WEAVING ARBITRAL AWARDS FROM THE MILLS OF ARTIFICIAL INTELLIGENCE

Ahan Gadkari and Ankit Malhotra

Ahan Gadkari serves as a Research Assistant under Dr. Aniruddha Rajput, Member, UN International Law Commission. Ankit Malhotra is the President of the Jindal Society of International Law. They can be reached at 18jgls-ahan.mg@jgu.edu.in and 19jgls-ankit.m@jgu.edu.in respectively.

Abstract

While it has been predicted that artificial intelligence will rule the roost by monopolizing almost all aspects of life. It pinches the shoe necessarily when this argument is made to lawyers and their practice. Some conclude the answer is on the wall. The wall, however, is hidden with a shabby dark cloth covering the writing since humans have always been wary of industrial revolutions. This paper will show a stronger reliance is being placed on artificial intelligence in law making it a burgeoning industry. The objective of this paper is not to focus on procedure, however, it is, to focus on the decision-making process behind the awards rendered. To do so, a technical study becomes the need of the hour. This study will underline the implications and limits of artificial intelligence and therefore expose the human reliance on technology. As a result, the tete-a-tete of humans and robots vis-à-vis legal decision-making will become clear.

Keywords: Arbitration, Artificial Intelligence, Legal-Decision Making, Harmonization

1. Introduction:

Artificial Intelligence (hereinafter 'AI') has been predicted to be used in a wide variety of tasks in international arbitration, including the appointment of arbitrators, legal research, drafting and proofreading of written submissions, document translation, case management and document organization, cost estimations, hearing arrangements (such as transcripts or

simultaneous foreign language interpretation), and the drafting of standard sections of awards (such as procedural history).[1]

Nonetheless, most attorneys feel the effect on their profession will be minimal. This misses the fact that AI is being used in a variety of fields of law, including contract analysis, legal research, and electronic discovery.[2] For example, computer applications are available to assist attorneys in analyzing the opposing party's written filings and providing pertinent case law that was omitted or delivered thereafter. Unsurprisingly, artificial intelligence in law is a burgeoning industry.[3]

This article will not address those points but will instead concentrate on one of the most contentious components of the arbitral procedure: *the decision-*

[1] *See* Kate Apostolova & Mike Kung, *Don't Fear AI in IA*, Global Arb. Rev. (27 Apr. 2018); Adesina Temitayo Bello, *Online Dispute Resolution Algorithm: The Artificial Intelligence Model as a Pinnacle*, 84(2) Int'l J. Arb. Mediation & Dispute Mgmt. 159 (2018); Emma Martin, *The Use of Technology in International Arbitration*, in *40 Under 40 International Arbitration* 337–48 (Carlos Gonzalez-Bueno ed., Wolters Kluwer 2018); Paul Cohen & Sophie Nappert, *The March of the Robots*, Global Arb. Rev. (15 Feb. 2017); Sophie Nappert, *Disruption Is the NewBlack – Practical Thoughts on Keeping International Arbitration on Trend*, (2) ICC Dispute Resolution Bulletin 20, 25–36 (2018); Sophie Nappert, *The Challenge of Artificial Intelligence in Arbitral Decision- Making*, Practical Law UK Articles (4 Oct. 2018); Kathleen Paisley & Edna Sussman, *Artificial Intelligence Challenges and Opportunities for International Arbitration*, 11(1) NYSBA New York Dispute Resolution Lawyer 35 (Spring 2018); Christine Sim, *Will Artificial Intelligence Take over Arbitration?*, 14(1) Asian Int'l Arb. J. 1 (2018); Robert H. Smit, *The Future of Science and Technology in International Arbitration: The Next Thirty Years*, in *The Evolution and Future of International Arbitration* 365–78 (Wolters Kluwer 2016); Francisco Uríbarri Soares, *New Technologies and Arbitration*, VII(1) Indian J. Arb. L. 84 (2018); Gauthier Vannieuwenhuyse, *Arbitration and New Technologies: Mutual Benefits*, 35 J. Int'l Arb. 119–29 (2018); Mohamad S. Abdel Wahab, *Online Arbitration: Traditional Conceptions and Innovative Trends*, in *International Arbitration: The Coming of a New Age?* ICCA Congress Series 17, 654–67 (Albert Jan van den Berg ed., Wolters Kluwer 2013).

[2] *See* e.g. Richard Susskind, *Tomorrow's Lawyers: An Introduction to Your Future* (2d ed., Oxford University Press 2017); Philip Hanke, *Computers with Law Degrees? The Role of Artificial Intelligence in Transnational Dispute Resolution, and Its Implications of the Legal Profession*, 14(2) Transnat'l Disp. Mgmt. 1 (2017).

[3] Robert J. Ambrogi et al., *Ethics Issues in Lawyers' Use of Artificial Intelligence*, presentation at 44th ABA National Conference on Professional Responsibility (1 June 2018), www.americanbar.org/content/dam/aba/events/professional_responsibility/2018_cpr_meetings/20 18conf/ma... (accessed 9 March 2022).

making process itself.[4] It will examine if and how artificial intelligence may be used to assist or even replace arbitrators in their role of resolving disputes. Notably, this article is not about online arbitration, which refers to procedures in which processes are simplified via the use of technology, such as electronic filings, but where human arbitrators continue to make decisions.[5] Additionally, although this article focuses on arbitral decision-making, it draws on examples and research from a broad range of legal disciplines, and its findings apply to judicial decision-making more broadly, not only in international arbitration.

When contemplating the viability of using AI for arbitral decision-making, some have speculated on the plausibility of 'robot-arbitrators,'[6] but little study has been conducted on the possible consequences of this usage. Typically, authors either argue that AI is inevitable in the future[7] or express skepticism, based on the notion that some 'human aspect' is required to assure empathy and emotional justice.[8] This article will go further into the subject, examining the technical elements of artificial intelligence, their implications and limits, as well as the more basic influence they may have on human decision-making processes and theories thereof.

This paper is divided into seven sections. Section 2 defines AI and discusses its most salient characteristics. A firm grasp of the technical features of AI is required in order to examine its implications for legal decision-making adequately. Section 3 reviews current research on the use of artificial intelligence to forecast the result of judicial judgements. It critically appraises their methodology and findings, raising doubts about the degree to which those research demonstrate the broad application of AI for ex ante

[4] Susan D. Franck et al., *Inside the Arbitrator's Mind*, 66 Emory L.J. 1115 (2017); Maxi Scherer, *International Arbitration 3.0 – How Artificial Intelligence Will Change Dispute Resolution*, Austrian Y.B. Int'l Arb. 503 (2019).

[5] *See* e.g. Amy J. Schmitz, *Building on Arb Attributes in Pursuit of Justice*, in *Arbitration in the Digital Age* 182 (Maud Piers & Christian Aschauer eds, Cambridge University Press 2018); Pablo Cortés & Tony Cole, *Legislating for an Effective and Legitimate System of Online Consumer Arbitration*, in *Arbitration in the Digital Age, supra* n. 5, at 209.

[6] Paul Cohen & Sophie Nappert, *Case Study: The Practitioner's Perspective*, in *Arbitration in the Digital Age, supra* n. 5, at 126, 140–45. Cohen & Nappert, *supra* n. 1; José María de la Jara, Daniela Palma & Alejandra Infantes, *Machine Arbitrator: Are We Ready?*, Kluwer Arbitration Blog (4 May 2017).

[7] Apostolova & Kung, *supra* n. 1.

[8] Soares, *supra* n. 1, at 101; de la Jara, Palma & Infantes, *supra* n. 6. *See also* more nuanced Sophie Nappert, *The Challenge of Artificial Intelligence in Arbitral Decision-Making*, Practical Law UK Articles (4 Oct. 2018).

outcome prediction. Section 4 explores the fundamental constraints of the AI models used, which are based on the so-called four Vs of Big Data – Volume, Variety, Velocity, and Veracity – and their implications for legal decision-making. This section examines the need of having a sufficient amount of non-confidential case data, the demand for recurring fact patterns and binary outcomes, the issue of policy changes over time, and the concerns of bias and data diet vulnerability. Section 5 discusses a significant disadvantage of AI decision-making: the difficulty of giving reasoned legal judgements derived by AI. Section 6 discusses the implications of AI choices on legal theories of judicial decision-making. It demonstrates that AI would alter the normative foundations for decision-making, implying a fundamental paradigm shift from a theoretical standpoint. Section 7 summarizes the article's conclusions and major results.

2. Characteristics of Artificial Intelligence:

Lawyers often lack a working knowledge of artificial intelligence.[9] It is stated that AI-savvy attorneys are as scarce as vegan butchers.[10] Even if attorneys do not want to become computer scientists, it is critical for them to comprehend the fundamental characteristics of artificial intelligence. Only with a thorough grasp of AI can the potential ramifications for the legal profession and legal thought be assessed. The purpose of this part is to present some critical technical background knowledge on artificial intelligence.

Artificial intelligence is described as 'making a machine behave in ways that would be called intelligent if a human were so behaving'.[11] Indeed, this was the concept offered by John McCarthy, a late computer scientist who is credited with coining the term 'AI' in 1956. There are several meanings that are comparable. The Oxford Dictionary, for example, describes artificial intelligence as the '[t]heory and development of computer systems able to

[9] Queen Mary School of International Arbitration Survey, *The Evolution of International Arbitration* 33 (2018) ('As far as AI is concerned, the lack of familiarity translates into a fear of allowing technology to interfere excessively with the adjudication function, which is supposed to be "inherently human"').

[10] Marc Lauritsen, *Towards a Phenomenology of Machine-Assisted Legal Work*, 1(2) J. Robotics, Artificial Intelligence & L. 67, 79 (2018).

[11] John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (31 Aug. 1955), in *Artificial Intelligence: What Everyone Needs to Know*1 (Jerry Kaplan ed., Oxford University Press 2016), www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (accessed 15 April 2022).

perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages'.[12]

Human intelligence is used as a benchmark for AI in these criteria. The word 'intelligence' is not easily defined and has sparked disagreements among philosophers, psychologists, cognitive scientists, and other specialists.[13] At a fundamental level, intelligence may be defined as 'the ability to learn, understand, and make judgments or have opinions that are based on reason'.[14] It is this capacity that differentiates humans from other types of non-intelligent or less intelligent life.[15]

During the early days of AI research, computer scientists attempted to create algorithms that mimicked human intelligence by attempting to comprehend and recreate human cognitive processes.[16] For example, computer scientists sought to comprehend the mechanisms involved in language acquisition in order to create an algorithm - a series of exact instructions – that would allow computers to acquire a language. The results were disappointing, especially when it came to more complicated activities, such as language acquisition.[17]

---

[12] Oxford Living Dictionaries, https://en.oxforddictionaries.com/definition/artificial_intelligence (accessed 15 April 2022).

[13] *See* e.g. Shane Legg & Marcus Hutter, *A Collection of Definitions of Intelligence*, 157 Frontiers in Artificial Intelligence & Applications 17 (2007). In the context of AI, the distinction between fluid intelligence (i.e. the ability to reason and think flexibly) and crystallized intelligence (i.e. the accumulation of knowledge, facts, and skills that are acquired throughout life) seems important. *See* e.g. David F. Lohman, *Human Intelligence: An Introduction to Advances in Theory and Research*, 59(4) Rev. Educational Res. 333 (1989).

