PANEL 2

Reforming International Arbitration Through the Use of Artificial Intelligence: What to Expect in the Near Future?

READINGS:

1. **Thomas R. Snider, Artificial Intelligence and International Arbitration: Going Beyond E-mail**
2. **Bianca Berardicurti, '25. Artificial Intelligence in International Arbitration: The World is All That is The Case', in Carlos González-Bueno (ed), 40 under 40 International Arbitration (2021), (Dykinson, S.L. 2021) pp. 377 – 392**
3. **Paul Bennett Marrow, Mansi Karolt, et al., 'Artificial Intelligence and Arbitration: The Computer as an Arbitrator—Are We There Yet?', inGregory Kochansky (ed), Dispute Resolution Journal, (© Kluwer Law International; AAA-ICDR 2019, Volume 74 Issue 4) pp. 35 – 76**
4. **Maxi Scherer, 'Artificial Intelligence and Legal Decision-Making: The Wide Open?', in Maxi Scherer (ed), Journal of International Arbitration, (© Kluwer Law International; Kluwer Law International 2019, Volume 36 Issue 5) pp. 539 - 574**

Artificial Intelligence and International Arbitration: Going Beyond E-mail

Thomas R. Snider - Partner, Head of Arbitration - Arbitration / Construction and Infrastructure / Mediation

t.snider@tamimi.com - Dubai International Financial Centre

Sergejs Dilevka - Senior Associate - Arbitration

- Dubai International Financial Centre

Camelia Aknouche

c.aknouche@tamimi.com - DIFC, UAE

The last two decades have witnessed phenomenal advancements in information technology (IT), which have fostered a remarkable level of innovation in products and services offered across numerous industries. Yet, international arbitration, forming part of the legal services industry, has been so far practically unaffected by such developments.

To be sure, one can find examples where the international arbitration community has been improving its services through implementation of new IT: video conferencing, e-disclosure, use of online platforms and cloud-based technologies. However, these and other similar steps are improvements of an incremental nature, well known to most and successfully practised by many arbitration practitioners and institutions. Accordingly, this article will look forward and concentrate on a ground-breaking technology that within the next 100 years will, apart from affecting all aspects of our lives in ways that we cannot yet imagine, create genuine innovation in international arbitration – Artificial Intelligence (AI).

The Advent of AI

In 1955, John McCarthy suggested the term 'artificial intelligence' in a research project proposal, which described AI as a problem '*of making a machine behave in ways that would be called intelligent if a human were so behaving*'.

Sixty years later, after several hype cycles inevitably followed by 'AI winters', the general consensus is that scientists are finally creating systems sophisticated enough to fall within the meaning of AI. IBM's Watson, '*a technology platform that uses natural language processing and machine learning to reveal insights from large amounts of unstructured data*', is the most celebrated example of AI machinery. Watson may comprise multiple enabling technologies that may be added and configured depending on the required functionality. Two technologies are of fundamental importance to its 'intelligence': (i) natural language processing (NLP) and (ii) machine learning (ML). NLP may be described as a '*problem of understanding strings of characters that form words, sentences, or larger assemblages of text in "natural" language, such as English'.* ML may be defined as *'a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or perform other kinds of decision making under uncertainty (such as planning how to collect more data!)*'.

Since its triumph at Jeopardy in 2011, Watson has unequivocally established that AI machines can 'understand' text, including grammar and context, and 'learn' through making decisions in response to altering circumstances.

An international arbitration specialist may now begin to contemplate the potential use of such technologies in her/his practice.

Legal Research and Document Review

The practice of international arbitration often entails having a grasp of international law and several domestic legal systems at the same time. Moreover, parties submit to tribunals voluminous hard copy and electronic documents. Accordingly, international arbitration is a document intensive field of law that requires counsel and arbitrators to spend countless hours on legal research and document review. Due to an ever-growing demand for speed and efficiency, the present state of affairs cannot last.

A significant amount of legal research and document review has now shifted from libraries and client basement archives onto online platforms. However, in the hunt for exhaustive research/review, counsel and arbitrators still read through innumerable pages, frequently containing irrelevant text. Application of search terms to text is mostly of assistance but is regularly impeded by false positive results and, in any case, requires constant human supervision.

Use of AI for legal research and document review in the foreseeable future will cut the time necessary for such exercises from hours/days/months/years to seconds (in some instances to milliseconds).

AI: Trusted Assistant to International Arbitration

Speech Recognition

Arguably one of the most important enabling technologies for AI, after NLP and ML, is speech recognition (SP) technology. The technology has made great improvements in quality; now SP may not only recognise different accents and languages with very impressive accuracy but it can also identify the voices of particular individuals. Let us consider just a few potential uses for an SP enabled AI platform:

Transcripts: Generally, the parties in international arbitration prefer to use the services of transcription service providers at hearings. Instructing such specialists, of course, is an additional expense for the disputing parties that involves various logistical arrangements and complications. AI may render the use of court reporters obsolete as the AI platform would be able to record the hearing via microphones and provide a real-time transcript with speaker identification for all concerned.

Interpretation: Parties in international arbitration often need to present witnesses that may require the assistance of interpreters. This also involves time and costs that may be cut by using AI for interpretation purposes at the hearing.

Translation: A document-heavy international arbitration may force parties to translate evidence into the language of arbitration, thus incurring substantial costs and extending the arbitration process. AI will, of course, be capable of translating thousands of documents in seconds with very high accuracy, including scanned, hand-written or annotated documents.

Drafting of Awards

Arbitrators spend a lot of time on drafting standard sections of their arbitration awards, e.g., the parties, the procedural history, the arbitration clause, the governing law, the parties’ positions, and the arbitration costs. Arbitrators may save time and parties’ fees by delegating the drafting of such ‘boilerplate’ sections to AI machines.

Appointing Authority

When the parties fail to appoint arbitrators or when arbitrators fail to agree on a chair, generally a default appointing authority will come into play. AI may assist with such appointments by providing its list of potentially suitable candidates based on multiple variables, for example, knowledge and experience in particular areas of law, languages, number of pending and concluded arbitrations, level of party satisfaction in previous cases, time taken to render a final award (on average), and, importantly, potential conflict of interest that AI may identify by scanning through databases and the Internet.

AI – Expert / Arbitrator

It seems possible to suggest that AI designed for international arbitration will continue to rapidly improve, reaching a stage when it would be conceivable to request AI for an expert opinion and even to render an award. An easily identifiable area of discomfort and an obstacle to allowing AI to perform such functions is the lack of understanding of how the decision was reached. However, Watson's developers claim that when answering questions, Watson develops hypotheses and makes evidence based decisions (taking into account a degree of confidence in percentage terms based on the preponderance of evidence). Therefore, AI is capable of reasoning and, over time, AI will begin producing lines of reasoning logical to a person.

As mentioned previously, the costs and time involved in creating an AI-generated expert opinion or an arbitral award will be cut to an absolute minimum: a development which, without a doubt, will be welcomed by the international arbitration community.

Legal Framework

In order for AI to be successfully integrated into the system of international arbitration in the future, its definition should be crystallised and its use should be regulated. Perhaps, some would suggest creating a custom (state-of-the-art) legal framework for dispute settlement by AI. Yet, another route may be to amend the existing arbitration rules, domestic legislation, and international agreements. The suggestions below are of a general and non-exhaustive nature.

Arbitration Rules

Let us use the DIFC-LCIA Arbitration Rules (effective 1 October 2016) as an example to suggest amendments that would allow the DIFC-LCIA to offer AI to the disputing parties.

Article 5.2 of the DIFC-LCIA Rules may be amended by introducing a new defined term in the following manner:

The expression the "Arbitral Tribunal" includes a sole arbitrator or all the arbitrators where more than one. [An arbitrator includes Artificial Intelligence Software].

Additionally, drawing inspiration from Article 9B ('Emergency Arbitrator') of the DIFC-LCIA Rules, a new AI-specific article could be introduced into the DIFC-LCIA Rules containing a definition and provisions pertinent to the use of AI, including which articles of the DIFC-LCIA Rules do or do not apply to AI.

Consequently, and unless otherwise agreed by the parties, AI may become a default sole arbitrator in certain disputes, for example, disputes under a certain amount or of a certain complexity/sector (e.g., construction).

Domestic Legislation

Changes in domestic arbitration laws would be strongly recommended to provide certainty to the international arbitration community (arbitral institutions, counsel, and parties) that the use of such technology for settlement of disputes by arbitration is legal. Taking the DIFC-LCIA Law No. 1 of 2008 (DIFC-LCIA Arbitration Law) as an example of domestic legislation that may be amended to introduce and reflect the practice of dispute settlement by AI, one suggestion would be to amend Article 16 of the DIFC-LCIA Arbitration Law as follows:

The parties are free to determine the number of arbitrators provided that it is an odd number. *[An arbitrator includes Artificial Intelligence Software]*.

Alternatively, and possibly more appropriate, is to amend the Schedule – Interpretation, in its relevant part, as follows:

D. Defined Terms

Arbitral Tribunal | a sole arbitrator or a panel of arbitrators [and Artificial Intelligence Software]

However, it is not yet clear how AI would co-operate in 'mixed' arbitral tribunals (consisting of AI and human arbitrators). Therefore, it is possible that a separate section dealing exclusively with AI as a sole arbitrator may be required.

International Agreements

For AI to become commonly used by the arbitral community, it is essential, for reasons of legal certainty, that major international agreements concerning international arbitration recognise AI as equivalent to arbitrators or arbitral tribunals. Of course, the Convention on the Recognition and Enforcement of Foreign Arbitral Awards 1958 (the NY Convention), to which the United Arab Emirates and other GCC states are parties, is one of the most important international agreements of such a kind.

One might be courageous enough to advocate, for example, an amendment to Article I (2) of the NY Convention in the following way:

The term "arbitral awards" shall include not only awards made by arbitrators appointed for each case but also those made by permanent arbitral bodies to which the parties have submitted. *[For the avoidance of doubt, the term "arbitrator" shall include Artificial Intelligence Software.]*

However, it does not seem realistic, considering the amount of time that was required for 156 States to become parties to the NY Convention, to obtain signature to such an amendment in the near future.

Conclusion

Regardless how the above may sound, large international law firms already employ 'data scientists', 'legal solutions architects' and 'heads of strategic client technology', who focus on IT and AI solutions that would assist human counsel. New companies are being incorporated with a particular focus on AI solutions for legal research. One law firm went far enough to employ two computers as partners.

As is the case with any new technology, AI will require a significant influx of capital at the beginning. It will also require constant development and improvement. In this regard, software developers will have to work closely with arbitration practitioners to identify problems that may occur and to streamline processes. With time, AI will reach a stage when its use in arbitration will be universal and its cost will be no higher than one of the average office computers of today.

AI will become an assistant to arbitrators and, in some cases, even an arbitrator possessing vital qualities for human arbitrators as being relentless, consistent, systematic, impartial; and it will continue to improve and grow to be powerful. Yet, there exists some scepticism towards an idea of assisting an arbitrator with AI, and even more – replacing a human arbitrator with AI. To shift thinking in line with technological development, one should consider this: airplanes are being landed using autopilot; cars are being driven autonomously. If humans entrust their lives to machines/computers, why should not AI take care of, perhaps, less important matters like settling arbitration disputes?

Al Tamimi & Company's Arbitration team regularly advises on international investment and commercial arbitration. For further information, please contact Thomas Snider (t.snider@tamimi.com) or Sergejs Dilevka (s.dilevka@tamimi.com).

Document information

Publication

- 40 under 40 International Arbitration (2021)

Bibliographic reference

Bianca Berardicurti, '25. Artificial Intelligence in International Arbitration: The World is All That is The Case', in Carlos González-Bueno (ed), 40 under 40 International Arbitration (2021), (Dykinson, S.L. 2021) pp. 377 - 392

25. Artificial Intelligence in International Arbitration: The World is All That is The Case

Bianca Berardicurti

(1)

(1)

> '*A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
>
> *A robot must obey the orders given by human beings except when, such orders would conflict with the First Law.*
>
> *A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws*'.

Isaac Asimov, Runaround (1942)

The scope of this article is to investigate how artificial intelligence ('**AI**') is being used in the field of international arbitration.

In more detail, I will endeavour to navigate the legal, ethical, and philosophical problems that the use of AI tools is either posing, or likely to pose, in terms of the integrity and reliability of the arbitration system.

