "liar's dividend", by which the veracity of genuine evidence might be doubted by claims that it has been digitally altered.

35 Fabricated evidence and even judgments and arbitral awards may be facilitated by Gen AI's particular penchant for verisimilitude and remarkable skills of imitation. These will no longer be the province of the skilled forger.

36 I regard this phenomenon as a particularly significant risk in litigation and arbitration.

Delegation

37 Moving from fabrication to delegation, another category fraught with risk is the delegation of the decision-making and the very arbitral function to Gen AI, by which I mean either fully autonomous AI arbitrators, or human arbitrators assisted by AI. That delegation may occur with or without the consent of the parties. I will deal with each separately.

(Buzzfeed, **17** April 2018), available at <www.buzzfeed.com/craigsilverman/obama-jordan-peeledeepfake-video-debunk-buzzfeed>.

[39] S Jain, 'No, That's Not Tom Cruise Playing Golf: Deepfake Videos of Actor Go Viral' *(NDTV,* 4 March 2021), available at <www.ndtv.com/offbeat/no-thats-not-tom-cruise-playing-golfdeepfake-videos-of-actor-go-viral-238356o>.

[40] Burgstaller and Macpherson (n 12) at 861.

[41] See A S Bell (n 8) at [32]-[33].

Non-Consensual Delegation

38 If an arbitrator delegates a significant portion, or all of their decision-making function to AI, without the express agreement of the parties, that may compromise the validity of the award in the seat of the arbitration as well as its enforceability in other jurisdictions. There is, however, a considerable evidentiary difficulty with proving that an arbitrator used AI, and, more difficult still, the degree to which they did so.

39 A recent illustrative example is found in *LaPaglia v Valve* 3:2025cv00833*,* (**LaPaglia**)*,*[42] in which the District Court for the Southern District of California is currently reserved on the question of whether to set aside an arbitral award (issued by an arbitrator of the American Arbitration Association) based on an allegation that the arbitrator used AI to draft to the award.

40 In late 2024, Mr La Paglia brought a complaint against Valve, a video game developer, publisher, and digital distribution company, alleging antitrust and breach of warranty for a defective game he had purchased. The claim was consolidated with 22 other individuals who made similar allegations against Valve.

41 On 7 January 2025, the Arbitrator dismissed the claim, but Mr La Paglia sought to have the award set aside, partly on the ground that the arbitrator relied on artificial intelligence to such an extent when drafting the award that he had "outsourced his adjudicative role" and thereby exceeded his powers under section 10(a)(4) of the Federal Arbitration Act[43] by acting outside of the scope of the parties' contractual agreement.[44]

---

[42] See https://www.acerislaw.com/when-arbitrators-use-ai-lapaglia-v-valve-and-the-boundaries-of-adjudication/.
[43] 9 U.S.C §§ 1 et seq.
[44] *LaPaglia v Valve* 3:2025cv00833 (Motion to Vacate, filed 8 April 2025, Document Number 1) available at chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://storage.courtlistener.com/recap/gov.uscourts.casd.810531/gov.uscourts.casd.810531.1.0.pdf.

42 To support the contention that the arbitrator had used artificial intelligence to draft the award, Mr La Paglia relied on the following factors:

- The arbitrator had, in the course of the hearing, admitted to the parties that he used ChatGPT to write articles in the past, including, during a break in the arbitration, telling "a story about how he had been assigned to write a short article on an aviation club he was part of, and that he had used ChatGPT to write it to save time".[45]

- The award itself allegedly "ha[d] telltale signs of AI generation", including the hallucination or mixing up of facts and "seemingly random, uncited reference[s]", as well as general statements about the gaming industry which were neither "in the record or otherwise evidenced or even argued".[46]

- The law clerk for Mr LaPaglia's counsel had asked ChatGPT whether it believed a certain paragraph was written by a human or AI, and ChatGPT concluded it was most likely that "the passage was generated by AI", citing awkward phrasing, redundancy, incoherence, and overgeneralisations;[47] and

- during breaks in the arbitration, the arbitrator had told the parties that he wanted to issue a decision quickly because he had a trip scheduled to the Galapagos Islands. The hearing took place over 10 days, the award was 29 pages long, and it was issued 15 days following the final post-hearing brief, on January 7, 2025, apparently the day when the arbitrator was scheduled to leave for the Galapagos.[48]

[45] *LaPaglia v Valve* 3:2025cv00833 (Declaration of William Bucher, filed 8 April 2025, Document Number 1-6 of 3) chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://storage.courtlistener.com/recap/gov.uscourts.casd.810531/gov.uscourts.casd.810531.1.6.pdf.
[46] See (n 44).
[47] Ibid.
[48] See (n 45).