[14] Cambridge Dictionary, https://dictionary.cambridge.org/dictionary/english/intelligence (accessed 9 March 2022).

[15] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, 24 et seq. (Knopf 2017).

[16] Steven Levy, *The AI Revolution Is on*, WIRED (27 Dec. 2010), www.wired.com/2010/12/ff-ai- essay-airevolution (accessed 9 March 2022); Osonde Osoba & William Welser IV, *An Intelligence in Our Image – The Risk of Bias and Errors in Artificial Intelligence* 5 (Rand 2017); Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach* 693 (3d ed., Pearson 2010).

[17] Mathias Winther Madsen, *The Limits of Machine Translation* 5–15 (2009) Master Thesis University of Copenhagen, http://vantage-siam.com/upload/casestudies/file/file-139694565.pdf, cited in Harry Surden, *Machine Learning and the Law*, 89 Wash. L. Rev. 87, 99 (2014).

Similar models are being utilized to a lesser degree nowadays. These programs are referred to as expert systems or rule-based programs.[18] These systems are built on a collection of rules, often expressed as 'if-then' statements (e.g., if the light turns red, then stop), referred to as the knowledge base. They make use of logical conclusions based on the knowledge base's principles. There are a number of reasons why such programs are not as strong as the other models shown below. Most significantly, they are time consuming since the knowledge base must be manually constructed by defining the rules and programming the application appropriately.[19] Furthermore, ex ante rules, such as 'if/then' concepts, are often insufficient to correctly explain complicated and dynamic reality.[20]

As a result, many models were constructed. The quantum leap in AI research coincided with the development of massive volumes of data, dubbed the 'dataquake.' [21] This data explosion was attributed to a combination of improved computer processing speed (which, according to so-called Moore's Law, doubles every twelve to eighteen months)[22] and lower data storage costs (which, according to so-called Kryder's Law, also double at a comparable rate).[23] The advent of 'Big Data' facilitated a sea change in the development of artificial intelligence. Rather than constructing sophisticated algorithms for cognitive operations, artificial intelligence is being utilized to 'learn' from previously collected data.

Machine learning is an area of artificial intelligence study that is concerned with computer systems that learn from their experiences and continuously improve their performance.[24] The term 'learning' does not relate to the cognitive processes associated with human learning; rather, it refers to the functional meaning of learning: the capacity to modify behavior over time as a result of experience.[25] Machine learning has produced startling

[18] Ethem Alpaydin, *Machine Learning* 50–52 (MIT Press 2016); Margaret A. Boden, *Artificial Intelligence: A Very Short Introduction* 26–28 (Oxford University Press 2018).
[19] Alpaydin, *supra* n. 18, at 50–52.
[20] Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, 55 Communications of the ACM 78, 80 (2012).
[21] Alpaydin, *supra* n. 18, at 10–13.
[22] Gordon E. Moore, *Cramming More Components onto Integrated Circuits*, Electronics 114 (19 Apr. 1965), reprinted in 86 Proceedings of the Institute of Electrical and Electronics Engineers 82 (1998).
[23] Chip Walter, *Kryder's Law*, 293 Scientific American 20 (1 Aug. 2005).
[24] Russell & Norvig, *supra* n. 16, at 693.
[25] Surden, *supra* n. 17, at 89.

outcomes in a variety of fields.[26] Continuing with the previous example of language acquisition, computer translation algorithms are becoming more accurate. In contrast to the earlier attempts described above, no programmer is required to code a translation algorithm; rather, computer models, such as neural networks, use massive amounts of available data to 'learn' the relevant features and continuously improve as a result of immediate online feedback via user clicks. Boden observes that many networks have an astonishing ability to self-organize from a random start.[27]

At its heart, machine learning is based on inferring hidden elements or patterns from seen data.[28] Rather of coding the essential algorithms into the machine, the computer pulls them from vast volumes of sample data and adequate computational power. In many cases, it is impossible to define the algorithm in terms of accurate ex ante instructions.[29] For example, humans are capable of quickly determining whether email is spam, but they are unable of providing accurate and complete instructions for this categorization job. However, if the software is given a large sample of emails classified as 'spam' or 'not spam,' the program will be able to discover the required classification method. It does this by identifying repeated patterns in spam emails and inferring that future emails with the same characteristics should be categorized as spam as well.

The phrase 'data mining' refers to the process of discovering hidden patterns. The comparison is that one must sift through tons of soil from a mine in order to discover valuable stuff.[30] In the context of artificial intelligence, the computer sifts through massive volumes of data in search of an appropriate model. Once the hidden model is discovered, it may be used to forecast future instances (e.g., categorize a future email as spam or not), which is very useful in the legal environment, as explained further below.[31]

[26] For a recent example, *see* a live debate between a human and an AI-driven digital debater, www.research.ibm.com/artificial-intelligence/project-debater/live (accessed 12 March 2022).
[27] Boden, *supra* n. 18, at 70.
[28] Alpaydin, *supra* n. 18, at xi.
[29] Surden, *supra* n. 17, at 94.
[30] Alpaydin, *supra* n. 18, at 14.
[31] *Ibid.*

The capacity to recognize patterns is based on statistical and probability calculations.[32] In basic words, the computer software determines the likelihood of a certain event for each item or combination of variables it sees. For instance, if an email contains the words 'sex' and 'Viagra,' the likelihood that it is spam is high. Probabilistic theories, such as Bayesian networks, underpin machine learning AI's success.[33] The learning programs are structured similarly to a generic template with adjustable parameters, with the goal of adapting the model's parameters based on the knowledge retrieved from the sample data. As Alpaydin puts it, '[i]ntelligence seems not to originate from some outlandish formula, but rather from the patient, almost brute force use of simple, straightforward algorithms'.[34]

As a result, AI models can generate 'intelligent' outputs that, when done by people, are believed to entail complex cognitive processes (e.g. understanding emails in order to classify them as spam).[35] However, this conclusion is obtained without the use of 'intelligent' human-cognitive processes, but rather via the use of probabilistic models. As one author puts it, 'research has shown that certain … tasks can be automated – to some degree – through the use of non-cognitive computational techniques that employ heuristics or proxies (e.g. statistical correlations) to produce useful, "intelligent" results'.[36] The consequences for legal decision-making that follow from this transition away from early models that focused on human-like processes and toward statistical or probabilistic models that provide human-like results without relying on 'intelligent' processes are examined in further detail below.

Researchers in artificial intelligence identify many forms of machine learning based on the degree of human involvement. Supervised learning necessitates human interaction: the programmer trains the algorithm by specifying a set of desired outcomes (e.g. spam/no-spam categorization) given a range of input.[37] This implies that the training set's data must be properly labeled (e.g., emails must be classified as spam or not) and that some sort of human feedback is necessary (e.g. when the program wrongly

[32] Boden, *supra* n. 18, at 39–40.
[33] Alpaydin, *supra* n. 18, at 63–64, 82–84.
[34] *Ibid.*, at xii.
[35] Gary Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins* (John Murray 2017).
[36] Surden, *supra* n. 17, at 95.
[37] Peter Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data* 2 (Cambridge University Press 2012).

classifies an email). On the contrary, unsupervised learning needs little, if any, human intervention. There are no pre-defined assumptions or outputs; rather, the software discovers co-occurring characteristics that imply that they will co-occur in the future.[38] This is true, for example, of several recent language translation algorithms outlined above.

Notably, there is no one AI system, but rather a collection of distinct models.[39] The distinctions between the two methodologies discussed above are critical for the present investigation. On the one hand, expert models are rule-based and adhere to logic as their guiding principle of reasoning. Additionally, they may be regarded as taking a forward approach, since they apply pre-defined rules to observable data. The technique is causal, deducing the conclusion from the algorithm's pre-defined, predefined rules. On the other hand, machine learning models, such as neural networks, often lack predefined rules and instead rely on pattern recognition and are constructed using probabilistic approaches as a guiding principle. They are sometimes referred to as inverse algorithms, since they derive the algorithm from observable data. The approach is predictive in nature, assessing the probability of any given result based on the extracted, and continuously developing, algorithm.

3. Legal Decision-Making and Artificial Intelligence: The Application Of Quantitative Prediction

To the majority of attorneys, the notion that AI-driven systems might forecast the result of legal decision-making is counter-intuitive. Lawyers intuitively assume that legal decision-making includes cognitive processes – such as comprehending the parties' legal statements and rationally identifying the correct conclusion – that cannot be accomplished by computer algorithms. However, as noted above, computer models are capable of producing 'intelligent' outputs that are regarded to require high-level cognitive processes if conducted by people.

[38] Boden, *supra* n. 18, at 40.
[39] *Ibid.*

Numerous research may bolster the argument that computer systems are superior than humans in predicting the result of judicial decisions.[40] For example, an early research shown that computer systems outperformed human experts in forecasting individual US Supreme Court justices' votes in impending 2002 term judgments. The computer model properly predicted 75% of votes, whereas the human expert panel of prominent attorneys and law professors correctly predicted just 59.1% of votes.[41]

The fundamental rationale for this seeming victory of AI is because human brains have 'hardware' constraints that computer programs readily overcome.[42] In the following years, consumer-level computers are likely to achieve storage capacities of several petabytes.[43] Fifty petabytes is sufficient to hold all of humanity's written works from the dawn of recorded history in all languages. As a result, computers can easily store large quantities of data and retrieve information – or experience – considerably more rapidly and effectively than humans can.[44]

This section examines the methodology and findings of two recent studies on the prediction of legal decision-making. Section 3.1 examines a 2016 research relating to European Court of Human Rights rulings, while Section 3.2 examines a 2017 analysis predicting US Supreme Court outcomes.

3.1 **Predicting European Court of Human Rights judgments**

[40] *see* Roger Guimerà & Marta Sales-Pardo, *Justice Blocks and Predictability of U.S. Supreme Court Votes*, 6(11) PloS One (2011); Andrew D. Martin et al., *Competing Approaches to Predicting Supreme Court Decision Making*, 2(4) Persp. Pol. 761 (2004); Theodore W. Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Sciences Approaches to Predicting Supreme Court Decision making*, 104 Colum. L. Rev. 1150 (2004). Generally on forecasting, *see* Philip E. Tetlock, *Expert Political Judgment: HowGood Is It? How Can We Know?* (Princeton University Press 2005); Philip E. Tetlock & Dan Gardner, *Superforecasting: The Art and Science of Prediction* (Crown 2015).

[41] Ruger et al., *supra* n. 40, at 1152.

[42] Tegmark, *supra* n. 15, at 27–28.

[43] *How Much Is a Petabyte*?, Mozy BLOG (2009), cited in Daniel M. Katz, *Quantitative Legal Prediction*, 62 Emory L.J. 909, 917 (2013).