After a brief introduction, I will firstly address the issue of the definition of Artificial Intelligence. I will then give an overview of AI tools that are being used in the International Arbitration field. Thereafter, I will examine the main ethical and legal problems raised by the use of AI tools, by focusing on different phases of the arbitral proceedings. Lastly, I will present the conclusions.

1 Introduction

Over the course of recent decades, technological developments, including the impressive improvement of AI, have triggered a revolution comparable to that
P 378
of the Industrial Revolution, and which is destined to have a disruptive impact over our lives.

At the very core of such a revolution lies a profound change in the paradigm of language. Indeed, mathematical writing is now used alongside the roughly 53 centuries old writing which humanity invented through the Greeks' conversion of the Phoenician consonant alphabet into a vowel-consonantal system [(3)]. Aside our alphabetical language, which humanity has been using to interpret and describe the world thus far, now stands computational language, which is transformed from a non-verbal source and through a combinatory function is recomposed into a new form. As such, it is a dematerialised language.

Not only can numerical writing transmit messages rapidly, through vast diffusion and beyond territorial boundaries: this new form of writing represents a symbolic revolution. This means that it has changed the way humans form and build on the perception of objects and the perception of moral values: after all, symbols lie at the root of intelligent actions. [(4)]

The technological revolution that the world is now experiencing is already having a huge social, cultural and even political impact, there for all to see in our daily life. Currently, new technologies are also proving pivotal in handling the Covid-19 pandemic outbreak, and in allowing people to adjust their lives to the new normal.

As far as the legal field is concerned, artificial intelligence tools are progressively taking hold and AI is already touching many areas of the law. Indeed, AI is already significantly affecting the manner in which legal business is conducted (including block chain and other technologies), transactions are entered into (including smart contracts) and disputes are raised and resolved. [(5)]

International arbitration makes no exception in this respect, although lawyers seems to be somehow reluctant to acknowledge the fact —I myself was strongly biased when I initially approached this subject. Yet, AI tools are already commonplace throughout most of the arbitral proceeding.

P 379

The debate on the entry of artificial intelligence into the field of arbitration has, in recent years, been very lively. Although the discussions have been largely focusing on (a) the pros and cons of the use of artificial intelligence in international arbitration, mainly in terms of time, efficiency and costs, and (b) the lawyers' concerns that artificial intelligence tools tailored to work in the field of law may eventually turn them into Silicon Valley's next victims, (6) yet the international arbitration community seems to be pretty much aware of all the challenges that the use of machines is already raising and is likely to change in the future, also in terms of ethical and legal problems. (7)

Indeed, it is precisely our responsibility, as international arbitration practitioners, (as it is of all humankind) to ensure in-depth discussion on such significant issues in order to best prepare for their ultimate arrival: the future is just around the corner.

2 What do we Mean by Artificial Intelligence?

Defining AI intelligence is no easy task.

The Oxford Dictionary, used as a starting point by prominent authors with extensive dealings on the subject, (8) defines artificial intelligence as the '[t]heory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making and translation between languages'. (9)

In order to simplify, beyond the strictly technical definition of AI, which falls outside the scope of this article, AI may be broadly defined as the general process whereby large amounts of data (the so called 'big data') are combined with a powerful iterative data processing system and intelligent
P 380
algorithms, thereby enabling the software to learn automatically from patterns in the data. (10)

Some distinctions are generally used in the AI field which might help to navigate this vast new world, which lawyers are mostly unfamiliar with.

One such distinction is between 'Strong AI' and 'Weak AI'.

Whilst a Weak AI System basically mimics human reasoning without actually having it, a Strong AI system is able to think or reason independently, without using pre-programmed ways of human thinking or reasoning. (11) In other words: Strong AI assumes that machines do or ultimately will have minds, while Weak AI asserts that they simulate real intelligence: the question seems thus to be whether machines can be truly intelligent, or simply act as if they were intelligent. (12) After all, the very person who has coined the term 'artificial intelligence' (13) defined it as the process of 'Making a machine behave in ways that would be called intelligent if humans were so behaving'. (14)

A further relevant distinction to be taken into account is between these two types of AI: rule-based learning and machine learning —the latter being a mechanism which is able to identify patterns and vary algorithms on the basis of already existing data and user feedback. Deep learning models are a specific subset of machine learning: these are modelled on the structure of a human brain and are able to learn themselves without human intervention from massive volumes of data.

It might be interesting to note however, that remarkably, the reference point for defining artificial intelligence still is human intelligence —which makes the question even more third-rate, considering how difficult defining human intelligence may also be.

P 381

3 The Array of Artificial Intelligence Tools in International Arbitration

Many AI tools are being used already in the field of international arbitration, and the trend reflects that users are increasingly optimistic as to the introduction of AI applications.

Indeed, the survey conducted in 2018 by the Queen Mary University shows, (15) *inter alia*, that 78% of respondents indicated that 'AI' is a form of IT worth using more.

It is beyond the scope of this article to enter into detail on all the AI applications that are used or may be used in the context of arbitration.

A useful and very clear classification has been made by some authors (16) who divided AI tools used in arbitration into four categories, based on their functional complexity. (17)

More specifically, a first category of AI tools can be used to carry out legal research more quickly and with more precise or focused results. A second group of AI tools may be used for the selection of suitable professionals, such as counsels, experts and arbitrators. A third group of AI tools may be used to facilitate certain procedural phases. By way of example, voice recognition devices may at some point substitute transcripts, AI tools may be used for evidentiary searches, for summarising pieces of evidence. Also, some of the compilatory parts of the awards may be drafted with the aid of AI devices. Finally, a fourth category may be used and qualified as tools of predictive justice. AI systems used for predicting the outcome of a dispute or even applied to the decision-making process fall under this fourth category.

Most of the AI tools described above proved very useful in terms of reducing costs and timing of arbitration and in supporting lawyers in those activities that are generally highly time consuming and expensive for the clients, such as document review and document production. (18)

P 382

However, it is clear to all that the use of AI application in international arbitration poses some questions both at a legal and ethical level. Obviously, the intensity of the issues possibly arising from the use of artificial intelligence vary, depending on the specific tool and the specific phase of the arbitral proceedings which is concretely concerned, or on the specific interests or rights at stake.

By way of example: tools aimed at assisting in the document production phase may pose an issue in relation to the access to justice, as those parties which have not sufficient resources to procure the facilities could be highly affected. Tools aimed at supporting the selection of arbitrators, form the one hand proved very useful in mapping the relationship between arbitrators and council in terms of conflict check; yet, these may lead to some manipulation strategies by the parties and raise some concerns in case the relevant tools are used with predictive purposes. Tools for the selection of witnesses then raise even more serious dilemmas from an ethical perspective to the extent that they might lend themselves to the manipulation of the evidence-taking phase. Finally, predictive tools suitable for use in the field of justice are by far the most problematic of the AI resources.

Some of those problems will be dealt with in the following paragraphs, in relation to three specific phases of the arbitral proceedings.

4 Predicting the Outcome of the Decision

Predicting the future and reducing uncertainties in advance has always been, and still is, an innate need for human beings. Different times, different methods: while in the ancient times haruspices' divination was common practice, in 2017 a Turkish entrepreneur created Falladin, a fortune telling app transporting the tradition of Turkish coffee grinds straight into the age of AI. (19)

As far as the field of arbitration is concerned, recent years have seen the launch of several tools for data analytics, aimed at predicting the outcome of disputes.

Nowadays, there are several such products on the market, although each of them seems to achieve results by different methods, including the ‘game theory’ application. (20) Such tools are likely to be increasingly used in the future by lawyers, as well by litigation funders, whose interest in the outcome of an arbitration is merely financial. (21)

P 383

Undoubtedly, predicting the outcome of a dispute through artificial intelligence mechanisms may bring with it many benefits. Just by way of example, when a lawyers' opinion is supported by the output of an AI machine, parties could be more inclined to settle a dispute, since they have a clearer idea of which way their arbitration could go.

However, the use (not to mention the delegation) of the predictive function to AI raises more than one question, and on many levels, either practical or political and ethical.

First, it should be considered that the reliability of any data-driven AI system lies in the so-called four Vs: Volume (scale of data), Variety (different forms of data), velocity (analysis of streaming data) and veracity (uncertainty of data). (22)

While the most important arbitral institutions have already taken many steps in the direction of making the award public, at least in part, arbitration —especially commercial— is still confidential. The scarcity of public data, which is typically inherent in arbitration, materially affects the first V: indeed, machine learning programs, which are based on probabilistic inferences, are data hungry. The less data available, the less accurate the prediction: for as much as international arbitral institutions may be tackling the transparency issue with obvious good will, the amount of case data generated from commercial arbitration is nevertheless completely inadequate as a tool for enabling AI to render an accurate prediction.

Here would appear to lie the precise difference between arbitration and some prediction experiments made in recent years.

In 2016, researchers at UCL, the University of Pennsylvania and the University of Sheffield, developed AI software which analysed the language used in submissions and previous judgments to predict the outcomes of the European Convention on Human Rights (ECHR). The machine was correct in 79% of cases. Likewise, a group of researchers worked on the prediction of US Court decisions, obtaining very accurate results.

Although the two experiments differed in several aspects, the enormous amount of data reviewed was the same for both models: the dataset for the ECHR project amounted to 584 decisions, while the US Supreme Court cases were more than 28,000.

P 384
It is plain to see that relying upon such a huge amount of data input is just not possible as far as international arbitration is concerned, at least for the time being.

Second, changes in law over time affect the Velocity of the incoming data to be processed: this raises the problem of how AI models which are, by definition, based past data, may deal with policy changes. Remarkably, this problem is inherent to all those systems which use the past to predict the future: after all, the creator of the Falladin App himself, stated that the tool is aimed at reading the future of a person 'by evaluating a person's past'.

However, luckily enough, people (and even arbitrators) might still be somewhat unpredictable.

From an ethical perspective, there are also issues in cataloguing adjudicators' beliefs, tendencies, and decisions. This is certainly more problematic where national court judges are concerned, as the cataloguing and prediction of the judges' decision could somehow clash with the fundamental principle of the *juge naturel* and the freedom of choice of the judiciary system. However, it cannot be denied that from an ethical perspective, the arbitrators' profiling tendencies may also cause issues in the decision making and give rise to abusive conduct from either or both parties.

Furthermore, using AI to predict the outcome of a dispute could raise some concerns over the appointment of arbitrators and the efforts made by the arbitration community to boost diversity and transparency: indeed, should the AI tools be able to predict the arbitrators' decisions, that would probably lead to the reinforcement of fixed patterns in the appointing of certain specific arbitrators in certain specific disputes. (23)

Finally, a material (and provocative) question may be posed with respect to the possibility to foresee the outcome of an arbitration: shouldn't risk be an inherent part of the dispute?

5 Making the Decision

Using AI tools for carrying out legal decision making might seem more distant than it actually is.

P 385
Indeed, artificial intelligence adjudicators are to some degree already being used where smart contracts (24) and blockchain are at stake. (25)

Also, some AI tools have been used in courts in assisting the adjudication phase already. By way of example, in Wisconsin v. Loomis the court relied on the decision supporting tool COMPAS to deny the indicted individual's request of parole.

The array of ethical dilemmas raised by the delegation of the decision-making process to a machine is so vast that it is almost impossible to address all of them.

At the very core of the topic lays a fundamental question: is it a basic right to have justice rendered by a human being? (26) To a certain degree, Constitutions and even arbitration laws basically assumes that there is an inherent value in being heard by a fellow human, who is subject to duties of fairness and respect. (27)

Indeed, although constitutions and legislations of most countries might not actually address the question, it is deemed reasonable to reply in the affirmative: humans should make justice, not machines, in accordance with the fundamental principles upon which democratic legal order was founded.

After all, most arbitration laws expressly provide that arbitrators must be persons having full capacity, (28) needless to say, machines do not fall under this definition.

Finally, and most importantly, as of today, machines are still unable to deliver the reasoning for their decisions, both in terms of causal chain, but also in terms of contextual explanation. Again, this is a fundamental difference with the ECHR experiment, the outcome of which went both ways: application/non-application of the sanctions. No reasoning for such outcome was provided by the system.

P 386
However, providing a thought through decision is one of the fundamental features of legal decision-making and a fundamental right of the advanced legal orders. (29)

Hence, the need for reasoned decisions is likely to be the most significant barrier for AI-decision making, as AI basically works on a probabilistic basis. (30)

On the other hand, it could be argued that the bright side of using machines for decision-making, is that machines should not be affected by human bias.