Delegation by Consent

43 But what if there has been a form of consent to the arbitral tribunal's use of Gen AI either as part of the arbitration agreement or subsequent to the referral?

44 There has been much academic attention in recent years on the use of machine learning in judicial decision-making by the likes of an AI arbitrator,[49] and one theorist has suggested the delegation of human judgment to AI arbitrators will occur in three phases: [50]

(1) AI playing an "advisory" function "so that human arbitrators can double-check their conclusion"- a variant of this is human arbitrators checking AI's conclusion – the AI first draft model which I think is particularly insidious;

(2) a joint tribunal of AI and human arbitrators, with the final decision made by joint decree; and

(3) a sole AI arbitrator.

45 Various jurisdictions, in both judicial and arbitral contexts, have already reached phase 1. Overseas judges have used AI to assist with:

- discerning the ordinary meaning of words: in *Snell v United Speciality Ins. Co*.,[51] Newman J, in a concurring judgment, referred to ChatGPT for the common, everyday meaning of the word "landscaping" in what was a relatively ordinary civil insurance dispute.

- Clarifying areas of law: at the Law Society's Dispute Resolution Conference on 14 September 2023, Lord Justice Birss, a Judge of the UK Court of Appeal and the Deputy Head of Civil Justice, admitted to

[49] See, eg, I Ng (Huang Ying) and V Benedetti del Rio, "Chapter 8: When the Tribunal is an Algorithm: Complexities of Enforcing Orders Determined by a Software under the New York Convention" in in K F Gomez and M L Rodriguez, *60 Years of the New York Convention* (Wolters Kluwer, 2019).
[50] S Shih & E C Chang, "The Application of AI in Arbitration: How Far Away Are We from AI Arbitrators?" (2024) 17(1) *Contemporary Asia Arbitration Journal* 69 at 76.
[51] 102 F.4TH 1208 (11TH Cir. 2024).

using ChatGPT to summarise an area of law and inserting the summary into a judgment. His Honour described the technology as "jolly useful".[52]

- Clarifying regulations: a Columbian Judge sitting in Cartegena, Juan Manuel Padilla, admitted in February of 2023 to using ChatGPT "when deciding whether an autistic child's insurance should cover all of the costs of his medical treatment". [53] His Honour asked the chatbot:

> "Is an autistic minor exonerated from paying fees for their therapies"

and the response was:

> "Yes, this is correct. According to the regulations in Columbia, minors diagnosed with autism are exempt from paying fees for their therapies".

- Making findings of fact: a Dutch judge in 2024 used ChatGPT "to assist in estimating the lifespan of solar panels in a civil matter".[54] In the US, AI tools, such as the COMPAS system, have been utilised in bail hearings to assist in predicting the risk of absconding.[55]

- Drafting judgments: The Frankfurt District Court recently incorporated an AI tool from IBM to assist with the handling of approximately 10,000 to 15,000 cases a year regarding passenger rights arising from delays and cancellations in the airline industry.[56] The tool works as follows:

> "IBM's AI tool analyses written documents and then helps with the preparation of a draft judgment for the judge to review. Based on submissions in a case, the AI tool suggests facts and reasoning

---

[52] See https://www.theguardian.com/technology/2023/sep/15/court-of-appeal-judge-praises-jolly-useful-chatgpt-after-asking-it-for-legal-summary.
[53] The judge insisted that "by asking questions of the application, we do not stop being judges, thinking beings": see https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling.
[54] CLI:NL:RBGEL:2024:3636 (2024), cited in M Lou and E Lefley, "Chapter 25: Generative AI and Article 6 of the European Convention of Human Rights" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 461.
[55] See A Liptak, "Sent to Prison by a Software Program's Secret Algorithms" (New York Times, May 1, 2017) available at https://nyti.ms/2qoe8FC; *Loomis v State* (Wis 2016) 881 N.W.2d 749, 767.
[56] I S Szalai, "Stranger Disputes: When Artificial Intelligence Turns Arbitration Upside Down" Journal (2025) 25(2) Pepperdine Dispute Resolution Law 133 at 140.