[44] Interestingly, France has recently prohibited, under threat of criminal sanctions, the use of certain data from published decisions for predictive analytics. A newly introduced provision states that '[t]he identity data of magistrates and members of the judiciary cannot be used with the purpose or effect of evaluating, analysing, comparing or predicting their actual or alleged professional practices'. *See* Law No. 2019-222 (23 Mar. 2019), Art. 11, www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000038262498&dateTexte=20190604 (accessed 23 March 2022).

The study conducted by a group of researchers in 2016[45] examined decisions by the European Court of Human Rights (hereinafter the 'ECtHR') in the English language regarding three provisions of the European Convention on Human Rights (hereinafter the 'Convention'), namely Article 3 on the prohibition of torture, Article 6 on the right to a fair trial, and Article 8 on the right to respect for private and family life.[46] These clauses were selected because they resulted in the greatest number of Convention judgments and so offered adequate data for the research.[47] The analysis chose an equal number of judgments in which the ECtHR found a breach as in which it did not. This resulted in a total of 584 decisions: 250 in the Article 3 category, 80 in the Article 6 category, and 254 in the Article 8 category.[48]

The study's technique concentrated on the textual information included in choices via the application of natural language processing and machine learning.[49] The analysis used text extracted from the judgements, which followed the standard framework of ECtHR rulings, which includes parts on process, factual background, and legal reasoning.[50] The operative parts of rulings, in which the Court declares the case's conclusion, were omitted from the input.[51] The output objective was a binary classification job determining whether or not the ECtHR found a breach of the Convention's underlying Article.[52] A 10% subset of the dataset was used to train and test the model.[53]

As a consequence, the model achieved an overall accuracy of 79 percent in predicting the outcome of the Court's ruling.[54] The portions outlining the

[45] Nikolaos Aletras et al., *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, PeerJ Computer Science 2:e93 (2016).
[46] The 'European Convention on Human Rights' refers to the Convention for the Protection of Human Rights and Fundamental Freedoms, signed in Rome on 4 November 1950, as amended and supplemented by subsequent Protocols Nos. 1, 4, 6, 7, 12, 13, 14, and 16, www.echr.coe.int/Documents/Convention_ENG.pdf (accessed 23 March 2022).
[47] Aletras et al., *supra* n. 45, at 6.
[48] *Ibid.* at 8.
[49] *Ibid.* at 1.
[50] *Ibid.* at 4-6.
[51] *Ibid.* at 8.
[52] *Ibid.* at 2.
[53] *Ibid.* at 9.
[54] *Ibid.* at 10.

factual circumstances and procedural background had the highest predictive value (76 percent and 73%, respectively), but the section outlining legal reasoning had a lower predictive value (62 percent).[55] Additionally, the research included the most commonly used terms for a variety of themes, along with their respective predictive weight for a violation or non-violation. For example, the most commonly used terms with a high predictive value are: 'injury', 'damage', 'Ukraine', 'course', 'region', 'effective', 'jail', 'well', 'ill treatment', 'force', and 'beaten'.[56] 'Appeal', 'execution', 'limit', 'copy', 'employee', 'January', and 'fine' are all defined in Article 6 of the Convention;[57] and 'son', 'body', 'result', 'Russian', 'department', 'attack', and 'death' as defined in Article 8 of the Convention.[58]

The study's authors assert that their findings may pave the way for ex ante prediction of the result of future ECtHR cases. They state as follows:

"[o]ur work lends some initial plausibility to a text-based approach with regard to *ex ante* prediction of ECtHR outcomes on the assumption that the text extracted from published judgments of the Court bears a sufficient number of similarities with, and can thereforestand as a (crude) proxy for, applications lodged with the Court as well as for briefs submitted by parties in pending cases."[59]

Additionally, the authors consider the above-mentioned findings as evidence of legal realism theories, according to which judges make decisions largely on non-legal, rather than legal, grounds.[60] They conclude that 'the information about the case's factual background as expressed by the Court in relevant subsections of its judgment is the most critical component for obtaining on average the strongest predictive performance of the Court's decision outcome' and thus suggest that 'the rather robust correlation between the outcomes of cases and the text corresponding to fact patterns … coheres well with other empirical work on judicial decision-making in hard cases and backs basic legal realist intuitions'[61] In Section 7 below, the conclusion on the endorsement of legal realism theories will be examined in

[55] *Ibid.*
[56] *Ibid.* at 13.
[57] *Ibid.* at 14.
[58] *Ibid.* at 15.
[59] *Ibid.* at 2.
[60] *Ibid.* at 12.
[61] *Ibid.* at 16.

depth. This section discusses the methodology and findings of the research, as well as the assertion that the study demonstrates the feasibility of ex ante outcome prediction.

To begin, it is unclear whether aspects of the ECtHR judgements were included into the study's input. As mentioned before, the decision's operative portion, in which the Court announces the case's conclusion, is plainly omitted,[62] since the prediction exercise would be moot.

What is less apparent is whether or not the portion of the legal section that contains the Court's reasoning is included. The paper states that the objective was to 'ensure that the models did not include knowledge about the case's result, but this proviso seems to apply only to the operative parts of the rulings.[63] According to the research, the law part is included[64] and this often comprises the Court's legal rationale.[65]

If the legal reasoning of the Court is included into the data input, the study's overall prediction findings are unsurprising. After being told the Court's rationale, any experienced lawyer – and possibly the majority of non-lawyers – would be able to determine, in almost 100% of circumstances, whether the Court finds a violation or not. Thus, the study's total prediction rate of 79% must be considered in this light. Additionally, the study's claim to pave the way for plausible ex ante result prediction is considerably undermined by the inclusion of the Court's legal reasoning. Because the Court's reasoning is not known ex ante, it cannot be used to forecast future cases.

Second, one could wonder if the Court's decision's factual background section does not already include 'hints' about the decision's result. The analysis admits the possibility that the Court's formulation would be tailored to achieve a particular desired result.[66] Without implying any bias or lack of impartiality on the part of the ECtHR judges, the facts presented in the judgment may represent a selection of those facts necessary for the legal

[62] *Ibid.* at 8.
[63] *Ibid.*
[64] *Ibid.* at 8-10.
[65] *Ibid.* at 5.
[66] *Ibid.*

rationale and conclusion of the decision, setting aside other irrelevant facts asserted by the parties. As a result, one may express reservations about the study's assertion that the text extracted from published Court judgments bears a sufficient number of similarities to, and thus serves as a (brute) proxy for, applications filed with the Court as well as briefs submitted by parties in pending cases.[67]

Third, any ex ante prediction model must include the most commonly used phrases for diverse themes with a high predictive value as defined in the research.[68] This seems to be troublesome for a variety of reasons. Certain terms – such as 'result', 'employee','region', 'copy', or 'department' – seem to be random, and it's difficult to understand how they may be used to forecast the outcome of future instances ex ante. Others are very case-specific and would provide difficulties if employed in future projections, such as 'Ukraine', 'January', or 'Russian'. Using these terms to forecast future outcomes may result in information about those nations or dates having a decisive effect on the outcome. The implications of potential text-based prediction systems are examined in further detail below.

While the study's overall finding of 79 percent accuracy in predicting the outcome of ECtHR rulings is remarkable at first glance, a deeper examination of the methodology and assumptions applied casts doubt on the study's claims of probable ex ante outcome predictions.

3.2 Predicting Supreme Court rulings in the United States

Another set of academics concentrated on forecasting Supreme Court judgments in the United States and released their final findings in 2017.[69] While their analysis relied on prior research on US Supreme Court forecasts,[70] it was novel in numerous ways. To begin, the study's objective

[67] *Ibid.* at 2.
[68] *Ibid.*
[69] Daniel M. Katz, Michael J. Bommarito II & Josh Blackman, *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, 12(4) PloS One (2017).
[70] Guimerà & Sales-Pardo, *supra* n. 40; Martin et al., *supra* n. 40; Ruger et al., *supra* n. 40. *See also* Michael A. Bailey & Forrest Maltzman, *Does Legal Doctrine Matter? Unpacking Lawand Policy Preferences on the U.S. Supreme Court*, 102(3) Am. Pol. Sci. Rev. 369 (2008); Stuart M. Benjamin & Bruce A. Desmarais, *Standing the Test of Time: The Breadth of Majority Coalitions and the Fate of U.S. Supreme Court Precedents*, 4 J. Leg. Analysis

was to develop a model that would be generally and consistently relevant to all US Supreme Court judgments across time, not simply in a particular year or with a particular composition of the Court.[71] Second, the research followed the idea that 'all information required for the model to produce an estimate should be knowable prior to the date of the decision'.[72] As explained above, this is to guarantee that the model is capable of ex ante result prediction.

To accomplish these goals, the research used US Supreme Court cases spanning almost two centuries, from 1816 to 2015. This resulted in the inclusion of over 28,000 case results and over 240,000 individual justices' votes as input data.[73] Rather of relying on the textual information included in the judgments, like the ECtHR research did, this study labeled the data associated with each decision using specified attributes.[74] To begin, some characteristics are unique to the particular case, such as the parties' identities, the issues at stake, or the date of the judgment to be made. Second, other aspects elicit data from the lower court's ruling that must be reviewed. This includes, but is not limited to, the identification of the courts of origin (i.e. which circuit), the disposition and directives of the lower court, as well as which lower courts are divided on the matter at hand. Thirdly, another set of characteristics focuses on the Supreme Court's composition, including the justices' identities, their prior rates of reversal votes or dissents, as well as their political inclinations. Fourth, a final set of characteristics pertains to the Supreme Court's process, including the method in which the Court exercised jurisdiction and the grounds for

445 (2012); Lee Epstein et al., *Ideological Drift Among Supreme Court Justices: Who, When, and HowImportant*, 101 Nw. U. L. Rev. 1483 (2007); Edward D. Lee, Chase P. Broedersz & William Bialek, *Statistical Mechanics of the US Supreme Court*, 160 J. Statistical Physics 275 (2015); Andrew D. Martin & Kevin M. Quinn, *Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999*, 10(2) Pol. Analysis 134 (2002); Jeffrey A. Segal & Harold J. Spaeth, *The Supreme Court and the Attitudinal Model Revisited* (Cambridge University Press 2002); Jeffrey A. Segal & Harold J. Spaeth, *The Influence of Stare Decisis on the Votes of United States Supreme Court Justices*, 40 Am. J. Pol. Sci. 971 (1996); Jeffrey A. Segal et al., *Ideological Values and the Votes of U.S. Supreme Court Justices Revisited*, 57(3) J. Pols. 812 (1995); Carolyn Shapiro, *Coding Complexity: Bringing Lawto the Empirical Analysis of the Supreme Court*, 60 Hastings L.J. 477 (2008).

[71] Katz, Bommarito & Blackman, *supra* n. 69, at 2–3.

[72] *Ibid.* at 3.

[73] *Ibid.* at 5

[74] *Ibid.* at 4-6.

granting certiorari,[75] whether or not oral argument was planned and, if so, the period between the argument and the ruling.