Indeed, considerable time has passed since psychologists discovered systematic patterns of deviation from rational judgment, which have been catalogued in a continuously evolving list of cognitive biases. [(31)] This seems to be inherent to the way the human brain actually works: a distinction has been made by psychologists and neuroscientists between two kinds of thinking, one that is intuitive and automatic (System 1), and another that is reflective and rational (System 2). [(32)]

Whilst implicit biases, such as cognitive, cultural, ethical and gender bias, may have a distorting effect in decision making made by human beings, machines should instead be immune from such deviations. Were this to be the case, it would bring to bear two main consequences.

If there is a cloud here, it clearly does have a silver lining: decisions would be free of irrational deviations forever. On the other hand, decisions could also be deprived of all those intuitions, which the human 'touch' generally provides: the taking into consideration of grey areas, fairness, the ability to understand whether a witness is actually telling the truth and to pinpoint contradictions, the appreciation of extra-legal factors and the application of general principles such as that of good faith, can be seen as inherent to human thinking rather than to machine processes.

Besides, can we be really sure that machine decisions are completely free from bias?
P 387
Data-based systems are good and reliable so long as the data they are fed are good and reliable. Hence, on closer inspection, should the input data be affected by human bias, not only machines would extract biased decisions as well, but these would also end up working as a bias multiplication, possibly perpetuating the systemic distortions.

This leads to a further negative aspect which could potentially affect the use of artificial intelligence for the decision making process in arbitration. Indeed, using the past to make the future (i.e., using the data related to past cases) would lead to conservative decisions, perpetuating trends and stifling the developmental process of change in human thinking and perception.

That would eventually restrain evolutionary jurisprudence, inevitably depriving justice of one of its most important social functions.

6 Challenging, Recognizing and Enforcing the Decision

The possible breach of fundamental rights or principles of public order as discussed above, could eventually raise difficulties during the phase of recognizing and enforcing a decision. On the other hand, depending on the specific seat where the award is made, such a violation could also provide grounds for challenging the decision.

So far as recognition/enforcement is concerned, the starting point is Article V(2) b of the Convention on the Recognition and Enforcement of Foreign Arbitral Awards (the '**NY Convention**'), pursuant to which 2. recognition and enforcement of an arbitral award may be refused if the competent authority in the country where recognition and enforcement is sought finds that the recognition or enforcement of the award would be contrary to the public policy of that country.

Hence, should the use of AI intelligence in one or more stages of the arbitration proceedings be considered as violating the public policy of the country where the award should be recognised/enforced, that would amount to solid grounds for refusal according to Article V(2) b of the NY Convention.

The questions to ask would therefore appear to be (a) what by public policy, provided that an internationally recognised notion actually exists, and (b) whether
P 388
and to what extent could the use of AI in the context of an arbitration proceedings theoretically breach public policy rules.

From an initial perspective, it is common knowledge that 'public policy' is a broad and variable concept, which changes considerably over time, also on the basis of the cultural, social and political context in which it resides. After all, the NY Convention itself does not define public policy, nor does it give any indication as to how to build the notion. Moreover, in practice, courts have varyingly used national, international and even transnational interpretations of the public policy exception. [(33)]

It is beyond the scope of this article to investigate whether a notion of public policy globally based, and on global values actually exists, and in the event that it did, of what would it comprise. [(34)]

However, whether assuming a transnational perspective of public policy or reasoning at a national level, it can be broadly said that public policy rules include those laws, the observation of which is necessary for the safeguard of political, social and economic organisation, in dealing with basic principles which are inherent to the legal system.

That being said, and although it has been observed that for the purposes of Article V(2)b of the NY Convention, the notion of public policy should be narrowly construed, (35) it cannot be excluded that the use of artificial intelligence in arbitral proceedings may somehow clash with certain public order principles.

Indeed, as discussed, depending on the specific tool used and the specific phase of the actual arbitration concerned, the principle of due process, for example, could be affected.

Such a set of circumstances extends even beyond the level of consent which parties might be able to provide regarding the use of certain AI tools: by way of example, while the use of AI systems would raise no significant issues in the document review phase, as long as the parties have given their consent, the lack
P 389
of a clear and logical reasoning of the award, could certainly raise due process violations and provide grounds for a refusal of enforcement.

At the same time, as most of the national, legal orders allow the challenge of the award in accordance with the violation of public policy rules, the use of AI could also represent grounds for setting aside the awards.

7 Conclusions

The considerations outlined above do not claim to be exhaustive, nor to provide definitive answers to a problem which is both delicate and still deep in the process evolution.

However, some conclusions can be clearly drawn.

The first, is that no absolutely reliable answers can be provided with respect to the numerous issues raised by the use of artificial intelligence in the international arbitration sector. It is important, therefore, that the subject be approached without any ideological bias or prejudice.

Carlo Rovelli, an Italian physician, put it brilliantly: ‘Our prejudices about reality are the result of our experience, and our experience is limited. We cannot take the generalisations that we have made in the past as gospel. Nobody said it better than Douglas Adams, with its ironic tone: *There are some oddities in the perspective whit which we see the world. The fact that we live at the bottom of a deep gravity well, on the surface of a gas-covered planet going around a fireball 90 million miles away, and think this to be normal, is obviously some indication of how skewed our perspective tends to be, but we have done various things over intellectual history to slowly correct some of ours misapprehensions*. Let's expect to have to change our metaphysical-provincial outlook. It's time we take the new concepts we learn about the world seriously, even if they clash with our prejudices about how things really are’. (36)

The second, is that the use of artificial intelligence can be defined as a true technological evolution, which can prove extremely effective in terms of time and cost savings in the course of arbitration, but that its application pose serious ethical and legal problems, which can interfere with the integrity of the arbitration system and which must therefore be used with caution.

P 390
To this extent, guidance can be sought from clear instruction provided by the Ethical Charter for the use of AI in judicial systems and their environment, and as adopted by the European Commission for the Efficiency of Justice (CEPEJ) on 3-4 December 2018, which set out some very sharp principles that must be met in using artificial intelligence in the legal field, namely:

a)Principle of respect for fundamental rights (ensure that the use of AI tools does not conflict with fundamental rights);

b)Principle of non-discrimination (by way of example in terms of access to justice);

c)Principle of quality and security (in terms of certified sources, intangible data and secure technological environment);

d)Principle of transparency, impartiality and fairness;

e)Principle ‘under user control’ (ensure that users are duly informed and in control of the choices made).

The third, is that human interaction is still, to this day, fundamental to the appropriate, wise and well-considered use of artificial intelligence in the field of arbitration.

As mentioned above, the first use of the term ‘artificial intelligence’ is to be attributed to John McCarty who defined artificial intelligence as the process of ‘Making a machine behave in ways that would be called intelligent if humans were so behaving’. (37)

However, humans and machines do not behave the same way. Just by way of example, and to put it with the words of John Searle, [38] computers themselves cannot think. Indeed, 'thinking' in its broader and most noble sense, is not a mere interconnection among neurons: rather it includes consciousness, the feeling of experiencing things (the so-called 'qualia', i.e., basically the subjective and conscious experience), sentience, discernment, judgment, empathy, intuition.

As a result, while machines are indeed able to manipulate symbols (sometimes even better than humans do) human beings are still the ones interpreting said activity at the end of the day.

P 391

After all, the basis of law is essentially social and political, and justice is also based on equity and fairness. [39] This is why the conduct of arbitration should still be handled by humans, although with the support of Artificial Intelligence in case needed for boosting efficiency.

There is absolutely no doubt, at least for the time being, that the contribution of a human being in the use and interpretation of machine-driven output is still necessary to the safeguarding of fairness and dependability in the justice system.

Hence, after all, machines are going to steal our jobs just not yet.

P 391

References

1)

Bianca Berardicurti: [2] Managing Associate Legance Rome

2)

I wish to thank Cecilia Carrara, for her invaluable help, and M.d.M. for giving me, as always, a different perspective from which moving forward.

1)

L. WITTENGSTEIN, 'Tractatus Logico-Philosophicus', 1922.

3)

The origins of language though are still unclear and much debated. Greek language has been taken as a reference point as it is the first language which was somehow potentially mechanisable.

4)

A. NEWELL & H. SIMON, 'Award Acceptance Lecture of the 1975 Turing Award'.

5)

K. PAISLEY, E. SUSSMAN 'Artificial Intelligence Challenges and Opportunities in International Arbitration', NYSBA, New York Dispute Resolution Lawyer, 2018, Vol. 11, No. 1, p. 35.

6)

S. LOHR, 'A.I. Is Doing Legal Work. But It Won't Replace Lawyers. Yet. The New York Times', March 19, 2017

7)

C. CARRARA, 'Chapter IV: Science and Arbitration, The Impact of Cognitive Science and Artificial Intelligence on Arbitral Proceedings Ethical Issues', Austrian Yearbook on International Arbitration 520 (K. Christian *et al*. (eds.), 2020).

8)

M. SCHERER, 'Artificial Intelligence and Legal Decision Making: The Wide Open?', Journal of International Arbitration 36 no. 5, 2019, p. 539.
M. SCHERER, 'International Arbitration 3.0-How Artificial Intelligence Will Change Dispute Resolution', Austrian Yearbook on International Arbitration (2019), p. 503.

9)

(Available at <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960>).

10)

K. PAISLEY, E. SUSSMAN, 'Artificial Intelligence Challenges and Opportunities for International Arbitration', NYSBA New York Dispute Resolution Lawyer, Spring 2018, vol. 11, no. 1, p. 35.

11)

M. SCHERER, see supra note 8 §2, p. 505.

12)

J. KAPLAN, 'Artificial Intelligence, What Everyone Needs To Know', Oxford University Press (2016), p. 68

13)

The term Artificial Intelligence is to be attributed to John McCarty, the assistant professor in mathematics at the Dartmouth College in Hanover, who, along with three other researchers, organised a conference during the summer of 1956, approaching the subject through the prism of symbolic logic, the branch of mathematics used by McCarthy which deals with representing concepts and statements by symbols.

14)

J. MCCARTHY, M.L. MINSKY, N. ROCHESTER & C. E. SHANNON, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', 1955.

15)

Available at <http://www.arbitration.qmul.ac.uk/research/2018/>).

16)

C. CARRARA, see supra note 7, p 522.

17)

A. SINGH CHAUHAN, 'Future of AI in Arbitration: The Fine Line Between Fiction and Reality', Kluwer Arbitration Blog, September 26, 2020.

18)

C. MOREL DE WESTGAVER, 'Artificial Intelligence, A Diver For Efficiency In International Arbitration-How Predictive Coding Can Change Document Production', Kluwer Arbitration Blog.

19)

Available at <https://restofworld.org/2020/faladdin-turkey-coffee-fortune/>.

20)

K. MAXWELL, *Computers Says No: Data Analytics in Arbitration, Practical Law*, February 9, 2018.

21)

K. MAXWELL, see supra note 20.

22)

Ex multis, F. URIBARRI SOARES, *New Technologies and Arbitration*, 7, Indian J. Arb. p. 84, 2018. See also, The Four V's of Big Data, (available at <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>).

23)

F. URIBARRI SOARES, see supra note 22.

24)

The term 'smart contract' generally refers to 'self-executing electronic instructions drafted in computer code', using blockchain technology as a platform. Smart contracts function according to an 'if-then' logic, enabling the self-execution of payments or the release of funds once certain predetermined conditions are fulfilled. W. MAXWELL & G.VANNIEUWENHUYSE, 'Robots Replacing Arbitrators: Smart Contract Arbitration' in ICC Dispute Resolution Bulletin, Issue 1, 2018.

25)

D. MOLINA, 'Las Nuevas Tecnologias Extinguiran El Sistema Arbitral? Kleros: Una Mirada AL Futuro del Arbitraje Internacional', Kluwer Arbitration Blog, September 30 2020.

26)

Put it differently: *Accepteriez-vous d*'être *jugé par des algorithms?*, see A. GARAPON, J. LASSÈGUE, 'Justice Digitale', (2019), PUF Paris (eds.).

27)

K. MAXWELL, 'Summoning the demon: robot arbitrators: arbitration and artificial intelligence', Practical Arbitration Blog, January 17, 2019, (available at <http://arbitrationblog.practicallaw.com/summoning-the-demon-robot-arbitrators-arbitration-and-artifi...>).

28)

See, by way of example, article 812 of the Italian Procedural Civil Code.