for a draft opinion, with a human judge retaining ultimate control and responsibility for editing the AI's suggested judgment and issuing a final decision. In proposing suggestions to the judge, the AI tool also appears to rely on prior decisions from the court involving similar facts."[57]

46 The Singapore Judiciary is also exploring and experimenting with AI assistance in judgment writing, including in relation to a function known as "red-teaming", as Justice Aidan Xu, Judge of the High Court of Singapore, and Judge in Charge of Transformation and Innovation, recently explained:[58]

> [21] Nonetheless we think it is fruitful to explore AI assistance in judgment writing. A meaningful distinction can be drawn between the making of a decision and writing a judgment. As it is, we do have bench memoranda and parties' submissions to assist us in our work. Having AI generate drafts that we can look at, consider and adapt would seem to be worth having, as it would save time, and help with the process of composition. So we are not ruling it out, but we will be trying it out carefully, and impose appropriate training, safeguards and supervision.
>
> [22] In our experiments thus far, the results have been quite promising. When we give the appropriate prompts, with sufficient detail, and documents, gen AI has been able to create drafts that are fairly readable and fairly accurate, albeit with some hallucination or creative leaps. It is likely with the continued progress in the development of models, better prompts and scrutiny of products that we will be able to obtain fairly accurate and well-written, and seemingly well-reasoned drafts on a regular basis.
>
> [23] One particular area of use of this capability is for red-teaming, i.e. to generate drafts for different outcomes, to allow the judge to test his or her reasoning and conclusions, to identify weaknesses and strengthen his or her own work. Again, this will require some guidance and monitoring, but we think that this will be good to explore.

47 The Judicial Yuan of Taiwan is reported to have planned to roll out AI-supporting tools for judges to draft criminal judgments, but the rollout was

---

[57] Ibid.

[58] The Hon A Xu "Legal and Regulatory Issues with Artificial Intelligence: The Use (and Abuse) of AI in Court" at [21]-[23], available at https://www.judiciary.gov.sg/news-and-resources/news/news-details/justice-aidan-xu--speech-at-the-it-law-series-2025--legal-and-regulatory-issues-with-artificial-intelligence.

postponed in response to public concerns.[59] Estonia is also reported to have trialled an AI system to resolve small claims disputes under €7000.[60]

48 Moving from litigation to arbitration, the following developments may be observed.

49 The Guangzhou Arbitration Commission in China utilises AI in what are described as "smart courts" to automate realtime transcription and translation, blockchain recognition of evidence, and recommend decisions, as well as monitoring the consistency of judgments with past case law.[61] The AI is estimated to improve the efficiency of resolving disputes fourfold.

50 British Columbia's Civil Resolution Tribunal is an online alternative dispute resolution system, which provides "end-to-end" virtual solutions to small claims disputes. It utilises an AI system called "Solution Explorer" which "autonomously handles case intake, administration, and correspondence" and "gives disputants access to a negotiation platform".[62] The Tribunal has four phases, the first two of which are supported by AI:[63]

> "1. A purpose-built expert system, Solution Explorer, asks parties questions to understand the legal claim, classify and narrow the matters in dispute, and provide tailored legal information and appropriate forms.
>
> 2. An automated negotiation tool is used to support interparty communication and prepare draft agreements.
>
> 3. A facilitation phase is undertaken with an expert facilitator to help parties reach a consensual agreement.
>
> 4. If parties are still unable to reach agreement the matter proceeds to adjudication by a Tribunal Member."

---

[59] J S Reddy and V Singh, "Soft Law, Hard Justice: Regulating Artificial Intelligence in Arbitration" (November 2024) 17(2) *Contemporary Asia Arbitration Journal* 191 at 201.
[60] https://www.smh.com.au/technology/from-hallucinations-to-high-court-can-ai-deliver-justice-20250909-p5mtpi.html
[61] Reddy and Singh (n 59) at 200.
[62] Ibid at 200.
[63] Victorian Law Reform Commission, "Artificial Intelligence in Victoria's Courts and Tribunals: Consultation Paper" (17 October 2024) at 4.82, available at https://www.lawreform.vic.gov.au/publication/artificial-intelligence-in-victorias-courts-and-tribunals-consultation-paper/4-ai-in-courts-and-tribunals/#footnote-268.