The study's output objective was twofold: to forecast the result of decisions and to forecast each justice's vote.[76] The categorization problem was binary in nature, determining whether the Supreme Court overturned or confirmed the lower court's ruling.[77] In a few (though rare) instances, the Supreme Court declines to reconsider a lower court's ruling and instead resolves a matter as the original court of jurisdiction.[78] Those examples were omitted from the decision result prediction because they do not fit within the scope of a binary classification job.[79]

The researchers trained the model on a subset of the dataset using machine learning and then applied the model to the remaining, out-of-sample, data.[80] On average, the model predicted individual justices' votes with 71.9 percent accuracy and decision outcomes with 70.2 percent accuracy.[81] While year-to-year or decade-to-decade variation occurred, the research asserts that the model maintained 'stable performance' over time.[82] Additionally, the report asserts that the model 'significantly outperforms' other baseline comparison models.[83]

By comparing the research's methods and findings to its stated objective of developing a generic model for ex ante outcome prediction, it becomes clear that the study has many significant flaws.

---

[75] A petition for a writ of certiorari is the most common procedural device to invoke the US Supreme Court's appellate jurisdiction. *See* 28 U.S.C. § 1254(1), § 1257, § 1259. *See also* Steven M. Shapiro et al., *Supreme Court Practice* 59 et seq. (10th ed. 2013).
[76] Katz, Bommarito & Blackman, *supra* n. 69, at 4.
[77] *Ibid.*
[78] The US Supreme Court has original (i.e. acts as a court of first instance) exclusive jurisdiction over controversies between States, and concurrent original jurisdiction over proceedings involving ambassadors and certain other foreign officials, controversies between the United States and a State, and proceedings by a State against citizens of another State or aliens. 28 U.S.C. § 1251; *see also* US Const., Art. III, § 2.
[79] Katz, Bommarito & Blackman, *supra* n. 69, at 4.
[80] *Ibid.* at 5-8.
[81] *Ibid.* at 8-9.
[82] *Ibid.* at 9.
[83] *Ibid.* at 15.

To begin, although the research adheres to the idea that 'all information required for the model to produce an estimate should be knowable prior to the date of the decision,'[84] several of the input data attributes are only available soon before the decision is made. For example, if an oral argument is planned and, if so, how much time will pass between the argument and the judgment is often not revealed until late in the process.[85] This severely restricts the use of those characteristics for ex ante outcome prediction.

Second, the bulk of the labels on the input data pertain to appellate or Supreme courts charged with reviewing lower court rulings. As mentioned before, several characteristics of the study are connected to the lower court judgment under consideration (e.g., which circuit, the lower court's disposition and directives), as well as the Supreme Court justice's treatment of prior lower court decisions (e.g. reversal rates). Few of the input factors are unique to the disagreement, such as the parties' identities, the topics at stake, or the procedural steps leading up to the conclusion. As a consequence, it is debatable whether the approach or model can be used successfully in situations when the court resolves a disagreement rather than reviewing a lower court's judgment.

Thirdly, and somewhat relatedly, the prediction of decision result is limited to binary classification tasks involving whether the Supreme Court reverses or confirms the subordinate court's ruling. As previously stated, the analysis excludes situations in which the Supreme Court resolves a matter in its capacity as the initial court of jurisdiction. This is because 'the Court and its members may adopt technically nuanced stances, or the Court's judgment may somehow result in a complicated consequence that does not translate to a binary outcome,' the report states.[86] The same may be true for the vast majority of occasions in which a court resolves an issue on its own,[87] rather than reviewing another court's judgment. In certain instances, the court will be faced with technically sophisticated and nuanced issues of fact and law that are difficult to categorize using a binary approach. The topic of binary-tasks for AI models is examined in further detail below. At this point, suffice it to mention that the technique used in the research is not readily transferable to lower court rulings, which have the responsibility of resolving disputes rather than evaluating earlier decisions by another court.

[84] *Ibid.* at 3.
[85] *Ibid.* at 5.
[86] *Ibid.*
[87] *Ibid.*

Fourth, it's worth noting that Supreme Court rulings in general, and the US Supreme Court in particular, are sometimes very political. Justices of the United States Supreme Court are actually appointed based on their political leaning, among other factors.[88] The Supreme Court often decides on questions of law on which attorneys from opposing political parties disagree, such as the feasibility of gun regulation.[89] On the contrary, lower court decisions are often more fact-driven and less legally grounded. As a result, some factors (for example, the judge's political inclination) are less likely to be outcome-determinative, or at the very least, the relationship between the trait and the conclusion will be less straightforward.

As a result, the aforementioned research have significant intrinsic limitations in terms of their generalizability for ex ante outcome prediction. Nonetheless, they raise the possibility that AI-driven and machine learning-based outcome prediction tools might complement human decision-making. Max Radin said in 1925 that the judge's 'business is prophecy, and if prophecy were certain, there would not be much credit in prophesying'.[90] If AI models are capable of forecasting or assisting in prediction, shouldn't they be used to replace, or at the very least supplement, human decision-makers? The subsequent portions of this paper will attempt to address this issue.

4. Limitations on the use of AI in Legal Decision Making: The Four 'V'S Of Big Data

The four Vs of Big Data — Volume, Variety, Velocity, and Veracity – are often referred to as the pillars of data-driven initiatives by data professionals.[91] The four Vs denote the difficulties associated with the

[88] Neal Devins & Lawrence Baum, *Split Definitive: How Party Polarization Turned the Supreme Court into a Partisan Court*, 2016 Sup. Ct. Rev. 301, 331 (2016).
[89] *See District of Columbia v. Heller*, 554 U.S. 570 (2008); *McDonald v. Chicago*, 561 U.S. 742 (2010). However, challenging the assumption that US Supreme Court justices vote on the basis of one-dimensional policy preference, *see* Joshua Fischman, *Do the Justices Vote Like Policy Makers? Evidence from Scaling the Supreme Court with Interest Groups*, 44 J. Legal Stud. S269 (2015).
[90] Max Radin, *The Theory of Judicial Decision: Or HowJudges Think*, 11 ABA J. 357, 362 (1925).
[91] Initially, the focus was on only three Vs (volume, variety, and velocity). *See* e.g. Max N. Helveston, *Consumer Protection in the Age of Big Data*, 93 Wash. U. L. Rev. 859, 867 (2016). Veracity was added in the mid-2000s. *See also* Margaret Hu, *Small Data*

utilization of Big Data. Additionally, they aid in the evaluation of data-driven AI algorithms such as those discussed in the preceding section, as well as their use in the legal field. This section covers the four Vs in turn and the inherent limits of data-driven models for legal decision-making using artificial intelligence.

4.1 Volume - Requirement for an adequate amount of non-confidential case data

Any data-driven AI program need data access first and foremost. Machine learning models, which are built on probabilistic conclusions, are data-hungry: the higher the sample size, the more accurate the prediction value of the model. The amount of data necessary in the legal field may impose a dual constraint for AI algorithms.

To begin, case data is not always readily available. Certain aspects of the law are secret, and hence not accessible to other parties. Confidentiality may be justified on the basis of safeguarding the interests of concerned parties or the underlying transactions. For example, international business arbitration decisions are seldom disclosed, making the creation of a database for the purpose of developing an AI model challenging.[92] However, this does not exclude the use of AI models in international commercial arbitration. There are initiatives to regularly publicize commercial prizes, often in a redacted version.[93] In any case, even if organizations do not announce secret awards,

Surveillance v. Big Data Cybersurveillance, 42 Pepp. L. Rev. 773, 795 (2015); Todd Vare & Michael Mattioli, *Big Business, Big Government and Big Legal Questions*, 243 Managing Intell. Prop. 46 (2014). More recently, some have suggested a fifth V in the form of 'value'. *See* e.g. Amy Affelt, *Big Data, Big Opportunity*, 21 Austl. L. Libr. 78 (2013). In the legal context, this last point is of less relevance and thus not discussed here.

[92] Queen Mary School of International Arbitration Survey, *The Evolution of International Arbitration* 3, 24 (2018) ('87% of respondents believe that confidentiality in international commercial arbitration is of importance'); Queen Mary School of International Arbitration Survey, *Improvements and Innovations in International Arbitration* 6 (2015) (respondents cited 'confidentiality and privacy' as one of the top five most valuable characteristics of international arbitration, with the in-house counsel subgroup rating it as the second most valuable characteristic).

[93] *See* e.g. ICC, *Note to Parties and Arbitral Tribunals on the Conduct of the Arbitration Under the ICC Rules of Arbitration*, paras 42–43 (1 Jan. 2019), https://cdn.iccwbo.org/content/uploads/sites/3/2017/03/icc-note-to-parties-and-arbitral-tribunals- on... (accessed 15 March 2022).

they may gather and make them accessible for the purpose of developing AI models.

Second, when case data is available, it is critical to have a high sample size. While there is no hard and fast rule about the sample size necessary, the more data available, the more accurate the derived model. As a result, areas of law with a high volume of judgements on a single subject will be more amenable to AI models. Although there are no reliable statistics on the number of awards rendered each year in international investment arbitration, on the basis of approximately sixty new cases initiated each year[94] the number of arbitral awards should also be in the double digits,[95] which does not make for a particularly large sample size.

4.2 Variety: Repetitive patterns with binary outcomes are required

Along with the required data amount, there is a concern concerning the diversity of the supplied data. Range of data, in data-research parlance, refers to the fact that data originates from a variety of sources and may be structured (e.g., a file including names, phone numbers, and addresses) or unstructured (photos, videos, social media feeds).[96] In a legal environment, the same question is almost always phrased differently. The variation will come not so much from alternative sources or formats – given that the incoming data will almost certainly be confined to prior judgments – but rather from the substance addressed in those decisions. Two distinct issues come to mind when considering AI-driven decision-making.

[94] According to UNCTAD statistics, sixty-two new treaty-based investor–State dispute settlement cases were initiated in 2016, sixty-five in 2017 and at least seventy-one in 2018. *See* UNCTAD, *Investor-State Dispute Settlement: Reviewof Developments in 2016* 1 (May 2017); UNCTAD, *Investor-State Dispute Settlement: Reviewof Developments in 2017* 1 (June 2018); UNCTAD, *NewISDS Numbers: Takeaways on Last Year's 71 Known Treaty-Based Cases* (13 Mar. 2019), https://investmentpolicyhubold.unctad.org/News/Hub/Home/1609 (accessed 15 April 2022).

[95] *Ibid.*

[96] *See* e.g. EY, *Big Data: Changing the Way Businesses Compete and Operate*, Rpt. 2 (Apr. 2014); Lieke Jetten & Stephen Sharon, *Selected Issues Concerning the Ethical Use of Big Data Health Analytics* 72 Wash. & Lee L. Rev. Online 486, 487 (2016); Uthayasankar Sivarajah et al., *Critical Analysis of Big Data Challenges and Analytical Methods*, 70 J. Bus. Research 263, 269 (2017).

The first challenge concerns data intake and whether AI-based decision-making models need repeated fact patterns or, alternatively, if they can handle complicated and non-repetitive themes. The computer algorithm used in the aforementioned research on US Supreme Court rulings was built for decisions spanning almost two centuries and addressing a wide array of concerns.[97] Nonetheless, the greater the number of outliers or non-repetitive situations, the greater the difficulty for the AI model. Thus, AI systems are more likely to be used in international arbitration to international investment arbitration (which often presents a number of well-known concerns) than to international commercial arbitration (which deals with diverse and often unique issues).