29)

M. SCHERER, 'How Artificial Intelligence will change dispute resolution', cit. p. 563.

30)

M. SCHERER, 'International Arbitration 3.0. Artificial Intelligence and Legal Decision Making', p. 565, cit.

31)

Ex multis, A. TVERSKY, D. KAHNEMAN (September 1974). 'Judgment under Uncertainty: Heuristics and Biases'. *Science*.; Kahneman, D., Slovic, P., & Tversky, A. 'Judgment under uncertainty: Heuristics and biases' (1st ed.). Cambridge University Press (1982); D. Kahneman, A. Twersky, Choises, 'Values and Frames'. Cambridge University Press (2000); D. Ariely, 'Predictable Irrationality, Harper Collins Publishers', 2008; G. Rojas Elgueta, 'Razionalità limitata ed efficienza nel procedimento arbitrale', Errori Cognitivi e Arbitrato (Azzali, Morera & Elgueta eds. 2019).

32)

D. KAHNEMAN, Thinking Fast and Slow (2011).

33)

M. MOSES, 'Public Policy: National, International and Transnational', Kluwer Arbitration Blog, November 12, 2018.

34)

G. ARGERICH, M. B. NOODT TAQUELA, 'Could an Arbitral Award Rendered by AI Systems be Recognized or Enforced? Analysis from the Perspective of Public Policy', Kluwer Arbitration Blog.

35)

L.A. MISTELIS AND S. L. BREKOULAKIS, 'Arbitrability: International and Comparative Perspectives', *International Arbitration Law Library*, Volume 19, pp. 3 and 4.

36)

C. ROVELLI, 'Helgoland', Adelphi (eds.) (2020).

37)

J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, E. SHANNON, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', 1955

38)

J. SEARLE, 'Can Computers Think?', Minds and Brains, (1983) Harvard University Press

39)

C. CARRARA, ses supra note 7, p. 529

Document information

Publication

- Dispute Resolution Journal

Bibliographic reference

Paul Bennett Marrow, Mansi Karolt, et al., 'Artificial Intelligence and Arbitration: The Computer as an Arbitrator—Are We There Yet?', in Gregory Kochansky (ed), Dispute Resolution Journal, (© Kluwer Law International; AAA-ICDR 2019, Volume 74 Issue 4) pp. 35 - 76

Artificial Intelligence and Arbitration: The Computer as an Arbitrator—Are We There Yet?

Paul Bennett Marrow; Mansi Karolt; Steven Kuyan

(*)

(†)

(‡)

> Indeed, short of authorizing trial by battle or ordeal or, more doubtfully, by a panel of three monkeys, parties can stipulate to whatever procedures they want to govern the arbitration of their disputes; parties are as free to specify idiosyncratic terms of arbitration as they are to specify any other terms in their contract.
>
> *Baravati v. Josephthal, Lyon & Ross, Inc.*, 28 F.3d 704, 709 (7th Cir. 1994) (Posner, C.J.).

In this age where big data is commonplace and computers are becoming more powerful every day, Artificial Intelligence ("A.I.") has become a fact of life and is here to stay. Computer scientists posit that with enough data and properly designed algorithms, the "well-trained" computer should soon be able to produce an acceptable arbitration award. The necessary data set would hopefully include thousands of transcripts from actual arbitration proceedings; hundreds of thousands of actual awards; all known reported judicial opinions issued by courts throughout the United States embodying
P 36
the complete state of arbitration jurisprudence; all relevant statutes and rules used by judges, lawyers, arbitrators, and the administrators of the arbitration process; and all known journal and law review materials. [(1)] If these computer scientists are correct, and the authors of this paper believe they are, the results would unquestionably be a game-changer in resolving legal disputes in many parts of the third world as well as in many industrialized countries. Today, millions either have no access to an existing system of justice or have access only to find the system badly choked by bureaucratic inefficiencies, costs that are beyond their reach, and/or corruption. The disputants are left to either take the law into their own hands, often resulting in a violent confrontation, or abandon their claims altogether. Either way the result is a distaste for the system under which they live and a disregard for the rule of law.

Arbitration by computers without human intervention offers an exciting alternative. Computers have no schedules. They can handle numerous tasks simultaneously. Computers need only electricity and a connection to the internet in order to perform any assigned task. Once up and running, a computer-based system for arbitration would be relatively inexpensive and thus within the reach of most disputants, no matter their economic status. And most important, computers don't do personal favors, demand fidelity and/or take graft.

As will be discussed throughout this paper, A.I. provides the bridge connecting the disregarded and/or overlooked disputant to a meaningful disposition of a dispute. This will be welcomed news that in the end will serve to encourage a respect for arbitration as a meaningful way to resolve disputes based on the rule of law and available to anyone. While it is true that even with the advantages of A.I., the outcomes may never be absolutely perfect, still a determination by computer is far preferable to no determination at all.

Humans have routinely believed that they can always make improvements on the performance of computers, and this is reflected in the way
P 37
that computers have evolved over the last 100 years. Computer science is a never-ending process striving for perfection. Imperfections, once identified, are the subject of study and research and, more times than not, result in the discovery of a way to eliminate the imperfection. The demand for bigger and better never ceases. Breakthroughs such as the development of hardware capable of supporting "big data" and computers using quantum physics for the processing of complex algorithms and huge data sets confirm this. With time, the concept of arbitration by a computer will very likely become an acceptable norm, providing considerable advantages over courtroom-based

dispute resolution.

Achieving this goal, however, will be challenging, and numerous questions will need to be addressed and answered. Here is a sampling:

1.Is there a risk that data can be biased, *i.e.* skewed by any number of conscious and/or unconscious factors; and if so, can bias be identified and eliminated?

2.Is the computer a "black box" operating in ways that are beyond our ability to understand, and if so, are there algorithms that can assist us in understanding how the computer is actually operating?

3.Assuming A.I. driven awards become a reality, will the award be persuasive, insightful, and timely, or mechanical, predictable, and rigidly correct?

4.Will state and federal laws that govern arbitration need to be amended to allow the use of A.I.?

5.Can a party consent to using a system that he/she doesn't fully understand?

6.Which disputes are best served by A.I. acting as an impartial neutral?

For the moment at least, the computer operates in a robotic manner, and this is likely to remain so for the foreseeable future. [(2)] Theorizing, creative
P 38
thinking, and robust understanding, the things humans do best, remain beyond the reach of computer science. Today humans determine the problems the computer is called upon to solve and humans define the instructions needed to solve those problems. That is not to say that the computer is to be dismissed as a device good only for crunching numbers and/or performing other robotic-like tasks. Quite the contrary. The human brain isn't designed to absorb and process great quantities of materials used in advocacy. No single person could address, digest, process, and evaluate a data set composed of over one million emails, whereas a properly "trained" computer does this with ease. Human thought processes are slowed measurably in the face of large amounts of data. Efficiency and speed are the primary reasons humans invented the computer. A properly equipped and "well trained" computer can digest new data, look for patterns, and make predictions and recommendations. And it is not uncommon for a computer to spot an unknown trend or pattern.

Today's A.I. is used for tasks such as legal research, drafting of contracts, corporate records, preparation of research memos, drafting of pleadings, facilitating document discovery, and providing language translation and interpretation, to name only a few. [(3)] Computers can review existing documents, detect and report on deficiencies, and make recommendations on ways to improve what it has reviewed. Computers can review briefs before they are filed; update research; and eliminate grammar, spelling, and formatting errors. A computer can recommend variations of any argument and even propose new arguments that may not have been previously considered. Case management is incorporating A.I. for tasks such as scheduling of meetings, telephone answering services, docket control, and the creation and mailing of standard form letters. Support services for courts and arbitrators now incorporate A.I. A downside is that jobs associated with these tasks are being lost. The ranks of administrators, secretaries, law librarians,
P 39
and paralegals are thinning with each passing day as computers take over many duties humans formerly performed.

On the upside, there are examples of AI being used to handle routine, transactional matters. For example:

•Online Dispute Resolution ("ODR") is readily available at sites such as Modria. These services provide a structure for processing a dispute with access to a human mediator and tools for evaluating the merits of claims.

•The American Arbitration Association offers the Modria Resolution Center for certain kinds of disputes. [(4)]

•eBay has a dispute resolution center and offers advice on how to best use it.

These services and others like them stop short of offering a computer as an arbitrator.

Many states and local jurisdictions have begun to implement programs designed to allow citizens access to dispute resolution online:

•Michigan has a program allowing citizens to resolve certain civil disputes online. [(5)]

•In Ohio, [(6)] New York, [(7)] and Texas, [(8)] there are online programs allowing citizens to submit real estate taxation disputes and traffic ticket challenges, with humans making the final determinations.
P 40

•Utah has a program for online resolution of small claim disputes. [(9)]

These efforts have begun to address the current shortfalls of the judicial system(s) in the United States and elsewhere, [(10)] suggesting that continued research will enhance the possibilities A.I. presents. For the moment, at least, perhaps the nomenclature "artificial intelligence" should be refashioned "augmented intelligence," [(11)] *i.e.*, robotic intelligence to be used by humans to assist in reaching as accurate and unbiased a judgment as quickly as possible.

This article looks beyond "augmented intelligence" and asks the question: Can A.I. be trained to a level that allows A.I. to replace an arbitrator and make final, binding awards? From the perspective of arbitration as a process, the answer is, if not now, then soon. If it is assumed that within a given class of disputes, every conceivable outcome has already been determined and resolved, and all relevant data points were known, the task would be simple. Presented with a factual scenario and asked to compare it to identical cases labeled and described in a data set, the computer would easily make a prediction of the likely outcome. But the likelihood of the circumstances being identical for any two cases is low. In contested disputes, by definition the parties see the facts and the law from dramatically different vantage points, and resolution is often difficult for the human arbitrator. For the human, what is important are the facts, the evidence being produced to prove the facts and, to some extent, applicable law. The human arbitrator processes this information using deductive reasoning, logic, established rules and, when necessary, common sense. For the most part, human arbitrators are totally unaware of the thousands or even millions of similar cases and any statistically relevant patterns in those cases. That there is a data set containing similar cases with claimants winning 52% of the time
P 41
and respondents 48% of the time is of no interest and has no influence on the arbitrator's decision. On the other hand, computer science instructs computers to make determinations using only pattern recognition and statistical formulations, with no consideration given to deductive reasoning, logic, or common sense. It's not surprising that computers and humans may literally see and process identical facts differently. The task ahead is figuring out how to unify these perspectives, so it is faithful to our systems of law *and* allows for an analysis tailored to the individual case at hand.

Today, there is no shortage of unresolved disputes of all kinds: large and small, simple and complex. And there is nothing suggesting that in the foreseeable future the number will not grow, perhaps even at an exponential pace. Driven in part by judicial and legislative findings that virtually any dispute is arbitrable, arbitration has become a major go-to alternative to the courthouse. Unfortunately, arbitration, once championed because of cost savings and other efficiencies, has become just the opposite, placing arbitration beyond the reach of many who seek an efficient, cost effective, disciplined, and fair alternative to the courthouse. Decision-making machines can address these deficiencies, making arbitration appealing. [(12)]

As the reader will see shortly, not every case is likely to lend itself to arbitration using A.I. Disputants and their lawyers involved in what have been called "bet the company cases" are the least likely to want to entrust such issues to a computer. In all likelihood, the best matches will prove to be small claim disputes stalled by a clogged judicial system that is bureaucratic, overwhelmed, and/or encrusted with corruption, and cases involving similar facts. In addition, remote access to a decision-making machine using internet technologies should eliminate bureaucratic delays and increase the ability of many who lack the means, or the time needed to travel distances to a courthouse, to resolve their disputes. The inability to obtain an unbiased and timely determination of a dispute creates a real risk that the rule of law will break down with parties even taking the law into their own hands. Recent history has shown us that one characteristic
P 42
of a failed state is widespread frustration across a population because of a lack of access to a practical, timely, and fair process for resolving matters important to them. No matter the type of case, it is reasonable and in fact, necessary, to see if A.I. can address the challenges presented. Only time will tell.

Acceptance of an award fashioned by A.I. requires that parties trust what the computer appears to be doing, *i.e.*, processing information and reaching an appropriate determination. Realistically though, the public's current perception is that the computer is a "black box," whose operations are beyond comprehension, principally because the computer cannot account for how and why it reached a given outcome. This lack of transparency has resulted in demands for a so called "right to explanation." Addressing this demand has generated extensive research into the development of appropriate procedures, including algorithms capable of compelling accountability and transparency, and has triggered governmental responses and demands:

•In 2016 the European Union's Parliament adopted the General Data Protection Regulation ("GDPR") [(13)] that became effective in 2018.