51 Singapore has experimented with AI-assisted arbitration to unclog backlogs,[64] including by introducing a Gen AI assistant in their Small Claims Tribunals as a part of a pro-bono collaboration with Harvey AI which is "designed to help self-represented litigants understand their rights and tribunal procedures".[65] The Honourable Justice Aidan Xu described the initiative as follows:[66]

> "[27] …We want to explore and test what benefits AI can provide for self-represented persons, who would face a number of challenges in navigating the justice system, as well as for the judicial officers involved, who can often have to deal with a lot of unstructured, less organised information.
>
> [28] What we have been able to get in place is translation of documents, and we hope to complete summarisation of materials shortly. Such summarisation will assist the parties in understanding each other's case, see how their own case comes together and also it will assist the tribunal magistrate in making senses of what is before him or her. We do need to make sure that the system is as accurate as possible, and we want to make sure that it is properly integrated with the tribunal case management and filing system. So it has been a good learning experience for us, and even as we continue our exploration with Harvey AI here, we are thinking of what else can be done. We do hope to have a substantial update by the time of the TechLaw Fest in September. My personal dream is to have AI assist in presenting the case, and nudging towards settlement, but we will have to see whether we can get there, and indeed whether we should."

52 Another development in Singapore is Project Sea-Lion (Southeast Asian Languages in One Network), the region's first LLM which is said to cater to the region's diverse culture and languages:[67]

> 'As technology evolves rapidly, there is a strategic need to develop sovereign capabilities in LLMs. Singapore and the region's local and regional cultures, values and norms differ from those of Western Countries, where most large language models originate. A cornerstone of this initiative is the development of multimodal and localised LLMs for

---

[64] https://www.smh.com.au/technology/from-hallucinations-to-high-court-can-ai-deliver-justice-20250909-p5mtpi.html

[65] See https://insight.thomsonreuters.com.au/legal/posts/how-are-courts-adopting-ai-in-the-asia-pacific-region.

[66] The Hon A Xu (n 58).

[67] J G Allen and J Loo, "Chapter 10: Singapore's Evolving AI Governance Framework" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 165.

> Singapore and the region to understand context and values related to the diverse cultures and languages of Southeast Asia, for example, managing context-switching between languages in multimodal Singapore."

53 Australia's approach to the use of Gen AI, on the other hand, has been described as "more cautious" and "guideline-driven".[68] This may change, given the pace at which AI technology is developing, and, as States like Singapore and China continue to capitalise on Gen AI in their arbitral institutions, there will be increased pressure "to modernize…arbitration laws under the influence of regulatory competition". Indeed, regulatory competition is a defining characteristic of the market for international commercial arbitrations.[69]

54 As against that, it has been said that the international arbitration community has reportedly displayed a reticence to adopt new technological advancements "for fear that doing so may result in the setting aside or non-enforcement of an award".[70] That concern is not unfounded. The enforceability of an arbitral award is, of course, the single most significant consideration for parties choosing to pursue arbitration.[71]

55 What are the implications for the use of Gen AI in the enforceability of an arbitral award issued (with the consent of the parties) by a fully autonomous AI judge, or a human judge assisted by AI? On its face, given the prominence given to party autonomy in international arbitration, the answer should be high.

56 The NYC only mentions an "arbitrator" twice[72] and does not include any reference to the nature of the arbitrator or the tribunal, thus neither implicitly nor explicitly precluding AI arbitrators.[73] Also, whilst certain states prescribe a

---

[68] https://insight.thomsonreuters.com.au/legal/posts/how-are-courts-adopting-ai-in-the-asia-pacific-region.

[69] Eidenmüller and Varesis (n 1) at 43, citing C Rogers, "Is International Arbitration in a Race to the Top?" (Kluwer Arbitration Blog, March 18, 2018).

[70] G H Kasap, "Can Artificial Intelligence ("AI") Replace Human Arbitrators? Technological Concerns and Legal Implications" (2021) (2) *Journal of Dispute Resolution* 209 at 252.

[71] Eidenmüller and Varesis (n 1) at 28.

[72] Articles I.(2) and V(1)(b).

[73] Ng and Benedetti del Rio (n 49) at 123.

mandatory rule that that arbitrators must be natural persons, including France,[74] the Netherlands,[75] Taiwan,[76] and Scotland,[77] Australia does not.

57 The NYC prescribes that contracting States shall recognise and enforce agreements to arbitrate and arbitral awards subject only to limited grounds for refusing to do so. The onus will rest on the party resisting enforcement to establish one of those grounds. In *IMC Aviation Solutions v Altain Khuder LLC*,[78] Warren CJ said, *''*that in all but the most unusual cases, applications to enforce foreign arbitral awards should involve only a summary procedure*''*.