The second query concerns the model's output. The preceding research on legal prediction all employ binary classification as the output task. The ECtHR's binary classification duty was to determine whether or not a breach of the relevant Convention provision occurred, while the US Supreme Court's binary classification goal was to determine whether or not the Court confirmed the lower court's ruling. As previously stated, this begs the issue of whether those models, or others of a similar kind, may be used to do more diversified, non-binary activities.[98]

One would object that each legal judgment can be broken into a plethora of binary classification jobs, such as determining whether (1) the tribunal has jurisdiction: yes/no; (2) the parties entered into a contract validly: yes/no; and (3) one party violated the contract: yes/no. Lord Hoffman famously used a binary analogy to explain a standard of evidence issue:

"If a legal rule requires a fact to be proved (a 'fact in issue'), a judge or jury must decide whether or not it happened. There is no room for a finding that it might have happened. The law operates a binary system in which the only values are 0 and 1. The fact either happened or it did not. If the tribunal is left in doubt, the doubt is resolved by a rule that one party or the other carries the burden of proof. If the party who bears the burden of proof fails to discharge it, a value of 0 is returned and the fact is treated as not having

[97] Katz, Bommarito & Blackman, *supra* n. 69, at 4.
[98] *Ibid.*

happened. If he does discharge it, a value of 1 is returned and the fact is treated as having happened."[99]

While it is true that many legal matters of fact or law may be reduced to a 0/1 or yes/no binary work, the issue is that each case will have a plethora of such binary tasks, and resolving them all will be situation-specific. To enable an AI model to extract the necessary patterns and algorithms from the input data, having a single distinct output question simplifies the model-building process. This is why the research group expressly omitted from the examination of US Supreme Court judgments those in which the Supreme Court was the court of original jurisdiction and did not correspond to a straightforward binary categorization job.[100]

4.3 Velocity: The issue of policy evolution through time

The term "velocity" refers to the rate at which data is received and processed. Big Data is often difficult to manage due to the sheer volume and frequency of incoming data. Such a danger is quite minimal in a legal situation. As previously stated, the issue is likely to be one of scarcity rather than availability of data in terms of volume.[101] As a result, choices may become less frequent over time, and when they do occur, there may have been a change in policy, rendering past data obsolete. At times, these policy shifts might be dramatic and abrupt. To use an international arbitration example, the Court of Justice of the European Union's judgment in *Achmea* overnight substantially altered the compatibility of investor-state arbitration with European law.[102]

This begs the issue of how AI models, which are defined by their reliance on knowledge derived from prior data, can cope with such policy changes. True, the core of machine learning is the capacity to continuously improve the algorithm. Nonetheless, such enhancements are always based on historical data. Changes in policy need deviations from historical data, i.e. prior instances. As a result, AI models are likely to maintain 'conservative' methods consistent with prior examples.

[99] *In re B* [2008] UKHL 35.
[100] Katz, Bommarito & Blackman, *supra* n. 69, at 4.
[101] Lieke Jetten & Stephen Sharon, *Selected Issues Concerning the Ethical Use of Big Data Health Analytics* 72 Wash. & Lee L. Rev. Online 486, 487 (2016).
[102] Case C-284/16 *Slovak Republic v. Achmea B.V.* (CJEU, 6 Mar. 2018).

4.4 Veracity: Bias risk and susceptibility to data diet

Finally, veracity refers to the data's correctness and dependability. In the context of artificial intelligence, the concern is if there are any hidden data vulnerabilities that might compromise the model's accuracy. The robustness and trustworthiness of AI are often discussed in discussions about the technology.[103]

To begin, one would think that AI models have the benefit of algorithmic impartiality and infallibility over humans, who will unavoidably make errors and are impacted by subjective, non-rational elements. Humans often behave irrationally, as shown by research in psychology, cognitive science, and economics.[104] Most notably, Nobel laureates Daniel Kahneman and Amos Tversky have conducted research on heuristics and cognitive biases in human decision-making.[105] Their investigations demonstrate several instances in which heuristics (i.e. cognitive shortcuts for otherwise intractable issues) and biases (i.e. elements that seem to be unrelated to the quality of our choices but have an effect on them) emerge in everyday human judgments.[106]

---

[103] *See* e.g. European Commission Press Release, *Artificial Intelligence: Commission Takes Forward Its Work on Ethics Guidelines* (8 Apr. 2019), http://europa.eu/rapid/press-release_IP-19- 1893_en.htm (accessed 15 April 2022).

[104] *See* e.g. Christine Jolls, Cass Sunstein & Richard Thaler, *A Behavioral Approach to Lawand Economics*, 50 Stan. L. Rev. 1471 (1998); Avishalom Tor, *The Methodology of the Behavioral Analysis of Law*, 4 Haifa L. Rev. 237 (2008). Regarding the idea of ecological rationality (rationality is variable and depends on the context), *see* e.g. Vernon L. Smith, *Constructivist and Ecological Rationality in Economics*, 93(3) Am. Econ. Rev. 456 (2003).

[105] *See* e.g. Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 Cognitive Psychol. 430, 431 (1972); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 Science 1124 (1974); Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 Cognitive Psychol. 207 (1973). Further research has emphasized the fact that the use of intuitive, non-rational decision-making is both a source of error and a factor of success for humans in their daily choices, and that humans have at least an intuitive logical and probabilistic knowledge. *See* e.g. Wim De Neys, *Bias and Conflict: A Case for Logical Intuitions*, 7(1) Persps Psychological Sci. 28 (2012); Jonathan Evans & Keith E. Stanovich, *Dual-Process Theories of Higher Cognition Advancing the Debate*, 8(3) Persps Psychological Sci. 223 (2013).

[106] For instance, a series of studies on the so-called anchor-effect has shown that people, when estimating an unknown quantity, are affected by a number given to them, even if it is obvious that this number is random. *See* Daniel Kahneman, *Thinking, Fast and Slow*119–128 (Penguin 2011). *See also* Edna Sussman, *Biases and Heuristics in Arbitrator Decision-*

By applying this study to the legal sector, a team of Israeli and US scholars provided insight on the significance of external influences in judicial decision-making.[107] The research examined over 1,100 judgments made by Israeli judges during a ten-month period in regard to 40% of the country's parole petitions[108] and found that the majority of applications are refused on average.[109] However, the likelihood of a favorable verdict is much greater immediately after the judge's regular meal break.[110] While the findings do not support the widely held adage that 'justice is what the judge had for breakfast,' they do 'indicate that judicial judgments may be impacted by whether the judge took a break to eat'.[111] This study demonstrates how extraneous circumstances like as meal breaks, which should be unrelated to the merits of the case, may alter human decision-making.[112]

As a result, some writers have claimed that AI-based decision-making is better to human decision-making because computers are immune to cognitive biases and the excessive effect of external circumstances.[113] However, an unquestioning deference to algorithmic impartiality and infallibility is misguided. Over the last several years, research on artificial intelligence has emphasized the dangers of misbehaving or biased algorithms. Significant research address issues about bias in computer systems that perform a range of functions, including travel listings, credit ratings, and online advertising.[114] Referring to a 'scored society,' some have

---

Making: Reflections on Howto Counteract or Play to Them, in *The Roles of Psychology in International Arbitration* (Tony Cole ed., Wolters Kluwer 2017).

[107] Shai Danziger et al., *Extraneous Factors in Judicial Decisions*, 108(17) PNAS 6889 (2011).

[108] *Ibid.* at 6889.

[109] *Ibid.*

[110] *Ibid.* at 6890.

[111] *Ibid.* at 6892.

[112] *See also* Chris Guthrie, Jeffrey Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 Cornell L. Rev. 777 (2001).

[113] Hanke, *supra* n. 1, at 8.

[114] *See* e.g. Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM Transactions on Information Systems 330 (1996); Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* (paper presented to the Data and Discrimination: Converting Critical Concerns into Productive Inquiry Preconference of the 64th Annual Meeting of the International Communication Association, 22 May 2014); Latanya Sweeney, *Discrimination in Online Ad Delivery*, 11(3) ACM Queue 10 (2013); Nicholas Diakopoulos, *Algorithmic Defamation: The Case of the Shameless Autocomplete*, Nick Diakopoulos (6 Aug. 2013), www.nickdiakopoulos.com/2013/08/06/algorithmic-defamation-the-case-of-the-shameless-autocomplete (accessed 9 March 2022).

suggested that uncontrolled and secret algorithms generate authoritative ratings for people that serve as a conduit for access to possibilities.[115] Other scholars have stated that 'procedural consistency is not synonymous with impartiality.'[116]

Any data-driven computer model is only as good as the data it is fed. Data diet vulnerability has a detrimental effect on the derived model. The underlying data used to train the algorithm, in particular, may have been 'contaminated' with human prejudices. The machine learning algorithm will be guided by these biases and may even exacerbate them by assuming them to be 'true' for future choices or result predictions.

For example, in the field of investment arbitration, concerns have been expressed that arbitral courts are intrinsically and excessively pro-investor.[117] I will not debate the validity of this objection here,[118] but will instead assume for the sake of this proof that such human bias occurs. In this situation, an AI model built on investment arbitration data will very certainly perpetuate such (claimed) investor favoritism. In a disproportionate number of circumstances, the model would likely anticipate favorable results for investors versus States.

Even if the model does not explicitly refer to human biases in the underlying data, it may draw patterns from it and extrapolate them in ways that might result in systemic errors. For example, research in the United States have shown that the use of algorithms in criminal risk assessment results in

[115] Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1 (2014).
[116] Osoba & Welser IV, *supra* n. 16, at 2.
[117] *See* e.g. Pia Eberhardt et al., *Profiting from Injustice: HowLawFirms, Arbitrators and Financiers Are Fuelling an Investment Arbitration Boom* 8 (Corporate Europe Observatory 2012); George Kahale III, *Is Investor-State Arbitration Broken?*, 9(7) Transnat'l Disp. Mgmt. 1, 1–2 (2012); Gus van Harten, *Part IV Chapter 18: Perceived Bias in Investment Treaty Arbitration*, in *The Backlash Against Investment Arbitration* 433 (Michael Waibel et al. eds, Wolters Kluwer 2010).
[118] *See* e.g. Gloria Maria Alvarez et al., *A Response to the Criticism Against ISDS by EFILA*, 33(1) J. Int'l Arb. 1, 4 (2016); Carolyn B. Lamm & Karthik Nagarajan, *The Continuing Evolution of Investor- State Arbitration as a Dynamic and Resilient Form of Dispute Settlement*, V(2) Indian J. Arb. L. 93, 96–97 (2016); Stephen M. Schwebel, *Keynote Address: In Defence of Bilateral Investment Treaties*, in *Legitimacy: Myths, Realities, Challenges*, 18 ICCA Congress Series 1, 6 (Albert Jan van den Berg ed., Wolters Kluwer 2015).

racially skewed outcomes.[119] In the United States, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) method is commonly utilized to analyze offenders' recidivism risks. According to studies, '[b]lack defendants were... twice as likely as white defendants to be misclassified as having a higher risk of violent recidivism,' whereas 'white violent recidivists were 63 percent more likely to have been misclassified as having a low risk of violent recidivism, compared to black violent recidivists.[120] It is unknown if the computer program's racial prejudice was based on pre-existing human biases in the training data. It might also be that the algorithm incorrectly identified black offenders as having a greater recidivism rate due to this ethnic group's overrepresentation in specific types of offenses. The computer model may have made the incorrect assumption of a greater recidivist risk based on this pattern.