•In late 2017, New York City, suspicious of algorithms used to determine the allocation of everything from food stamps to firehouses created a fact-finding task force to determine if the algorithms were performing in a fair and equitable manner. (14) The task force released its report in November, 2019. (15)

I. Arbitration as the ideal candidate for computer-driven dispute resolution.

The first step is finding out if A.I. and arbitration are indeed compatible. The academic literature to date has had little to say about this relationship.
P 43
The focus of research to date has been on the relationship between A.I. and courthouse litigation. Arbitration is quite different from courthouse litigation. Arbitration is not intended to be a *substitute* for litigation. Arbitration is a meaningful *alternative*. This article argues this difference is so consequential as to make arbitration a far better candidate for A.I. applications than litigation. Arbitrators and A.I. are subject to constraints on how each operates, constraints not imposed on a judicial system. For A.I., performance is limited by the quality and quantity of training data and the quality and robustness of the algorithms. For the human arbitrator, performance is limited by the terms of the arbitration clause and governing law.

First consider the limitations placed on the arbitrator. Not being judges, arbitrators are rarely, if ever, granted authority to consider factors external to the case at hand; factors such as political trends and changes in societal norms. Arbitrators cannot disregard, modify, or defang existing law and/or precedent. Arbitrators can do no more than consider a prescribed factual scenario and apply the law as required by the parties. The arbitrator's first obligation is to the parties and to the terms of the agreement to arbitrate. By contrast, judges have a first obligation to the law without concern for any agreement by parties limiting the ability to apply, interpret, and even nullify statutes, precedents, and rules and regulations. Judges are free to consider changing societal norms and conditions. Judges can resolve a case of first impression, *i.e.*, a case not known to have been evaluated by any judge, and proclaim rationales for overturning precedent or voiding statutes and administrative rules and regulations. (16)

Next consider limitations placed on A.I. Unless allowed by the designer, A.I. can't operate outside its instructions. A.I. must obey the mathematical
P 44
and structural limitations humans impose. The human designer restricts A.I.'s understanding of our world to the training data provided. Unless instructed to do so, A.I. is unable to access data on its own initiative. A.I. doesn't even know there is a world beyond what humans define for it. (17) While arbitrators can think and reason, they are constrained by the legal limitations of the process. A.I. can't think independently and reason, but it can mimic arbitrator performance if trained by humans about the restrictions imposed on arbitrators. The aggregate of all these limitations leads to the supposition (to be confirmed) of a "perfect match," meaning that arbitration is a superior platform for integrating A.I. into the dispute resolution process.

A. The limitations imposed on arbitrators and A.I. make arbitration an ideal candidate for a computer-driven process.

These limitations on an arbitrator and on A.I. come into play when issues involving existing law require consideration. An established rule requires that unless the parties provide otherwise, the arbitrator can only apply the law as it actually exists and cannot add conditions because the arbitrator believes given circumstances appear to be slightly out of line with the law. In addition, the arbitrator isn't allowed to modify the law by considering evolving needs created by societal changes and pressures. Here are two examples of situations where these constraints are brought into sharp focus.

Suppose a buyer and seller contract for the purchase and sale of tires to be shipped to the buyer in Newark, New Jersey from Houston, Texas. The seller, in violation of the Jones Act, contracts to use a ship that isn't registered in the United States and is manned by citizens of the Philippines. The Coast Guard seizes the ship and holds it for ten days. The contract between the parties calls for arbitration "of any and all disputes" arising from their agreement to buy and sell tires. Query: Are the consequences of the delay caused by the seizure of the ship and therefore the substance of
P 45
a dispute within the meaning of the arbitration clause, *i.e.*, is this dispute "arbitrable"? And who decides the issue, an arbitrator or the court?

In *AT&T Technologies v. Communication Workers of America, et al.*. (18) the United States Supreme Court declared: "Unless the parties clearly and unmistakably provide otherwise, the question of whether the parties agreed to arbitrate (a substantive issue) is to be decided by the court, and not the arbitrator." [(19)] But the Court stopped short of addressing the question of "who" shall decide if there is clear and unmistakable evidence about the intent of the parties, and what standard must be used when making that determination. In *First Options of Chicago v. Kaplan*, [(20)] the U.S. Supreme Court established that the answer to this "who" question was the same: it depends on whether or not there is clear and unmistakable evidence that the parties want the arbitrator and not the court to decide.

Before the *First Options* case, an arbitrator had to defer to a court for an answer the question of whether clear and unmistakable evidence existed, and to do otherwise would have resulted in a declaration that the arbitrator had exceeded his or her authority. This would be so even if the arbitrator believed clear and unmistakable evidence actually existed. Similarly, a computer would have had to defer to a court for the answer. But the day after *First Options* was handed down, the arbitrator could determine the presence of clear and unmistakable evidence and therefore so could a computer.

First Options doesn't address the role that societal pressures play in the evolution of jurisprudence. Consider *Brown v. Board of Education,* 374 U.S. 483 (1954). This groundbreaking case and the abolishment of the "separate but equal" standard came about because the Supreme Court was willing to consider societal pressures and changing norms. In May of 1954, on the day prior to the handing down of the decision an arbitrator would have had to apply the "separate but equal" standard even if the arbitrator
P 46
believed that standard offended the 14th Amendment of the U.S. Constitution. Similarly, if charged to report on the law of desegregation as of the day before the *Brown* decision came down, the computer could do no more than recommend applying the "separate but equal" standard. If asked to review an arbitrator's ruling, issued the day before *Brown*, declaring the separate but equal standard unconstitutional, the computer would have to deem the ruling defective and an example of an arbitrator who has exceeded authority.

In the discussion that follows, we focus on two types of cases likely to be presented to a decision-making computer. *First Options* and *Brown* are both examples of a law case: one where the facts are not in dispute. The second is a fact case: one where the arbitrator and therefore the computer is called upon to determine the facts, determine the credibility and authenticity of written evidence and witness testimony, and apply applicable law. Here is the example:

> Hadley had his tailor hand-craft a dress shirt to be worn the day Hadley was scheduled for an audience with Queen Elizabeth. The tailor used only the finest quality fabrics and the shirt fit Hadley perfectly. It cost $450 plus sales tax at 7%. Hadley tried wearing it a week before his audience and inadvertently spilled mustard on the right sleeve. He took the shirt to his local dry cleaner and asked to have the stain removed. The dry cleaner accepted the shirt and issued a receipt containing an arbitration clause requiring arbitration before the Technically Savvy Arbitration Association, New York, New York using the Association's Commercial Rules then in effect. (Those rules are identical to the Commercial Rules of the American Arbitration Association (the "AAA").) When the shirt was returned, the left sleeve was badly burned. Hadley determined that the shirt was a total loss.
>
> The case manager has advised that it would be less expensive for the parties to submit the dispute to a specially equipped
> P 47
> arbitration model [(21)] containing an extensive, clean, and structured data set with over 100,000 dry cleaner burned shirt cases resolved by small claims courts throughout the New York Metro area and by arbitrators. The computer also has a complete library of New York law, including all reported judicial decisions of every kind, all New York statutes and governmental administrative rules and regulations of all kinds, and an extensive collection of secondary legal materials including digests and journals, form books, and other materials relied on by judges, arbitrators, and the arbitration administrators doing business in the New York Metro area. The case manager also advises that the model has been outfitted with extraordinary algorithms designed by the most sophisticated and respected computer scientists in the world, algorithms tested in simulated and real-world environments that have received international and national

> recognition and awards.
>
> Hadley submits an affidavit swearing as truthful the shirt was brand new, never worn until the day he put it on and spilled mustard on the right sleeve. He states emphatically the left sleeve was in perfect condition when he left the shirt with the dry cleaner. The dry cleaner submits an affidavit claiming the shirt wasn't burned by any of the machinery in his store. He further swears that the pressing machine couldn't burn a shirt because of a fail-safe mechanism designed to shut the machine down if the temperature of the iron exceeds a safe level. Attached to his affidavit is a copy of the manual for the pressing machine and a statement from a repair man that the machine was in proper working order the day of the incident.
> P 48

B. A brief look at the technology involved with A.I.

Few professionals engaged in arbitration are also computer scientists, so getting the tangle of weeds that computer scientists routinely trouble over is not necessary for our purposes here. This discussion is focused on the practical, not the genius, involved in making A.I. a valuable tool.

For our purposes, A.I. is defined as the creation and use of basic algorithms designed to allow a computer to use training, data provided by the architect, to search through yet unseen sets of data looking for patterns and trends to answer a query with an almost certain probability. Overseeing and managing a model's activities are "predictive" algorithms, i.e. instructions telling the computer the steps needed for both dependent and independent performance. [(22)]

The science behind the creation of predictive algorithms is called "machine learning." There are three types:

1.Supervised learning,

2.Unsupervised learning, and

3.Reinforcement learning.

Deep learning is a subcategory of Unsupervised learning and today is the area receiving the most attention from the computer science community.
P 49
Deep learning involves the use of layered neural networks designed to mimic how human brains learn, by strengthening neurological pathways with repetition. Unlike Supervised learning that requires human control over data input, training, and the design of algorithms, Deep learning:

•Allows the computer to identify both apparent and hidden patterns embedded in very large sets of either labeled or unlabeled data, and

•Permits the machine to create algorithms necessary to make predic-tions. [(23)]

These models use pattern-recognition probabilistic methods. [(24)] To some this may sound like magic, but it actually involves credible and knowable mathematical methods to do this using a computer with suitable power to execute a huge number of calculations. The hope is that in short order new Deep learning methods will evolve allowing a computer to operate on its own, solving problems without any human support. Good data is at the core of any A.I. algorithm, and the quality of the algorithm is directly correlated to the quality of the data set used. The main function of A.I. is the unlocking of information and knowledge embedded in data. A.I. data only becomes useful if it's cleansed, well-labeled, annotated, and properly prepared for relevant input and analysis. This requires a significant investment in resources and time. Once data is prepared, it is ready for processing. See Figure 1 for further detail on the data lifecycle.

Computers can now receive, absorb, catalogue, and store data from many sources and in many forms, and can even receive and process data real-time without first having to store it. Advances in algorithm design now allow computers to:

•Interpret and understand human speech,

•Accept and read the written word,
P 51

•Translate text to speech,

•Receive and interpret visual information, and

•Permit computers to independently monitor and update data from sources on the internet and elsewhere. [(25)]

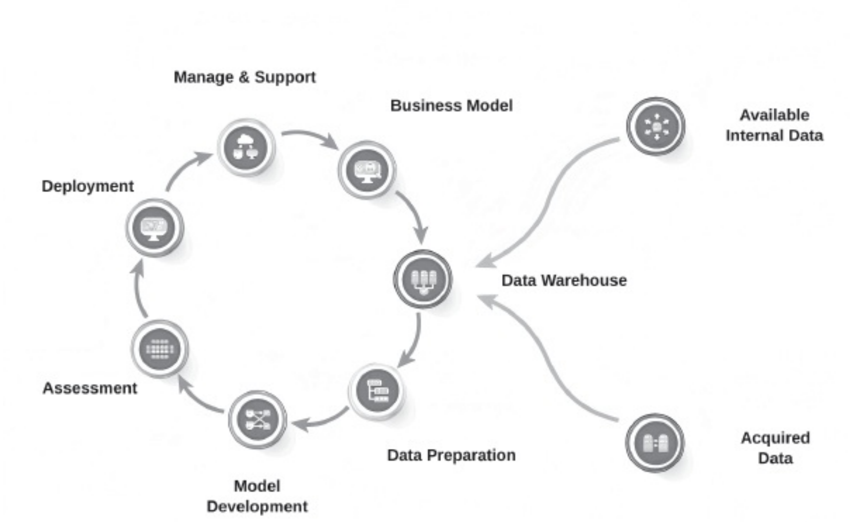

Figure 1. Cross-industry standard process for data mining

With Supervised learning, data is labeled by a human and algorithms provide the required processing instructions. The techniques used have interesting names like "decisions trees" and "random forests," code words for finding a pathway leading to any one of the many possible end points. For example, an algorithm can be developed and trained to identify what a judge looks like using a large labeled data set composed of photographs of individuals exhibiting known features unique to members of the American judiciary: their robes, a gavel in hand, and perhaps gray hair and glasses. Using this data, the computer can be asked to determine whether a previously unseen picture of a person is a judge. With Supervised learning, a series of step-by-step instructions can require the computer to compare points in the picture one at a time, with labeled training data to support the conclusion of the presence of an element of a feature. If the comparison falls short, the computer is instructed to abandon the effort and a new search is ordered. If the comparison is successful, the computer is instructed to store the result and proceed to an analysis of the next point in the picture until sufficient information supports a final conclusion that the feature is present. With Deep learning, a very large unlabeled data set is used. The machine "learns" from patterns that are identified, stored, and then used to reach a conclusion the picture is or is not a judge, with a degree of certainty as close to 100% as possible. [(26)] No matter the system for learning, the end result is
P 52
a model that, when shown pictures not in the training data, is capable of answering the inquiry as to whether or not a picture is of a judge. Unfortunately, this model can't tell you if it's a picture of an arbitrator since it was only trained on pictures of a judge. It's important to note the overwhelming majority of A.I. tools are designed for a specific purpose and are not models worthy of broad application.