58 Leaving aside the potential issues which may arise in relation to the formality requirements imposed by Article IV,[79] Article V(1)(d) provides that the recognition and enforcement of an award may be opposed if a party can prove that "the composition of the arbitral authority or the arbitral procedure was not in accordance with the agreement of the parties, or, failing such agreement, was not in accordance with the law of the country where the arbitration took place." Where both parties have agreed to use an AI arbitrator, no issue in relation to Article V(1)(d) should arise.[80]

59 But what of Article V(2)(b) where "the recognition or enforcement of the award would be contrary to the public policy of that country."[81] There is no common definition of public policy, given the meaning of that term varies from time to time.[82] Since the public policy of the place of enforcement might differ from the jurisdiction supervising the arbitral proceedings, to avoid doubt, section 8(7A) of the *International Arbitration Act* provides that the enforcement of a foreign award would be contrary to public policy if the making of the award was induced

---

[74] Article 1450 of the French Civil Code (although this provision is applicable to domestic and not international arbitrations).
[75] Article 1023 of the Dutch Code of Civil Procedure provides "[a]ny natural person of legal capacity may be appointed arbitrator".
[76] Arbitration Law of R.O.C. Art 6 (2015).
[77] Bayraktaroğlu-Özçelik and Barış Özçelik (n 70) at 12.
78. (2011) 282 ALR 717; 38 VR 303 at [3].
[79] See the discussion in Ng and Benedetti del Rio (n 49) at 124-127.
[80] Ibid at 130.
[81] Public policy is also a ground to set aside an arbitral award under Article 34(2)(b)(ii) of the UNCITRAL Model Law.
[82] G Bayraktaroğlu-Özçelik and S Barış Özçelik, "Use of AI-Based Technologies in International Commercial Arbitration" (2021) 12(1) *European Journal of Law and Technology* at 11.

or affected by fraud or corruption or a breach of the rules of natural justice occurred in connection with the making of the award.[83] One element of the rules of natural justice is that the tribunal not exhibit bias.

60 The exemption is often invoked but rarely granted; it being generally accepted that it follows from the exceptional character of the exemption that it should be construed narrowly.[84] Indeed, as Sir Anthony Mason observed (sitting as a member of the Hong Kong Court of Final Appeal) in *Hebei Import and Export Corp v Polytek Engineering Co Ltd*):[85]

> "…the object of the Convention was to encourage the recognition and enforcement of commercial arbitration agreements in international contracts and to unify the standards by which agreements to arbitrate are observed and arbitral awards are enforced (*Scherk v Alberto-Culver Co* 417 US 506 (1974); *Imperial Ethiopian Government v Baruch-Foster Corp* 535 F 2d 334 (1976) at 335). In order to ensure the attainment of that object without excessive intervention on the part of courts of enforcement, the provisions of art V, notably art V(2)(b) relating to public policy, have been given a narrow construction. It has been generally accepted that the expression 'contrary to the public policy of that country' in art V(2)(b) means 'contrary to the fundamental conceptions of morality and justice' of the forum."

61 One view is that, on the hypothesis that the parties expressly agreed to the determination of their dispute by an AI arbitrator (or assistant arbitrator), the enforcement of the award would not breach public policy. Ng and Benedetti del Rio argue that:

> "…in the case of an AI arbitrator or an algorithmic arbitrator, it is difficult to extend the argument that this would be contrary to public policy, unless explicitly expressed so, as parties did contract to pursue arbitration, let alone contracting to pursue arbitration and agree to a decision made by an algorithmic arbitrator".[86]

---

83. M Davies, A S Bell, P L G Brereton and M Douglas, *Nygh's Conflict of Laws in Australia* (LexisNexis, 10th ed, 2020) at 1031.

84 See the distinctions between domestic public policy, international public policy, and transnational public policy in M Moses, "Chapter 11: Public Policy under the New York Convention: National, International and Transnational" in K F Gomez and M L Rodriguez, *60 Years of the New York Convention* (Wolters Kluwer, 2019) at 169, 173.

85 [1999] 2 HKC 205 at 232–3. Cf. *Resort Condominiums International Inc v Bolwell* (1993) 118 ALR 655 at 677–8.