The possibility of systemic mistakes occurring as a result of hidden patterns in the underlying data is a significant danger. As shown before, in the analysis on ECtHR rulings, terms such as 'Ukraine' or 'Russian' have a strong predictive value.[121] This is most likely because a substantial percentage of ECtHR lawsuits are aimed towards and resolved against these nations.[122] According to statistics, a handful of countries get the greatest number of applications and condemnations.[123] A computer algorithm built on data having a larger percentage of condemnations against a particular country may extrapolate a higher probability of the country committing a violation in the future, biasing its result predictions against the country.

It is consequently critical to evaluate if and how to resolve systemic errors in algorithms. In systems where the algorithm is written by a human programmer, the error is often made in the algorithm's design. Once the error is discovered, it may be corrected.[124] On the other hand, in machine

[119] Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, ProPublica (23 May 2016), www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 9 May 2019); Jeff Larson et al., *HowWe Analyzed the COMPAS Recidivism Algorithm*, ProPublica (23 May 2016), www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed 15 April 2022).
[120] Jeff Larson et al., *supra* n. 19, at 2.
[121] Aletras et al., *supra* n. 45, at 6.
[122] *Ibid.*
[123] European Court of Human Rights, *Violations by Article and by State, 1959–2018* (2018) (finding that Turkey and the Russian Federation lead the list of countries with most judgments having found at least one violation of the Convention).
[124] Friedman & Nissenbaum, *supra* n. 114, at 331.

learning systems, the method is retrieved from the sample set's data, as mentioned above. Thus, errors are more likely to occur as a function of the input data and are thus more difficult to identify and correct.[125] Hiding sensitive information in the input date, such as ethnic origin or geographic origin, should be explored to assist avoid problems. Even if crucial traits are concealed, algorithms may implicitly reconstruct them using proxy variables.[126]

Additionally, as noted before, the goal of machine learning is for computer systems to learn from their experiences and continuously improve their performance. As a result, the algorithm is impacted not only by the first training dataset, but also by subsequent data input. As a result, users have some 'control' over the algorithms. A prominent example is the AI-chatbot Tay's cursing habit and other undesirable behavior in response to interactions with its Twitter followers.[127] One may also foresee individuals attempting to excessively influence or rig the algorithms in order to acquire favorable outcomes in a legal environment.[128] For example, if it was obvious that specific terms or clusters of words, such as those used in the ECtHR decision research, resulted in a favourable case prediction, the targeted use of such words in a party's legal filings may result in an improper impact on the outcome.

By and large, this section has shown that the employment of AI algorithms for legal decision-making has a number of intrinsic constraints. These constraints must be carefully explored before advocating for the use of artificial intelligence in this setting. Additionally, other more basic and far-reaching difficulties exist, which are described in the following sections.

[125] Jeff Larson et al., *supra* n. 19, at 2.
[126] Simon DeDeo, *Wrong Side of the Tracks: Big Data and Protected Categories* (2015), https://arxiv.org/pdf/1412.4643v2.pdf (accessed 15 April 2022) (for instance, income might be inferred from proxy variables such as postal codes).
[127] Ian Johnston, *AI Robots Learning Racism, Sexism and Other Prejudices from Humans, Study Finds*, The Independent (13 Apr. 2017), www.independent.co.uk/life-style/gadgets-and- tech/news/ai-robots-artificial-intelligence-racism-sexi... (accessed 15 April 2022) (Microsoft chatbot called Tay was given its own Twitter account and allowed to interact with the public; after twenty-four hours the chatbot used sexist, racist and profane language which it had learned from interaction with other Twitter users).
[128] *Ibid.*

5. The Dark Side of AI-Assisted Legal Decision-Making: The Importance of Reasoned Decisions

One of the key characteristics of legal decision-making is the ability to provide a reasoned judgment that details the premises upon which it is founded. One might conceptualize different aims for offering justifications in legal judgements. To begin, justifications assist the losing side in comprehending why it lost and making the decision more palatable (legitimacy objective). Second, reasons enable the disputants, and, if the decision is made public, third parties in comparable circumstances, to change their behavior in the future (incentive objective). Thirdly, reasons enable subsequent decision-makers to adopt the same reasoning or justify their deviation from it (consistency objective). While one may debate whether there is a market for unreasoned choices (e.g., in certain cases, parties may want 'quick-and-dirty' unreasoned decisions), legal decisions must include justifications unless the parties agree differently.

AI systems will have major challenges when it comes to making reasonable legal judgements and adhering to those rationales.[129] Indeed, questions have been expressed about the difficulty of AI systems to explain their outcomes, not only in the legal field, but more widely.[130] For instance, alarming findings were achieved from an AI software that was able to deduce a person's sexual orientation based on publicly available profile images.[131] While the accuracy percentages are concerning (83 percent for women and 91 percent for males), what is more concerning is the researchers' inability to determine the basis for the AI program's conclusions.[132] This exemplifies the broader issue confronting AI research: the so-called explainability or interpretability of its findings.[133]

---

[129] *See* Scherer, *supra* n. 4, at 511–12.

[130] *See* e.g. Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise*, 34:1 Berkeley Tech. L.J. 143 (2019).

[131] Michal Kosinski & Yilun Wang, *Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images*, 114 J. Personality & Soc. Psychol. 246 (2018).

[132] Cliff Kuang, *Can A.I. Be Taught to Explain Itself?*, New York Times (21 Nov. 2017), www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html (accessed 15 April 2022).

[133] Or Biran & Courtenay Cotton, *Explanation and Justification in Machine Learning: A Survey*, in *IJCAI-17 Workshop on Explainable AI (XAI) Proceedings* 8 (2017), https://pdfs.semanticscholar.org/02e2/e79a77d8aabc1af1900ac80ceebac20abde4.pdf

This challenge arises as a result of the characteristics of particular AI models. As mentioned before, expert models or decision trees adhere to pre-established norms.[134] It is therefore feasible to determine the reasons of a particular outcome using those principles, thus explaining the model.[135] On the other hand, as previously stated, other machine learning models, such as neural networks, frequently lack predefined rules and instead rely on pattern recognition to extract the required algorithm.[136] These systems may include hidden units that correlate to unobserved properties.[137] As a result, the mechanism through which such AI models produce outcomes is 'black-boxed' and difficult to describe.[138]

AI research is attempting to address these difficulties by developing Explainable Artificial Intelligence, or XAI.[139] One approach is to use counterfactual situations. The model picks alternative samples with distinct characteristics, compares the resulting results, and thereby determines how and why they vary.[140] For example, the model will be able to determine if the result would have been different in a particular situation had feature X been removed or feature Y been introduced. In other words, the model used to make the actual choice is complemented by another model that serves as an explanation.[141]

(accessed 23 March 2022) (defining interpretability as the ability for humans to understand operations either through introspection or through a produced explanation).

[134] Alpaydin, *supra* n. 18, at 14.

[135] *See* e.g. Bruce G. Buchanan & Edward H. Shortlie, *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley 1984).

[136] Alpaydin, *supra* n. 18, at 98.

[137] *Ibid.* at 100.

[138] *Ibid.* at 155.

[139] *See* earlier on Bruce Chandrasekaran, Michael C. Tanner & John R. Josephson, *Explaining Control Strategies in Problem Solving*, 4(1) IEEE Expert 9 (1989). *See* more recently Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 Harv. J.L. & Tech. 842 (2018). *See also* DARPA, *Explainable Artificial Intelligence (XAI) Program*, www.darpa.mil/program/explainable- artificial-intelligence (accessed 9 May 2019), full solicitation at www.darpa.mil/attachments/DARPA-BAA-16-53.pdf (2016) (accessed 15 March 2022); George Nott, *'Explainable Artificial Intelligence': Cracking Open the Black Box of AI*, Computer World (10 Apr. 2017), www.computerworld.com.au/article/617359/ (accessed 15 March 2022).

[140] Charlotte S. Vlek et al., *A Method for Explaining Bayesian Networks for Legal Evidence with Scenarios*, 24 Artificial Intelligence L. 285 (2016).

[141] *See* e.g. Michael Harradon, Jeff Druce & Brian Ruttenberg, *Causal Learning and Explanation of Deep Neural Networks via Autoencoded Activations* (2018),

The challenge in providing AI-generated reasoned legal judgements is twofold. To begin, it may be challenging to determine the exact components that contributed to a certain result prediction in the case of black-boxed models. Second, even if specific elements may be identified as causes of a certain result prediction, they may not provide an adequate explanation. For example, specific terms or clusters of words were discovered as having a high predictive value in the aforementioned research on ECtHR choices.[142] However, the evidence that phrases such as 'injury', 'Ukraine', 'copy', or 'January' contributed to the result prediction falls short of providing an explanation necessary for a legally reasoned judgment.

It is critical to distinguish between causal attribution, which is the process of extracting and displaying a causal chain to a person, and causal explanation, which is the social process of knowledge transfer between the explainer and the explainee with the goal of the explainee having the information necessary to understand the causes of an event.[143] The latter needs not only the identification of reasons, but also the provision of contextual explanation. Miller has shown that a good AI explanation must consequently take the human addressee into consideration.[144] This implies, among other things, that explanation selection is critical: normally, only a tiny subset of all potential causes is appropriate for an individual.[145] For example, based on the findings of the ECtHR research, the fact that an incident occurred in 'January' may be a factor in the choice, but it is a less relevant explanation than the fact that it constituted 'ill treatment'.

---

https://arxiv.org/abs/1802.00541 (accessed 9 May 2019); Bradley Hayes & Julie A. Shah, *Improving Robot Controller Transparency Through Autonomous Policy Explanation*, in *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2017); Pat Langley et al., *Explainable Agency for Intelligent Autonomous Systems*, in *Proceedings of the Twenty-Ninth Annual Conference on Innovative Applications of Artificial Intelligence* 4762 (AAAI Press 2017); Marco T. Ribeiro, Sameer Singh & Carlos Guestrin, *Why Should I Trust You?: Explaining the Predictions of Any Classifier*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135 (ACM 2016).

[142] Aletras et al., *supra* n. 45, at 6.

[143] Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, 267 Artificial Intelligence 1, at 17–18, 20 (2019).

[144] *Ibid.*

[145] *See* e.g. Denis J. Hilton, *Social Attribution and Explanation*, in *Oxford Handbook of Causal Reasoning* 645 (Michael Waldmann ed., Oxford University Press 2017).