Machine learning techniques have evolved to a point where it is now possible to teach a computer a great deal about litigation and arbitration. However, some have raised the objection that data sets may contain embedded biases capable of undermining the ability of A.I. to be objective. [(27)] This is a justified concern, and we will discuss it later in this article.

For our purposes, assume any data set composed of arbitration related case law materials can be separated into three categories:

1.Cases involving a question of law,

2.Cases involving issues of fact, and

3.Cases involving both questions of law and issues of fact.

We will use *First Options of Chicago v. Kaplan* and *Brown v. Board of Education* as examples of cases involving a question of law. The computer's task is to understand the law involved. The case involving Hadley's shirt and the dry cleaner is one in which there are central questions of fact and credibility. As a practical matter, all cases "arbitrated" by a trained computer will involve questions of both fact and law.

Let's now look at some concerns with and questions about the concept of machine-driven dispute resolution and final arbitral awards.
P 53

II. Question #1. How different is arbitration from courthouse litigation and why are the differences important?

Arbitration is *not* a substitute for litigation. (28) The only substantive similarity is a ruling on the merits. In litigation technical rules for pleading, discovery, and evidence are the norm. Not so in arbitration. In litigation, a final finding leads to a judgment and, in arbitration, an award. Judgments can be appealed as of right. Arbitration awards cannot be appealed with one exception: if the parties agree to submit the award to yet another arbitrator for review on the merits (this is very rare), whereas the finality of a judgment is not established until the right to an appeal is exercised or allowed to expire.

The most significant differences between arbitration and litigation involve reach and scope of authority. In the courthouse judges have broad discretionary powers. A judge's authority is found in the law they are sworn to apply and uphold. Not so for an arbitrator. The baseline for determining what an arbitrator can and cannot do is found in (a) the arbitration clause, (b) the limits imposed by statute, and (c) judicial interpretations of both the arbitration clause and controlling statutes. The arbitrator's first obligation is to the requirements specified by the parties, as long as their requirements do not offend public policy or are outright illegal. Judge Posner, quoted at the beginning of this article, reminds us that with the parties in full control, only trial by battle or ordeal is unacceptable.

Unlike in litigation, in arbitration the parties control several key issues by including or excluding any number of items in the arbitration clause. For example, they are free to provide rules for the selection of an arbitrator(s), rights to discovery, the law they want applied, the application of the rules of evidence, the timetable for the hearing, and the rules for the hearing. (29) They can even use a form of shorthand to provide for most of the items
P 54
on this list. In the Hadley scenario, the clause is so short as to appear to be incomplete: "arbitration before the Technically Savvy Arbitration Association, New York, New York using the Association's Commercial Rules then in effect." In reality, this clause is complete because the provision incorporates by reference the commercial rules. And those rules provide that the parties can modify or eliminate any rule. (30) In Hadley's case, the applicable rules haven't been amended by the parties.

Significantly, there is no mention any place in the AAA Commercial Rules of the need for the arbitrator to apply law, meaning that unless the parties say otherwise, the arbitrator may do that which is thought to be reasonable. Similarly, absent a contrary direction from the parties, AAA Commercial Rule 34(a), "Evidence" provides "Conformity to legal rules of evidence shall not be necessary."

The ability of parties to exercise control over the arbitration process has implications for the application of A.I. Consider a situation in which the parties have required application of the laws of a given state. Selecting data sets containing those laws along with judicial decisions interpreting them is, in theory, feasible and simple. But if the parties fail to demand application of any specific law, the human arbitrator and therefore a mirroring computer would be charged to do what is thought by humans to be "reasonable." Selecting appropriate sets of data tailored to this situation may prove a daunting task given the unbounded debate over what the term "reasonable" actually means.

All commercial contracts that include an arbitration clause are automatically subject to the Federal Arbitration Act ("FAA"). (31) Every state has its own version of an arbitration act, and parties are free to choose between the FAA and a state's arbitration act, subject, however, to the rule of federal preemption. (32) The FAA provides for the vacating of an award if the arbitrator runs afoul of Section 10(a) and all state arbitration acts have similar,
P 55
although not necessarily identical, provisions. Most important for our discussion is FAA. Section 10(a)(4) allows for vacating "where the arbitrators exceeded their powers . . . ." Arbitrator authority, discretion, and powers are limited by the agreement of the parties. Exceed those limitations and the award is subject to being vacated.

What constitutes exceeding powers is not always clear, and there's a large body of jurisprudence, federal and state, analyzing and discussing this topic. Libraries of this case law are readily available for training a computer. If a question arises about applying New York law, as opposed to some other law or no law at all, the computer would have to make that determination based on a review of a New York case law database. But suppose a computer isn't capable of deciding because no judicial teaching exists (the case is one of first impression), or there is a conflict of opinion between courts in the jurisdiction where the dispute is being heard? Algorithms can be

designed to instruct the computer to signal the need for human intervention if the computer determines circumstances exist requiring the computer to perform beyond its scope. (33)
P 56

FAA Section 10(a) (34) creates a unique condition not found in courthouse litigation. If the human arbitrator's authority and powers are restricted, so must be the authority and powers of any computer mimicking the human arbitrator. Ensuring the faithful emulation of a human arbitrator is the task of the designer of an algorithm. The ability of the designer to confirm the presence of an appropriate algorithm should serve to mollify fears that an algorithmic-driven computer might be dispensing what the computer, or its human masterminds, believe to be their version of justice. (35)

III. Question #2. Is a computer really just a black box? Exploring and resolving issues of embedded bias.

On numerous occasions courts have held that parties in an arbitration aren't entitled to a perfect hearing. They are, however, entitled to a fair hearing. (36) What constitutes a fair hearing isn't always an easy measure, but at the very least, the arbitrator must operate in the open where the parties can
P 57
observe demeanor and professionalism. No arbitrator can operate behind a curtain. (37) Transparency, meaning overt behavior consistent with a commitment to neutrality and freedom from prejudice and bias, presents unique complications for the use of A.I. in an arbitration. While a computer can be programmed to resolve a problem, for the moment, algorithms directing the computer to explain to the user why it is doing what it is doing are still a work in progress. (38) So, for now there is a perception of computers being a "black box" that may be doing the bidding of someone other than the user. If the computer cannot account for its actions, how can a user know that data sets and algorithms are not tainted by undisclosed, undetectable and/or unintentional biases? (39)

To put this into a proper context, first consider whether human arbitrators are likely to be biased. There's no shortage of cognitive psychological
P 58
studies pointing to "yes." Through observation, parties can detect some biases. An arbitrator who shows a preference or distaste for someone based on race, religion, gender, or some other unique factor should be easy to spot. For example, a showing that 30 out of 30 awards favor a white complainant over a black respondent does justify inquiring if the arbitrator harbors racial animus. But sometimes it's not so clear. For example, a pattern of awards disproportionately favoring banks over consumers seems to suggest a bias. But there may be other explanations. Perhaps the reason has to do with the number of respondent consumer defaults leaving the arbitrator no choice but to rule in the bank's favor. (40) The takeaway is that while a pattern of behavior may point to a known bias, there is still a need to use caution and do further analysis.

Cognitive unconscious biases are a different matter. Empirical studies have shown that arbitrators, judges, and juries bring to their roles hidden biases that often they themselves are unaware of. These biases, some call them blinders, (41) result from the human tendency to use heuristics—mental shortcuts—when making decisions. (42) Most individuals are totally unaware since these biases are embedded in the unconscious and are often further supported by cognitive dissonance. Anyone asking an arbitrator about an unconscious bias is not likely to get meaningful information because the arbitrator, unaware of the unconscious process, will in good faith deny falling victim to bias. For instance, a "coherence" heuristic was identified in a study involving judges who were asked to estimate their reversal rate relative to their peers. The study showed that judges are inclined to believe their own rulings are correct at a rate exceeding that of their peers. In the study, judges had to estimate their reversal rates by an appellate court. Fifty-six percent rated themselves in the lowest reversal rate group and 31% rated themselves in the next lowest reversal rate group. These results suggest
P 59
that 81% of the judges' thought that at least half of their peers had higher reversal rate records than they had. (43) But if you were to ask any of the participants if he or she had inflated their perception of judicial acumen, the general answer, given in good faith, would be "of course not." Another example involves studies showing how the "anchoring effect" heuristic impacts the ability to estimate an unknown quantity. If a participating judge attempting to settle a matter is given a number, even if this number is random, the participating judge will likely anchor to the provided number, making any estimate highly unreliable. (44) If asked if there is any factor influencing their determination as to what constitutes a reasonable number for settlement purposes, the most likely answer might be "no."

Factors unknown to a judge, even factors that seem totally irrelevant to the dispute at hand, can influence decisions. For example, a famous study involving Israeli judges hearing applications for parole revealed a direct connection between the likelihood of securing parole and the timing of a judge's lunch break. While applications for parole were infrequently granted, it turned out chances greatly improved at the very beginning of the working day and again after a judge's lunch break. [(45)]

Influencing and/or defining the algorithm design process is fertile ground for the insidious involvement of unconscious biases. The same is true for training data. [(46)] An algorithm needs a structure and the determination of the structure can easily involve a subjective criterion not apparent to the author or those using the output. To state the obvious, how an algorithm processes data dictates the objectivity and value of the output. [(47)] For example

P 60

, consider an algorithm designed for learning from data being contaminated because it allows for the inclusion or exclusion of certain parameters. The scope of this problem, however, is somewhat contained by the ability of the architect to physically reexamine the design of any algorithm. More insidious is the potential for the failure or inability to properly manage and screen input data.

Training data originates from two sources. It can be selected by third parties and/or by the architect of the algorithm. A major risk is training data contaminated with unconscious biases of the people involved in the selection process. One observer noted it's easier to detect error/bias in algorithm design than detect and correct them in training data sets. Since learning models also retrain and reinforce using prior results obtained with tainted data, the biases are likely to become self-perpetuating, further detracting from the value of the model. [(48)] As a result, there is a real risk over time that undetected data set errors and biases can become so deeply embedded as to take on meanings of their own and eventually change the algorithm without human detection. The experience of Staples provides an early example of our blindness to the outcomes of algorithm and data selection designs containing imbedded, albeit unintentional, bias. Staples deployed an algorithm that unwittingly discriminated against certain consumers based on social and economic status. The root of the problem was traced to the training data. Staples identified an unintended bias allowing the offering of reduced prices to buyers in more affluent neighborhoods of means when the purpose of the algorithm was to offer the lower prices to a less privileged population. The investigation discovered that the training data contained an assumption that those living closer to a brick and mortar location of a competitor such as OfficeMax would be less affluent and therefore more price sensitive than those living further away. It turned out that less affluent buyers actually lived further from the competition and yet were shown higher prices than the more affluent living nearby the competition.

P 61

The final conclusion was the algorithm was relying on skewed training data that contained an undetected bias favoring the affluent. [(49)]

Consider the example given earlier of training a computer to recognize a judge. If the photos shown to the computer include only elderly men with gray hair, glasses, and black robes, the computer, relying on the training data, will fail to identify pictures of young men and women and older women, no matter the color of their hair or their garb, as judges. [(50)]

Data may be biased simply because of outside and/or societal pressures that have helped to shape the data. For example, if the data contains information that is correct and yet unbalanced, any prediction made using that data will be unbalanced. Assume data sampled shows a ratio of four (4) apples to every one (1) orange, and an overall likely rotting rate of 10%. It follows that the outcome will reveal far more apples have rotted than oranges. Now consider a sampling of convicted criminals 85% who are black and 15% who are white used in a study to determine the likelihood of a convicted criminal committing another crime once released from prison. It's a certainty that the resulting statistics will show more recidivism among blacks in the sample than whites. As one observer put it: "The outcome is biased because reality is biased." [(51)]

Designers have a humbling responsibility that must be taken very seri-ously. [(52)]

What is being done to address these concerns? The simple answer is "a lot," but the solutions available to date are only a beginning. Research has led to a number of possibilities. [(53)] One suggested solution is asking the developer of the data set and/or algorithm for an opportunity to audit the inner workings before anything is deployed. While an examination of the

P 62

inner workings of an algorithm or the criteria for data set selection should expose errors and biases of all types, there is no guarantee. [(54)] Moreover, as a practical matter, there is a likelihood of resistance if the author is concerned about the trade secret value of the work. (Some algorithms are open source.)