86 Ng and Benedetti del Rio (n 49) at 131.

62 Nevertheless, numerous other scholars[87] have raised the following issues which may be raised in relation to Article V(2)(b).

63 First, partiality is an established category for setting aside awards on the basis that enforcement would be contrary to public policy.[88] The Paris Court of Appeal, for example, refused the enforcement of an award in 1998 where the arbitrator had been appointed by the same party in parallel arbitrations, one of which took place in France and the other in Italy. It was held that the arbitrator sitting on both tribunals had provided erroneous information to influence the Italian tribunal's decision on a question of jurisdiction, such that the impartiality of the arbitrator "created an imbalance between the parties, amounting to a violation of due process, so that the award rendered in Italy under such conditions violated Franch public policy".[89] The decision was later upheld by the *Court de Cassation*.[90]

64 Gen AI models are often accused of partiality on the basis that they are trained on datasets of human-generated content, themselves subject to present biases in society, which are then "amplif[ied] in [the AI's] outputs."[91] The following two examples are indicative of that dilemma: [92]

> "Reportedly, Amazon's resume screening tool – which used precursors to modern foundation models – learned to identify words that indicated if a candidate was a woman and then screened out those candidates. And recently legal scholars showed that when models were asked for advice on scenarios like car purchase negotiations and election outcome predictions, the advice 'systematically disadvantage[d] names that are commonly associated with racial minorities and women' with 'names

[87] See, eg, Eidenmüller and Varesis (n 1); Bayraktaroğlu-Özçelik Barış Özçelik (n 82); Kasap (n 70).
[88] See D Otto and O Elwan, "Article V(2)" in H Kronke, P Nacimiento, D Otto and N C Port (eds), *Recognition and Enforcement of Foreign Arbitral Awards: A Global Commentary on the New York Convention* at 369-372.
[89] A G Maurer, *The Public Policy Exception under the New York Convention: History, Interpretation and Application* (Juris, 2nd ed, 2022) at 109, citing Cour De Cassation [Supreme Court], March 24, 1998, *Excelsior Film TV srl v. UGC-PH*, reported in Yearbook of Commercial Arbitration Vol. XXIV (1999) at 643-4.
[90] Ibid
[91] Zou and Lefley (n 54) at 465.
[92] P Henderson, "Chapter 8: Challenges for Foundation Model Liability and Regulatory Regimes: An Analysis of US Law" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 124-5.

> associated with Black women receiv[ing] the least advantageous outcomes." (footnotes omitted)

65 A growing body of research has also identified that Gen AI outputs around non-Western cultural contexts entail "superimposed Western interpretations of those contexts".[93] Further, certain studies have indicated that AI arbitrators may be disposed to show bias "if previous awards reflect a pattern that is biased against consumers and in favour of companies."[94]

66 Some bias may even be intentionally worked into the algorithms; significant responsibility, for example, is given to data analysts with proprietary incentives, who engage in "feature selection": that is the process of deciding "which variables the model should observe and identify as most important when analysing correlations and patterns for a prediction task."[95]

67 The natural point in reply is that human judges also exhibit bias. But human bias is less pernicious than machine bias, because it is easier both to detect and to correct. Human bias is easier to detect because one has the benefit of observing the arbitrator's "human interactions", such as "how an arbitrator asks questions, or what issues the arbitrator focuses on, or the extent to which an arbitrator pays attention when one of the parties speaks."[96] Human bias is also easier to correct, since human arbitrators "can be trained to recognise and mitigate their bias", whilst AI arbitrators "may perpetuate and even exacerbate biases without the ability to self-reflect or adjust based on changing an nuanced ethical considerations."[97]

68 Secondly, and relatedly, the limited available *data* in respect of arbitral decisions poses a dilemma. The efficiency and accuracy of an AI arbitrator's decision depends heavily on the volume of training cases and data fed into it[98]

---

[93] Zou and Lefley (n 54)) at 465, citing D M Kotliar, "Data Orientalism: On the Algorithmic Construction of the Non-Western Other" (2020) 49 (5) *Theory and Society* 919.
[94] Kasap (n 70) at 225.
[95] Ibid at 227.
[96] S Shih & E Chin-Ru Chang, "The Application of AI in Arbitration: How Far Away Are We from AI Arbitrators?" (2024) 17(1) *Contemporary Asia Arbitration Journal* 69 at 77-8.
[97] Zou and Lefley (n 54) at 466.
[98] Kasap (n 70) at 221-2.