Generally, an explanation is offered in terms of the explainer's beliefs regarding the explainee's views.[146] Dworkin has underlined the critical nature of law's common environment. He created a philosophy of law as an interpretative process occurring among a community of interpreters in his magnum opus, Law's Empire.[147] Drawing on hermeneutical tradition, Dworkin argues that comprehending a social activity, such as law, needs a focus on the meaning it has for participants. Thus, the meaning of law can be recovered only within a common framework.[148] These contextual variables are expected to provide difficulties for legal explanation or reasoning based on artificial intelligence.

Additionally, social scientists have conducted research on the utility of probabilistic explanations.[149] By and large, the utilization of statistical or probabilistic links does not provide the same level of satisfaction as causal explanations. For example, if a student receives a 50/100 on an exam and inquires as to why, the teacher's response that the majority of the class achieved the same result is unlikely to satisfy the student. While discussing why the majority of students achieved this score is an improvement, it pales in comparison to explaining what this individual kid did to get this outcome.[150]

This example demonstrates the difficulty inherent in providing explanations or justifications for AI decision-making, which is often based on statistical or probabilistic models, as explained above.[151] Providing a 'explanation,' such as the 86 percent probability of a claim being rejected, would not satisfy the losing party. It does not satisfy any of the legal reasoning goals listed at the beginning of this section. To begin, the legitimacy aim is not accomplished, since statistical data is unlikely to assist the losing side in comprehending why it lost and therefore making the decision more palatable. Second, the incentive purpose fails since statistical data also prevents parties or third parties from adapting their behavior in the future. Finally, the consistency aim is not met because other decision-makers lack knowledge on why they should adhere to or diverge from the same logic.

---

[146] *See* e.g. Denis J. Hilton, *Conversational Processes and Causal Explanation*, 107(1) Psychol. Bull. 65 (1990).
[147] Ronald Dworkin, *Law's Empire* (Fontana 1986).
[148] *Ibid.*
[149] John R. Josephson & Susan G. Josephson, *Abductive Inference: Computation, Philosophy, Technology* (Cambridge University Press 1996).
[150] Miller, *supra* n. 143, para. 4.5.2.
[151] Alpaydin, *supra* n. 18, at 14.

Thus, the need for reasoned choices is likely to be a significant hurdle to AI-assisted legal decision-making. The influence of AI models' probabilistic nature, on the other hand, raises even more basic problems about the decision-making paradigm as a whole, as explored in the next section.

6. A Legal Decision-Making Paradigm Change: Probabilistic Inference Instead of Deductive Reasoning and Logic?

Evaluating whether AI is capable of assisting humans in making legal judgments unavoidably raises the topic of how people make legal decisions. As early as 1963, Lawlor predicted that computers will eventually be able to analyze and forecast court judgments, but cautioned that accurate prediction would need a 'scientific' knowledge of how the law and facts affect the judges' choice.[152] Even today, a 'scientific' knowledge of judicial decision-making is missing and is a source of contention among legal philosophers and theorists.

There are numerous theories of judicial decision-making, but a fundamental distinction exists between those that postulate the use of logic through deductive reasoning based on abstract, predetermined legal rules (collectively referred to as legal formalism) and those that emphasize the importance of extra-legal factors and the political dimension of the law (regrouped in the category of legal realism). This section demonstrates that the application of artificial intelligence in legal decision-making does not simply fall into either category. As this section demonstrates, AI models would raise probabilistic inferences to the level of legal decision-making, resulting in a dramatic paradigm change.

6.1 **Formalism in law and the use of deductive reasoning and logic**

In its purest form, legal formalism asserts that law is and should be a totally self-contained system in which judges are never confronted with decisions

[152] Reed C. Lawlor, *What Computers Can Do: Analysis and Prediction of Judicial Decisions*, 49 ABA J. 337 (1963).

or issues of interpretation that may be resolved via extra-legal factors.[153] Rather than that, as Max Weber put it, 'every concrete decision [is] the "application" of an abstract proposition to a concrete fact situation' and 'it must be possible in every concrete case to derive the decision from abstract legal propositions by means of legal logic'.[154]

Thus, a judicial judgment is the result of what seems to be a mechanical or mathematical application of pre-existing legal principles or norms to known facts via the use of logic.[155] The fundamental concept may be represented simply as 'R + F = C' or 'rule plus facts equals conclusion'.[156] More precisely, the legal syllogism will consist of a major premise in the form of a pre-established rule (e.g., 'if P then Q') and a minor premise attempting to establish that the required condition stipulated in the major premise (P) actually occurred. If such a requirement is satisfied, the judge concludes by deductive reasoning or subsumption that the legal consequence is satisfied. As a matter of logic, (Q) must be applied in this instance.[157]

Although 'pure' formalists are rare today, the central idea of legal decision-making as based on deductive reasoning and logic continues to be influential. Hart established a critical distinction in his seminal work The Concept of Law between clear cases, for which simple deductive reasoning applies, and hard cases, for which extra-legal moral and political considerations may come into play.[158] Drawing on Wittgenstein's philosophy, Hart stresses the indeterminacy of natural phenomena. Language and the open texture of the law, for example, by the use of broad principles such as 'good faith'.[159]

[153] Hans Kelsen, *Reine Rechtslehre* 478 (2d ed., Deuticke 1960).
[154] Max Weber, *Wirtschaft und Gesellschaft (Economy and Society)* 657–58 (Tü bingen 1922).
[155] French jurist Jean Domat saw the law as a logical, 'geometrical' demonstration, as any other scientific demonstration. *See* e.g. Marie-France Renoux-Zagamé, *La figure du juge chez Domat*, 39 Droits 35 (2004); Marie-France Renoux-Zagamé, *Domat, Jean*, in *Dictionnaire Historique des Juristes Français* (Patrick Arabeyre, Jean-Louis Halpérin & Jacques Krynen eds, Presses universitaires de France 2007).
[156] Neil MacCormick, *Legal Reasoning and Legal Theory* x (Oxford Clarendon 1977) (with revised foreword, 1994).
[157] *Ibid.* at 21-29.
[158] H. L. A. Hart, *The Concept of Law*(Oxford Clarendon 1961).
[159] *Ibid.*

Even in its more sophisticated manifestations, legal formalist theories continue to emphasize deductive, logical, rule-based reasoning as the guarantee of law's objectivity, impartiality, and neutrality. In 1994, MacCormick wrote:

"A system of positive law, especially the law of a modern state, comprises an attempt to concretize broad principles of conduct in the form of relatively stable, clear, detailed and objectively comprehensible rules, and to provide an interpersonally trustworthy and acceptable process for putting these rules into effect. […] [T]he logic of rule-application is the central logic of the law within the modern paradigm of legal rationality under the 'rule of law."[160]

If implemented in a legal setting, AI procedures might possibly contradict this concept of legal decision-making. As discussed in Section 2, certain computer models (for example, expert models) are truly rule-based, relying on causal logic and deductive reasoning to apply pre-established rules to observable data. Other AI models, on the other hand, have distinct characteristics. Machine learning models, in particular, such as neural networks, often lack predefined rules. Due to the fact that the machine learning software derives the algorithm from the observable data, deductive, causal reasoning is therefore substituted by an inverse approach. Rather of relying on logic, the AI model assesses probabilities, or the probability of any given result.[161]

Applying such machine learning methods to legal decision-making would therefore imply a break from the above-mentioned formalist view of judicial reasoning. A judgment made using such AI models would not be founded on predetermined legal principles, would not be the outcome of deductive reasoning, and would not follow the legal syllogism stated above. While this circumstance is troubling for legal formalists, it may be seen as vindication for those who have long challenged formalist ideas.

6.2 Legal realism and the significance of non-legal variables

[160] MacCormick, *supra* n. 156, at ix–x.
[161] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, 24 et seq. (Knopf 2017).

Over time, legal formalism has come under fire. Legal realists criticized the basic postulates of formalist theories in the first half of the twentieth century. [162] While realist theories differ greatly, they have several characteristics. Llewelyn and others argued against the notion that law was a mechanical application of predetermined rules by a judge using logic and logical reasoning.[163] Recognizing that legal certainty was a fantasy, realists such as Frank created what they termed rule skepticism, highlighting the reality that rules do not play a decisive role in legal decision-making.[164] Rather than that, judges determine cases based on irrelevant non-legal elements or their 'hunches' and then cover their judgment with an ostensibly rational rule-deferring coating ex post.[165] By exposing the hypocrisy and double standard inherent in judicial decision-making, realists claim that logic and rule-following are only a façade that obscures the underlying societal interests. This notion was subsequently expanded upon by the critical legal theory movement, which emphasized the political relevance of the law as a tool for empowerment and liberation.[166] Rather than being a mechanical and ostensibly unbiased application of rules, law lacks a 'correct answer' and instead conforms to opposing normative views.[167]

Even before the legal realism movement gained prominence, Justice Oliver Wendell Holmes expressed similar sentiments about decision-making. In 1897, he published his key book, The Path of Law, in which he challenged what he termed the 'logic fallacy':

---

[162] For an overview *see* e.g. Laura Kalman, *Legal Realism at Yale: 1927–1960* (University of North Carolina Press, 1986); Wilfrid E. Rumble, Jr., *American Legal Realism: Skepticism, Reform and the Judicial Process* (Cornell University Press 1968). *See also* more recently Pierre Brunet, *Analyse Réaliste du Jugement Juridique*, 147:4 Cahiers Philosophiques 9 (2016); Brian Leiter, *Naturalizing Jurisprudence. Essays on American Legal Realism and Naturalism in Legal Philosophy* (Oxford University Press 2007).

[163] *See* e.g. Karl N. Llewellyn, *Some Realism About Realism: Responding to Dean Pound*, 44(8) Harvard L. Rev. 1222 (1931). *See also* the later study, Wilfrid E. Rumble, Jr., *Rule-Skepticism and the Role of the Judge: A Study of American Legal Realism*, 15 Emory L.J. 251 (1966).

[164] *See* e.g. Jerome Frank, *Lawand the Modern Mind* (Brentano's 1930); Jerome Frank, *What Courts Do in Fact*, 26 Ill. L. Rev. 645, 645–66, 761–84 (1932).

[165] Joseph C. Hutcheson, Jr., *The Judgment Intuitive: The Function of the 'Hunch' in Judicial Decision*, 14 Cornell L. Rev. 274 (1929).

[166] *See* e.g. feminist critiques of adjudication, such as by Carol Gillian (e.g. *In a Different Voice* (Harvard University Press 1982)) and Catharine A. MacKinnon (e.g. *Feminism Unmodified: Discourses on Life and Law*(Harvard University Press 1987); *Toward a Feminist Theory of the State* (Harvard University Press 1989)).

[167] *See* e.g. Roberto Mangabeira Unger, *The Critical Legal Studies Movement* (Harvard University Press 1983). Compare Antonin Scalia, *The Rule of Lawas a Lawof Rules*, 56 U. Chi. L. Rev. 1175 (1989) (arguing to reduce the discretion given to courts).