Another approach is testing using queries designed to show the existence of an embedded bias. If a bias is detected, asking the author to account for it would allow for an explanation and even a solution and is unlikely to be threatening so long as there is no need for the disclosure of sensitive information. However, care must be taken to ensure that any testing techniques are broad enough in scope to be able to detect a broad range of possible biases.

Data sets present yet another challenge. Litigators frequently search on the name of a judge for any insights into how the judge is likely to rule on an issue. Given the nature of arbitration and the emphasis on confidentiality, the reality is that few arbitrators file awards that are available for public review. Some issue unreasoned awards, meaning that an award does not include an explanation of the reasoning behind the decision. And most arbitrators issue awards involving an assortment of topics. For these reasons, it is very difficult to test for a pattern suggesting an individual's bias.

Beyond testing, much effort is being given to what computer scientists call "explainability." It is commonplace for parties to demand a human arbitrator produce a reasoned decision, one that reveals the thinking of the arbitrator and explains the details about the law applied and the evidence found credible. Demanding an algorithm capable of providing this level transparency seems only reasonable. Is the science involved in A.I. capable of directing a computer to explain any award it issues? Programming models to provide an explanation of the parameters used for deciding is a nascent area of research, but one that has shown early promise. (55) The tradeoff has been between "explainability" and effectiveness. While explainable
P 63
A.I. exists (commonly referred to as XAI), the level of "explainability" is inversely proportionate to the complexity of the problem and the deployed model used to solve it. Simple statistical models, like those currently used to determine insurance premiums, credit card rates, or loan approvals are typically based on a decision tree analysis. But for more complex systems, like Deep learning, the research is only now beginning to show progress. As was noted earlier, Deep learning involves algorithms allowing a computer to evaluate data unsupervised. The decision process employed by the computer can be poorly defined and at times appear to the human being as based on nothing more than useless noise. In addition, because a neural network is multi-layered, tracing back the decision process is an undertaking that is highly complex, time consuming, and expensive.

A second concern is determining the magnitude of explainability within a given setting. The ethical and legal issues involved and the degree of explanation required differ for every application of A.I. While the concern has led to discussions among those designing algorithms that permit explanation, among those using them, and those who are considering how to regulate the need for explanations, there is still no agreement on how to set parameters flexible enough to embrace the full scope of a given A.I. application. (56)

A third area of concern involves expectations. Given the concern about the "black box," in an arbitration case, the ideal would be an audit function providing details about the computer's decision-making process indicating what factors, evidence, and law is being considered as well as the factors, evidence, and law disallowed. The ability of the science behind XAI to deliver this type of explanation is not yet fully developed. Most of the work in XAI has involved "simplified approximations of complex decision-making functions." These approximations appear to users to be more like scientific models than the "everyday" explanations the user community is looking for. (57)

P 64

Yet another area of concern is the philosophical implications of the differences between how humans and computers perceive and evaluate a dispute. Scherer (58) observes that humans have developed systems of law that are serviced through deductive reasoning, logic, and the application of known rules. This allows for some flexibility, as required by the circumstances of a specific dispute. A.I., on the other hand, is currently designed to resolve an issue using mechanistic formulas grounded in pattern identification and statistical probabilities derived from the learning data. No matter the training method or the model, A.I. can do little more than provide a predictive output as to a degree of mathematical certainty. Acceptance of this type of protocol could prove a hard sell since it would require users to knowingly abandon the ingrained notion that law is a function of deductive reasoning, logic, and application of known rules. (59)

On the other hand, the use of deductive reasoning and logic can allow the arbitrator a degree of flexibility that at times can entice an arbitrator to exceed authority, something § 10(a)(3) of the FAA prohibits. For example: The arbitrator is confronted with a case of first impression, i.e. a situation that no judge has ever considered and ruled on. An argument can be made that the arbitrator can only apply a known law and lacks authority to fashion new law to accommodate such a situation and that doing anything else amounts to dispensing his or her own brand of industrial justice. This risk is reduced substantially, if not eliminated altogether, using a computer
P 65
trained with a proper database and equipped with algorithms designed to have the computer alert the designer if confronted by such a situation. In addition, the computer's conduct is devoid of

human emotion and subjectivity. Over time these considerations may prove sufficient to allow arbitration by computer to become an acceptable alternative. (60) And finally, arbitration training data is anchored to the law applicable to arbitration. While the computer may be looking for patterns and applying mathematical formulas to obtain a probability, it is doing so using the same materials an arbitrator would when making a judgment leading to an award.

What is clear is the entire field involves many complexities, (61) and there is a need for much further research and work. While it is anticipated and expected that most applications of A.I. will eventually become explainable, we may not be able to answer whether or not the explanations themselves will be sufficient, auditable, and trustworthy.

Beyond testing and XAI, consider using a panel of three independently trained computers or two independently trained computers and one human being. Presumably the panel, however composed, would vote and the majority
P 66
would prevail. The format involving two computers and a human being seems to defeat the purpose of turning decision making over to machines. (62)

Some have challenged the appropriateness of using a "big data" set. The concern is: (1) Big data isn't objective. (2) Big data doesn't consider the evolutionary nature of our law and legal system. (3) Big data risks failing to reflect the true nature of our legal system and instead will reflect a system all its own. (63) These concerns reflect a misunderstanding of the role big data plays in Machine learning and, in particular, unsupervised Deep learning. A well-trained Deep learning computer receives data from any number of diverse sources. The human element is absent from the selection and labeling of data thereby eliminating human source subjectivity. The objection that data might not consider the evolutionary nature of our law shows a misunderstanding of the role A.I. can play in arbitration. Unless allowed by the parties to an arbitration, an arbitrator can never consider changes in societal norms. Indeed, this constraint is one of the most compelling reasons why arbitration is a suitable candidate for A.I. driven programs.

Whether a human or a computer, there is a risk bias will play a role in the decision-making process. But that doesn't necessarily mean an award is unfair. What gives the computer the edge is the ability to uncover and remove bias using simulation techniques.

What might a data set tailored to Hadley's case look like? The design of an appropriate data set would entail a great deal of forethought to ensure its appropriateness. The peculiarities of the basic elements of the case, (let's call them generalized data points) should drive the criteria for the data sets needed to train a computer. The particular model would need to be trained to understand what a shirt like the one involved looks like, a description of the type of fabric(s) used, what mustard is, how it stains fabric, what a mustard stain looks like, what the cleaning process is, what can cause a
P 67
burn on fabric during the cleaning process, what the tell-tale signs of a burn caused by a machine such as the one used by the dry cleaner are, what a burned shirt sleeve looks like, what a shirt that isn't burned looks like, what a dry cleaner is and how they operate, and what is the typical relationship between a dry cleaner and a customer. Photographic information if any, showing the shirt and its condition at any point in the timeline leading to the dispute would also be required. Other general data points might include information about the machine used to clean and press the garment along with the manual(s) describing any fail safe mechanism claimed to have been running the day the shirt was allegedly damaged and information about the proper process for removing a mustard stain. If available, a data set composed of cases involving damage claims against dry cleaners and verdicts together with a library of photographs showing similar burns might be considered. Other general data points to be considered might include information about either or both parties' prior lawsuit/arbitration activity. Addressing issues of credibility might require developing algorithms trained on data sets composed of examples of shirt owners and dry cleaners who are known to either be lying or telling the truth.

Dry Cleaner consumer disputes occur throughout the U.S. Each jurisdiction charged with resolving these disputes has its own rules and laws that often vary. Some jurisdictions may not have a population able to afford to purchase a shirt of the quality of the one Hadley bought, raising the question of whether or not it is appropriate to include a "proxy" for such shirts. If shirts aren't the measure, should other garments similarly priced be used as a proxy? While a $450 shirt may not be the norm, a $450 men's suit might be. The point becomes obvious; great care is needed when determining the scope and nature of the data.
P 68

IV. Question #3. How complete need the data set be before it can be used?

The more robust a data set the better A.I. and the computer will perform. [(64)] No data set ever is complete, there is always room for additional informa-tion. [(65)] Within the context of arbitration, having on hand all the relevant "book knowledge" available through services such as LEXIS and Westlaw and more is important. Access to most, if not all of this data is not a problem. However, finding "real" substantive information about the details of most arbitration proceedings is likely to prove to be difficult. Hearings are closed, and transcripts, if made, are confidential unless a party seeks to va-cate. [(66)] FINRA makes its awards available, unredacted, online, but doesn't reveal the supporting materials such as the exhibits and memorandums of law submitted by the parties or motion practice materials. Redacted awards issued by AAA Labor and Employment arbitrators are available at Lexis and Westlaw. Bloomberg Law makes available numerous international awards. Westlaw has a library of awards submitted by insurance arbitrators and a library containing international awards. Most awards that are available contain the name of the arbitrator as well as the result reached. When a party seeks to vacate an award, what is provided to the court is publicly available. Another source are the annual filings by any facilitator referred to in this article conducting business in California, Maryland, Maine, and the District of Columbia. While these filings are not complete, they do reveal the names of arbitrators, the number of cases heard, and the awards made.

In the past few years, courts throughout the U.S. have opened their files to the public at sites on the internet. While procedures in the courthouse may differ, these files should contain information relating to petitions to
P 69
confirm or vacate awards. Compiling a library with this information may prove expensive, but worthwhile.

As efforts are made to train computers using existing data sets, testing will no doubt reveal deficiencies needing to be addressed. Thinking through the details of any arbitration is arduous and time consuming for the parties and counsel. Training a computer with algorithms and data sets needed to address a broad range of conditions and situations is a far more complex task. Template style data sets describing in general terms the common factual elements of any number of disputes alone may not suffice. The training effort will probably require the development of algorithms designed to respond and provide structures addressing the peculiarities of any individual case.

Selling the idea of a machine-driven binding award may prove a daunting task. Along with transparency, the completeness of the data sets and the appropriateness of governing algorithms are likely to be two of the most controversial aspects of any program. Adding to the difficulties, the public will have to be educated about the use and meaning of any number of terms employing examples that are as non-technical, and yet as persuasive, as possible.

V. Question #4. What kind of cases would be best served by a computer acting as the arbitrator?

Today, the science involved with machine-driven final binding awards is at an embryonic stage. The types of disputes and legal issues being considered for arbitration are ever-growing and may prove to be without limit. Still, as a practical matter, not every dispute is a suitable candidate for A.I. There will always be a self-selecting process evidenced by the earlier suggestion that stakeholders in a "bet the company" case will more often than not reject the idea of machine-driven arbitration. But what about all the other disputes of great importance to the parties, even if the financial consequences aren't draconian? Will all or just some be candidates for arbitration using a computer? In all likelihood, most disputes will prove to be non-compatible. The more complex factually or at law a dispute becomes, the less likely the disputants and in particular counsel will accept a machine-driven process. Complex disputes usually require the absolute right to discovery in all allowable
P 70
forms. The same is true concerning the rules of evidence. These rights are not automatically available in an arbitration.