and "the risk of discrimination in question may decrease gradually with the increase in the amount of data to be processed by the AI".[99]

69 Although Australia, as a common law system (with publicly recorded judge-made law), is "better suited to generating sufficient data for machine learning applications than civil law systems",[100] international arbitral awards are rarely published and, even when they are, they are heavily redacted.[101] Notwithstanding efforts for increased transparency, "the vast majority of international commercial arbitral awards are still unpublished or published only sporadically".[102] It has also been observed that, leaving aside the question of publication, "arbitral decision-making is not sufficiently high-volume to make it an ideal candidate for automation with AI".[103]

70 Thirdly, the enforcement of an award may be considered contrary to public policy where there is a failure to give adequate reasons for the decision, since the lack of explanation for a decision substantially hampers a party's ability to identify grounds upon which to challenge it.[104]

71 AI-arbitrators may not be able to provide reasons for their decisions due to the so-called "black box problem", whereby their outputs are either opaque or too complicated for humans to comprehend and disentangle, including for its designers.[105] Where Gen AI chatbots do provide explanations, they are suspected of being generated post hoc, without reflecting the actual pathways by which the decision was made.[106] As Professor of Computer Science at the University of Oxford, Tom Melham has explained, this dilemma is a function of

---

[99] Bayraktaroğlu-Özçelik and Barış Özçelik (n 70) at 12.
[100] Eidenmüller and Varesis (n 1) at 17, 26
[101] Kasap (n 70) at 221-2.
[102] E Zlatanska, "To Publish, or Not to Publish Arbitral Awards: That is the Question" 81 *International Journal of Arbitration Media and Dispute Management* (2015) 25, 25.
[103] Kasap (n 70) at 222.
[104] According to Article 32(2) of the UNCITRAL Model Law, "[t]he award shall state the reasons upon which it is based, unless the parties have agreed that no reasons are to be given or the award is an award on agreed terms."[104] This requirement is also reflected in national legislation: Turkish International Arbitration Act, Art. 14/A(2); Belgian Judicial Code, Art. 1713(4). See also, eg, *Smart Systems Technologies Inc. v Domotique Secant* [2008] J.Q. No. 1782, 2008 QCCA 444 (Can. Que.), where the failure of an arbitral tribunal to provide reasons violated the public policy of Quebec,
[105] Kasap (n 70) at 229-39.
[106] Zou and Lefley (n 54) at 467.

the "deep neural network" machine learning models upon which Gen AI relies: [107]

> "Deep neural networks are highly complex models, and the 'knowledge' represented within them may be distributed across millions of numerical parameters, making it very difficult if not impossible to explain their behaviour – how and why they get the outputs they product – in human terms. The issue is the subject of much ongoing research, but for the time being, deep neural networks are commonly regarded as the 'epitome of black box techniques'."

72 Luis Greco expanded on this point, in his critique of "robo-judges": [108]

> "For insofar as algorithms function as black boxes, they deny individuals any justification for the outcome adversely affecting them. Put plainly, individuals are denied the very thing that constitutes the lawfulness of the results affecting them, the thing that distinguishes the decision from a fiat pronouncement."

73 One solution is for the chatbot to provide the *algorithmic* explanation for the decision by reference to its internal technology, but that runs into the obstacle that AI algorithms are the subject of trade secret protection and companies may legitimately refuse to disclose the relevant algorithm leading to the decision.[109]

74 Another emerging solution to the "black box" problem is the development of "explainable AI" research "which aims to create understandable AI models that can shed light on their decision-making processes".[110] Explainable AI techniques are, however, "still far from being optimal".[111] One study suggested that Chat GPT-3 explanations "degraded much more with example hardness

---

[107] T Melham, "Chapter 1: Generative AI: An Introduction" in M Zou, C Poncibò, M Ebers, R Calo (eds), *The Cambridge Handbook of Generative AI and the Law* (Cambridge University Press, 2025) at 7, citing ICO and the Alan Turing Institute, *Explaining Decisions Made with AI* (ICO, 2020).
[108] L Greco, "Judicial Power Without Judicial Responsibility: The Case Against Robot Judges" in D Moura Vicente, R Soares Pereira and A Alves Leal (eds), *Legal Aspects of Autonomous Systems: A Comparative Approach* (Springer, 2024)
[109] Kasap (n 87) at 230. See also T Wischmeyer, "Artificial Intelligence and Transparency: Opening the Black Box, Regulating Artificial Intelligence" 76, 89 in T Wischmeyer & T Rademacher (eds) *Regulating Artificial Intelligence* (Springer, 2020).
[110] Zou and Lefley (n 54) at 462. See also A Deeks, "The Judicial Demand for Explainable Artificial Intelligence" (2019) 119 *Columbia Law Review* 1829,
[111] J Schneider, "Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda", 15 April 2024, at 2, available at https://arxiv.org/html/2404.09554v1.