"certainty generally is illusion, and repose is not the destiny of man. Behind the logical form lies a judgment as to the relative worth and importance of competing legislative grounds, often an inarticulate and unconscious judgment, it is true, and yet the very root and nerve of the whole proceeding. You can give any conclusion a logical form."[168]

He emphasized the relevance of statistics for the future of law, insisting that it was imminently an issue of prediction. He defined his work as a study of prediction, more particularly 'the prediction of the occurrence of public force through the courts'. [169] He maintained that a'so-called legal responsibility is nothing more than a forecast that if a man [or woman] performs or omits certain things, he [or she] would suffer in this or that manner as a result of a court ruling; and therefore of a legal right.'[170] To make accurate predictions, he speculated on the use of statistics for future generations of lawyers, noting that '[f]or the rational study of law, the black-letter man [or woman] of the present may be the man [or woman] of the future, but the man [or woman] of the future is [one] of statistics and a master of economics', adding that the number of our predictions, when generalized and reduced to a system, is not unmanageably huge.[171]

In light of the implications of AI, Holmes's focus on prediction and statistics in judicial decision-making in 1897, rather than logic, casts a new light today. As said before, predictions based on statistics or probabilities are exactly the traits that AI machine learning models make use of. [172] Furthermore, the predictive AI research described above indicate the relevance of extraneous non-legal elements, as suggested by legal realists.[173] According to the ECtHR research, the portion of judgements with the best predictive value is not the legal component, but the factual background section.[174] Additionally, the US Supreme Court research includes extra-legal aspects such as the justices' political views in the computer model.[175]

---

[168] Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 Harv. L. Rev. 457, 466 (1897).
[169] *Ibid.* at 457.
[170] *Ibid.* at 458.
[171] *Ibid.* at 458, 469.
[172] *Ibid.*
[173] *See* Section 3.
[174] Aletras et al., *supra* n. 45, at 10.
[175] Katz, Bommarito & Blackman, *supra* n. 69, at 4–6.

So, are we to infer, as some have claimed,[176] that AI will justify legal realist theories? And that the potential use of machine learning models to judicial decision-making would be consistent with what human judges have historically done? Thus, will the dispute between formalists and realists ultimately be won by the latter? These results, however, overlook a critical point: the centrality of probability as a normative underpinning for machine learning in artificial intelligence. As explored in further detail in the next section, this extends well beyond legal realism ideas.

6.3 The application of probabilistic inferences: A Path to Legal Determinism?

When considering legal theories relating to judicial decision-making, it is critical to distinguish between their descriptive aspect (i.e., how judges reason and make judgments successfully) and their prescriptive or normative element (i.e. how they should reason and make decisions).[177]

Legal formalism is composed of both descriptive and normative components. As a matter of logic, deduction, and legal syllogism, formalists define the method by which judges apply the law.[178] Additionally, they contend that the law's self-contained character, the neutrality of legal reasoning unaffected by extraneous non-legal variables, is how it should be normatively. This is to maintain the law's independence from politics and morality[179] and to establish a 'modern paradigm of legal rationality under the "rule of law"'.[180]

On the contrary, legal realism is preoccupied on descriptive characteristics. Holmes, Frank, and others illuminate the realities of judicial decision-making — therefore the movement's name. They emphasize the importance of non-legal issues and criticize the formalistic, automated, mathematical rule-application method as utopian and disconnected from reality. They do not, however, argue that courts should consider extraneous, non-legal

[176] Aletras et al., *supra* n. 45, at 16.
[177] *See* e.g. H. L. A. Hart, *Essays in Jurisprudence and Philosophy* 103–05 (Oxford Clarendon 1983). *See also* e.g. Pierre Brunet, *Le Raisonnement Juridique: Une Pratique Spécifique?* 26(4) Int'l J. Semiotics L. 767 (2013).
[178] *See* Section 6.1.
[179] Kelsen, *supra* n. 153, at 478
[180] MacCormick, *supra* n. 156, at ix–x.

matters. To take the Israeli parole research as an instance, although it is true that judges are affected by external variables such as lunch breaks, no one really argues that this is a positive thing that should serve as the normative foundation for judicial conduct.

However, normative features are not alien to other theories, such as the critical legal theory movement. Unger and others have emphasized the political importance of legislation and the underlying social objectives. By emphasizing the normative dimensions, legislation is seen as a way of achieving successful radical social reform.[181]

When examining artificial intelligence models, the preceding results in a number of findings. AI models would not only make decisions based on probabilities, but would also serve as their normative foundation. As previously stated, a conclusion based on machine learning AI models would not constitute predetermined legal norms, would not be the outcome of deductive reasoning, and would not follow the legal syllogism indicated above. This is true both descriptively (i.e. how successfully these models determine) and, more significantly, normatively (i.e. how these models should decide). Thus, substituting probabilistic inferences for logical, deductive, and rule-based reasoning as the normative foundation for judicial decision-making would not only violate legal formalism, but would also go way beyond legal realists' ideas.

Indeed, realists acknowledge that judges, after deliberating on a range of variables, including non-legal, political, and moral concerns, encapsulate their judgment in a manner that adheres to logic, using rule-based deductive reasoning.[182] What realists condemn is the hypocrisy of such a mask, yet they acknowledge its existence. AI-assisted decision-making would eliminate this format. Artificial intelligence judgments would not be made using logical or causal reasoning based on legal norms. The difficulties associated with this deficiency of thinking have previously been discussed in Section 5 above.

[181] Unger, *supra* n. 167.
[182] Holmes *supra* n. 168, at 465–66.

However, more fundamentally, the lack of a logical framework in judicial decision-making has consequences that extend beyond the descriptive or normative dimensions described. Hart distinguished three levels of judicial reasoning: (1) the processes or habits of thought that lead judges to their decisions (descriptive psychology); (2) recommendations for the processes to be followed (prescriptive judicial technology); and (3) the standards by which judicial decisions are to be evaluated.[183] It is at the third level that the lack of logic generates worry, at the very least, since it contradicts the decision's judgment or rationale. Alternatively, as Hart puts it:

The problem is not one of how judges get at their choices, or should arrive at them; rather, it is one of the principles they adhere to in defending their decisions, regardless of how they arrived at them. Whether conclusions are made by calculation or intuitive leap, the presence or lack of logic in their assessment may be a reality.[184]

Additionally, to the degree that legal theories highlight the political relevance of legislation and the fact that decision-makers have discretion to 'fill in' broad norms such as 'good faith,' the issue of how these political or moral factors would be addressed in an AI model emerges. Who or what could have the ability to sway such political or moral considerations? One may point towards the programmer in a classic computer paradigm. However, as mentioned in Section 2, sophisticated AI models do not need a programmer to define the method; rather, the algorithm is derived from the observable data. As a result, even on ethically or politically contentious subjects, the sole foundation for decision-making will be historical evidence. As previously stated, AI models are thus likely to adopt a conservative stance, even in a machine learning scenario characterized by ever-improving algorithms.

Using statistics or probability as the normative basis for judicial decision-making seems to provide additional difficulties. Probabilities or statistics have not been recognised as legal grounds for decision-making in the past.[185] English and other common law attorneys are likely to be acquainted

[183] Hart, *supra* n. 177, at 105.
[184] *Ibid.* at 105. *See also* Richard A. Wasserstrom, *The Judicial Decision* (Stanford University Press 1961).
[185] *See* e.g. the discussion in the US Supreme Court case of *McCleskey v. Kemp*, 481 U.S. 279, 287 et seq. (1987).

with the phrase 'balance of probability,' which denotes a proof requirement.[186] It is critical to note, however, that this only pertains to the establishing of facts. For example, in *Miller v. Minister of Pensions*, the UK Supreme Court (then the House of Lords) elaborated on the balance of probabilities concept, stating that if 'the evidence is such that the tribunal can state "We believe it is more probable than not," the burden is discharged, but not if the probabilities are equal.[187] Once the facts have been proven in this manner, probability has no place in judicial decision-making. For example, a claim cannot be granted only on the premise that there is an 80% possibility that the given circumstances constitute a breach of contract.

The above scenario exemplifies the practical challenges associated with probabilistic decision-making frameworks. What is the proper level at which a claim is considered granted? Would anything greater than 50% suffice? Or would a larger threshold of, say, 80% be required? Even with a higher threshold, one deliberately accepts a 20% chance that the case will be ruled incorrectly.

In this context, it's also worth recalling the sensitivity to data diets and the related bias problems outlined before. Consider, for example, a scenario in which State X has been regularly found in breach of a substantive investment protection mechanism included in investment treaties. Is this a factor in State X's chances of losing a future investment claim made by another investor?

To summarize, employing probabilistic analysis as the normative foundation for decision-making not only represents a significant theoretical paradigm change, but also creates significant practical concerns. This novel technique may be referred to as legal determinism, since it bases future outcomes on probabilistic calculations based on historical facts. As shown in this essay, it has a variety of ramifications for judicial decision-making that must be properly evaluated.

[186] *See* e.g. Emily Sherwin, *A Comparative Viewof Standards of Proof*, 50 Am. J. Comp. L. 243 (2002).
[187] House of Lords, [1947] 2 All E.R. 372 (opinion delivered by Lord Denning).

7. Conclusion

The purpose of this article is to discuss the application of artificial intelligence in arbitral and judicial decision-making. After examining the technical components of AI and its implications and limitations, as well as the more basic influence they may have on human decision-making processes and theories, the following are the study's major findings and conclusions:

First, existing research on decision outcome prediction, despite achieving astonishing accuracy rates of 70–80%, have significant limits. An examination of the methods and assumptions used casts doubt on the notion that these models may be used to make ex ante result predictions. Among other reasons, it is debatable whether the models can be used equally well and provide effective outcomes in circumstances when the court resolves a disagreement directly rather than reviewing a lower court's judgment.

Second, the technological characteristics of artificial intelligence imply specific prerequisites for its application in judicial decision-making, at least for the time being. This includes the necessity for adequate non-confidential case data and, presumably, repeating fact patterns and binary outcomes. Because AI models are often constructed using knowledge collected from prior input data – even in ever-improving machine learning algorithms – they are likely to take 'conservative' methods and may be unable to adapt to significant policy changes over time. Additionally, a blind subservient attitude toward the impartiality and infallibility of algorithms is misguided. Any data-driven computer model is only as good as its input data, and so runs the danger of perpetuating pre-existing biases.

Third, the need for rational decision-making is expected to be a significant impediment to AI-based legal decision-making. In the case of black-boxed models, it may be impossible to pinpoint the components that contributed to a certain result prediction, at least at the present technology level. Furthermore, even if specific components are identified as causes of a certain result prediction, they may not provide an adequate explanation for human addressees in a particular situation.

Fourth, the application of artificial intelligence does not simply fit into legal conceptions of judicial decision-making. Artificial intelligence models elevate probabilistic findings to serve as the normative foundation for legal decision-making. This not only represents a paradigm change theoretically, but also raises significant considerations about whether and how future judgments should be made on probabilistic estimates based on historical evidence.

However, none of these findings should obscure the obvious: AI will radically alter the legal profession and legal operations, including judicial decision-making. It is critical to continue researching the best ways to employ AI, despite the limits, restrictions, and concerns outlined in this article. International arbitration, which is always criticized for being overly costly and time-consuming, must take the assertion made by certain AI developers that computers can accomplish the job of 360,000 attorneys seriously. Further study is required to determine the optimal technique to mix human decision-makers with AI to get the most efficient outcomes.