than human explanations".[112] In other words, the more difficult the question asked of the chatbot, the less likely it is to be able to explain its response. This does not portend well for AI's ability to explain its reasons in relation to complex, often multi-party, international arbitrations.

75 Fourthly, a court might observe that there is something inherent in having a human decision-maker, with their capacity for empathy and emotional intelligence, which is necessary to maintain public confidence in the administration of justice, thereby justifying the non-enforcement of the award on public policy grounds. As Gisem Halis Kasap has suggested:[113]

> "…Courts might find that an AI arbitrator's lack of emotional intelligence violates public policy even if there is no outright provision as to whether arbitrators need to be human. This is particularly true if the court finds that adjudication by a human is one of the most fundamental values of the country concerned."

76 There is some force in the suggestion that, although the parties to the arbitration have consented to the award, its enforcement would be cause for public concern. Mimi Zou and Ellen Lefley recently made a strong argument that the "trustworthiness" of a decision may be compromised if it is issued by AI.[114] One of the reasons for this is that "AI systems cannot be held accountable for their decisions in the way that human judges" (or arbitrators) can be.[115] Leaving aside the philosophical debates on whether autonomous AI has moral responsibility, it is also exceedingly difficult to attribute liability to a particular actor along the production pipeline of a Gen AI LLM, which consists of several actors in the areas of: data creation, data collection, model development, model deployment, and model use.[116]

---

[112] Ibid at 23, citing S Saha, P Hase, N Rajani et al, "Are Hard Examples also Harder to Explain? A Study with Human and Model-Generated Explanations" in Proceedings *of the Conference on Empirical Methods in Natural Language Processing,* available at https://aclanthology.org/2022.emnlp-main.137/.
[113] Kasap (n 70) at 252.
[114] Zou and Lefley (n 54) at 466-7.
[115] Ibid.
[116] P Henderson (n 92) at 123-4.

77 Inevitably the enforceability question will depend on the facts of the particular case, especially the degree to which the decision-making function was delegated to AI with the consent of the parties. Regrettably, that question is plagued by the same evidentiary obstacle in relation to which the District Court for the Southern District of California is currently reserved in *LaPaglia*: just how much of the judgment was written by AI?

78 Were detection technology to develop to a stage where it was possible to determine which parts of an award were written by a human and which parts were written by AI, the possibility of partial enforcement of an arbitral award may arise.[117]

79 The enforcement question may also depend on the nature of the dispute, and the extent to which it required a degree of human evaluation and discretion. It has been observed that:[118]

> "…it would be easier to see such systems operating for simple money claims or tax disputes where the outcome is based on the analysis of facts and the calculation of variables that are easily quantifiable. By contrast, cases involving 'hidden variables', such as social or economic considerations not evident in legal or factual documents, bring a degree of outcome-relevant uncertainty to the adjudicative processes, which, at the present stage of AI development, cannot properly be accounted for by available systems."

80 Finally, questions of enforceability in the arbitral context may depend on the development of Gen AI technology, and the extent to which utilising machine learning in judicial decision becomes accepted or is regarded as contrary to public policy, a concept which evolves over time and cannot be "stereotyped", even within a given common law country.[119] Thus, it has been observed that:[120]

> "since public policy reflects the mores and fundamental assumptions of the community, the content of the rules should vary from country to country and from era to era".

---

[117] *ACN 006 397 413 Pty Ltd v International Movie Group (Canada) Inc* [1997] 2 VR 31.
[118] Eidenmüller and Varesis (n 1) at 17.
[119] *Nordenfelt v Maxim Nordenfelt Guns and Ammunition Co Ltd* [1894] AC 535 at 553-4.
[120] *CBI NZ Ltd v Badger Chiyoda* [1989] 2 NZLR 669 at 674.

81 For now, suffice to say that with the advent of Gen AI, the international arbitration community should be alive to the dual issues of fabrication and delegation which I have explored in this paper.