The paramount benefit attributed to arbitration is a desire to keep things as simple and streamlined as possible. The default position is dispensing with as many courthouse formalities as possible and bringing the case to a conclusion as quickly as possible. Given the preference for simplicity, the best candidates are likely small claims matters, defined as involving disputes (1) at law, excluding equitable matters of any kind, (2) with a dollar value of no more than a set amount, probably less than five figures, (3) involving simple factual and legal issues (4) and limited legal defenses, and (5) that can be easily classified as typical. For example, a suit on a promissory note is usually straightforward. The debtor either hasn't paid and has no excuse or hasn't paid and has an excuse that can be established with document evidence. Also included are claims involving limited property damage, failure to comply with a clearly defined obligation, breach of a contract claims involving the delivery of goods or services, and minor negligence matters. In Hadley's dispute, the underlying facts and law involve the basic elements of bailment and negligence. To prevail Hadley must, at a minimum, show:

•Ownership of the shirt, including proof of payment

•His care for the shirt prior to having it cleaned

•The value of the shirt prior to having it cleaned

•Proof that the shirt wasn't burned before it was handed over

•The actual condition of the shirt when handed over

•A description of the take-in procedures involved when the shirt was turned over

The dry cleaner, in order to defeat the claim, will have to show at a minimum:

•What the take-in procedures were when the shirt was presented
P 71

•What the custom and usage standard was at the time among dry cleaners in the community concerning take-in procedures

•What method was used to clean and press the shirt

•What type of equipment was used to clean and press the shirt

•The condition of the equipment on the day and time when the shirt was cleaned and pressed

•What fail safe mechanisms or procedures were in place to avoid damage to garments processed

•What the standard of care by dry cleaners was at the time of Hadley's transaction

•Identification of the individuals at the dry cleaner who were involved at the time Hadley's garment was taken in and subsequently processed

These elements are likely to be present in any dispute involving a claim of damage to property left with a dry cleaner.

Similarly, the duties imposed by law in this type of claim are elementary. The dry cleaner's duties are to inspect and take in the garment using procedures common to dry cleaners in the community, processing the shirt in a reasonable manner to avoid damage, and finally to ensure the return of the cleaned garment in substantially the same condition as when received. The standard of care is reasonableness under the circumstances.

In Hadley's case, the human arbitrator's first task would be to evaluate the information provided and ask questions designed to (a) ensure a complete picture of what actually happened to the shirt while in the possession of Hadley and then the dry cleaner, and (b) establish the credibility of witnesses and tangible evidence offered to establish or defeat the claim. An arbitrator might want to know:

•Before this claim was there a history of disputes between Hadley and this dry cleaner?

•Did either party have a history suggesting a penchant for litigation?
P 72

•Did anyone other than Hadley see the shirt and evaluate its condition before Hadley brought it to the dry cleaner?

•Why did a repairman examine the equipment on the day of the incident?

•Who, if anyone, other than the dry cleaner himself actually handled the shirt when it was cleaned and pressed?

To ensure the legal issues being decided and factual contradictions needing resolution meet the requirements for arbitration by computer the parties would have to submit a "package" of materials for review by the provider of the service. If determined to be insufficient or incomplete, the party involved would have to address the deficiency before the matter could be presented to the computer for determination.

Other types of disputes might also lend themselves. Insurance companies have created mechanisms for resolving disputes concerning distributing liability between carriers who have insured a risk. Many of these disputes are pro forma involving few complexities. These disputes, along with others of a similar nature arising in the business to business community, are candidates with the benefit being cost savings realized by the elimination or substantial reduction of the human factors in the dispute resolution process.

VI. Question #5. Will it be necessary to amend or modify the FAA or any of the state arbitration laws?

The short answer is "probably."

The FAA doesn't define the term "arbitrator." Nevada is the only state with an arbitration statute that defines the term. [(67)] Few courts have looked into the issue. The U.S. Supreme Court, as far back as 1868, opined "An arbitrator is defined as a private extraordinary judge chosen by the parties who have a matter in dispute, invested with power to decide the same." [(68)]
P 73
While the Court in 1868 wasn't aware of computer technology and algorithms, the reference to a "private extraordinary judge" is quite broad and arguably could include new technologies. The Supreme Court of Wisconsin accepted the definition offered by *Webster's Third New International Dictionary (1967):* "one with absolute power of deciding disputes so as to bind the disputants" [(69)] (Can "one" be interpreted to include a computer?) Most other state courts speaking about arbitrators as actors refer to them as being a "person." [(70)] Perhaps most important though, FAA Section 10 makes no reference to an arbitrator being a person. The described missteps sufficient to warrant vacating do not appear to provide a basis for concluding Section 10 would have no application just because the arbitrator is a computer. However, there is good reason to wonder if a misstep by a person who has trained the computer would come within the scope of Section 10. Given this question alone, it appears some legislative action might be needed to insure recognition of a computer as an arbitrator within the meaning of the Act.

None of the rules of the major arbitration administrators [(71)] address the issue of who can serve as an arbitrator. FINRA rules define both a public and a private arbitrator as "a person." [(72)] The Business-to-Business rules of the National Arbitration Forum define an arbitrator as:
P 74

> An individual selected in accord with the Code or an Arbitration Agreement to render Orders and Awards, including a sole Arbitrator and all Arbitrators of an arbitration panel. [(73)]

All administrator rules use the term "persons" or "individuals," when referring to an action taken by an arbitrator. All administrator rules allow parties, upon mutual agreement, to modify any rule. Therefore, it appears parties are free to agree to designate a computer as the arbitrator. To date no administrator has opined on whether or not it would administer a claim where parties have designated a computer to serve as the arbitrator. Modifying administrator rules might best wait until the question of the need for legislation has been resolved.

Conclusion

In the opening paragraphs of this article, the question "Where to begin?" was asked and subsequently answered in the context of feasibility. Here, using the same words, we assume feasibility and now focus on what steps are necessary to start the research ball rolling. The successful creation of a program allowing a computer to act as an arbitrator is closer to becoming a reality than many believe, the concerns raised in this article notwithstanding. The basic technologies needed are rapidly falling into place though they are only part of the equation. The complexities involved go well beyond designing the architecture, curating the data set, and deploying the technology. There are any number of matters that will need to be looked into. Issues such as security and privacy, hacking, the ethics of using a computer to issue a binding award, and legal liabilities involving the provider of such services, are but a few.

The continuing evolution that is producing new and better algorithms and tools for the assembling of data sets, provide hope for capturing and
P 75
resolving technical unknowns. [(74)] Training and testing aside, there are at least three (3) unknowns connected to human behavior:

1.People will need to be persuaded to accept the benefits of a machine-driven system that doesn't apply or even recognize deeply engrained perceptions about what our laws are and how our laws operate.

2.Computer scientists will need to fuse the differences between how humans and computers perceive dispute resolution.

3.And developers will have to grapple with the question: "Will human beings be willing to ever accept the judgment of a machine that knows nothing about gut feelings?"

With all the problems and challenges ahead, there is still no doubt that the effort will yield benefits reaching far beyond the present applications of arbitration as a means to resolving disputes. The well-known maxim "Justice delayed is justice denied," applies to localities where the local judicial system is clogged, corrupt, or otherwise non-responsive to the demands being made. Arbitration is a recognized alternative means for delivering a timely and efficient resolution of disputes no matter the size or complexity. Arbitration by A.I. has the potential to quickly move the benefits of arbitration substantially ahead. Governments at all levels are tasked with providing a judicial system. With few exceptions, there are no limitations on demands for access and yet resources are sparingly provided by legislatures resulting in clogged calendars and over-worked judges and supporting staff. The structures of the systems and protocols require humans to serve other humans thereby adding pressures flowing from the personal needs of all involved and leading to delays brought on by scheduling conflicts and even a failure to have available adequate court room facilities. Frustrated by delay, some stoop to corruption while others take the law into their own hands. No matter the basis for the ill, the respect for law is undermined.
P 76
Without doubt, there is room and an opportunity for any reasonable program with the potential to overcome these problems. Arbitration using A.I. presents such an opportunity. If successfully implemented, the need for additional court rooms, judges, and staff will be significantly reduced, as will associated costs. Scheduling conflicts will be reduced. Delay times will be reduced as computers can handle hundreds, if not thousands of cases every day, 365 days a year. Corruption can be contained because of the difficulties associated with directly corrupting a computer. While not perfect, the issued awards will provide resolution.

The starting point is recognizing the need for research. Developing the data sets, designing the algorithms, and deploying the models will be challenging, requiring an intense effort, and a financial commitment, though the investment required will be less than the ongoing impact of proceeding solely with human arbitrators. The only limitations are a lack of focus on the topic and the willingness of someone to make the needed financial commitment.
P 76

References

*)

Paul Bennett Marrow is an attorney/arbitrator and a member of the American Arbitra-tion Association's Commercial Panel. He teaches Domestic Arbitration at New York Law School. He can be reached at pbmarrow@optonline.net.

†)

Mansi Karol is the Director of ADR Serves, Commercial Division at the American Arbitration Association in New York. She oversees administration of the large, commercial complex caseload, user outreach, and the panel of commercial neutrals in New York. She canbereachedatKarolM@adr.org.

‡)

tSteven Kuyan is the Director of Entrepreneurship at New York University's Tandon School of Engineering and Managing Director of the NYU Tandon Future Labs. He's an advocate for the responsible adoption of AI, founder of NYCai, adjunct faculty at NYU, as well as an investor and advisor to numerous startups. He can be reached at kuyan@nyu.edu.

1)

No such data set currently exists and once development efforts get underway, due to the evolutionary pressures inherent to our systems of law, the results will continue to be a work in progress with no data set ever containing the full body of subject knowledge.

2)

Vincent C. Muller et al., *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, 376 Fundamental Issues of Artificial Intelligence 553-71 (2016).

3)

William S. Veatch, Artificial Intelligence and Legal Drafting, American Bar Ass'n Legal Analytics Committee Newsletter (Apr., 2019)

https://www.americanbar.org/groups/business_law/publications/committee_newsletters/legal_analytics/2....

4)

https://aaa-nynf.modria.com/.

5)

Michigan Courts, MI-Resolve,
https://courts.michigan.gov/Administration/SCAO/OfficesPrograms/ODR/Documents/contact/index.html.

6)

Franklin County Municipal Court, Online Dispute Resolution, www.courtinnovations.com/ohfcmc.

7)

Modria Resolution Center, New York No-Fault Insurance, https://aaa-nynf.modria.com/.

8)

Texas Judicial Branch, eFile Texas Status (Aug. 29, 2017), http://www.txcourts.gov/media/1438816/efiletexas-status-jcit-20170829.pdf.

9)

Utah Online Dispute Resolution Steering Comm., Utah Online Dispute Resolution Pilot Project 3-4 (2017), https://ncsc.contentdm.oclc.org/digital/api/collection/adr/id/63/download.

10)

Vivi Tan, *Online Dispute Resolution for Small Civil Claims in Victoria: A New Paradigm in Civil Justice*, 24 Deakin L. Rev. 101 (2019).

11)

Jan W. Vasbinder et al., *Artificial or Augmented Intelligence? The Ethical and Societal Implications*, *in Grand Challenges for Science in the 21st Century*, 51-68 (Jan W Vasbinder et al. eds. 2018).

12)

Omar-Rabinovich-Einy & Ethan Katsh, *Access to Digital Justice: Fair and Efficient Processes for the Modern Age*, 18 Cardozo J. Conflict Resol. 637 (2017).

13)

Regulation 2016/679, GDPR art. 12, 2016 O.J. (L 119) 39-40 (EU) (addressing "Transparent information, communications and modalities for the exercise of the rights of the data subject").

14)

Julia Poweles, *New York City's Bold, Flawed Attempt to Make Algorithms Accountable*, The New Yorker (Dec. 20, 2017), https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algor....

15)

New York City, Automated Decision Systems Task Force Report (Nov. 2019), https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.

16)

An arbitrator's merit-based decisions are for the most part beyond judicial review. This is true even if the arbitrator makes a good faith mistake. Judicial review is limited to the vacating of an award on limited grounds specified by applicable federal and state statutes. *Major League Baseball Players Ass'n v. Garvey,* 532 U.S. 504, 509 (2001) The rule is followed by most state courts. *See, e.g.*, *Heimlich v. Shivji*, 7 Cal. 5th 350, 367 ("A court's power to correct or vacate an erroneous arbitration award is closely circumscribed."); *In re Santer*, 23 N.Y. 3d 251, 263 (2014) Courts, on the other hand are subject to review based on errors.

17)

David Ha & Jurgen Schmidhuber, *World Models*, https://arxiv.org/abs/1803.10122 (2018).

18)

475 U.S. 643 (1986).

19)

Id. at 656 (bracketed words added). What constitutes "Clear and unmistakable" evidence of what the parties want is yet another issue, one that has been handled on a case by case basis.

20)

514 U.S. 938, 943 (1995).

21)

For purposes of this paper, "model" is sometimes used to describe an algorithm and at other times an assemblage of two or more algorithms working in tandem.

22)

The Association for Computing Machinery U.S. Public Policy Council defines an algorithm:

> An algorithm is a self-contained step-by-step set of operations that computers and other smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

The Association identifies seven principles for transparency and accountability: awareness, access and redress, accountability, explanation. data provenance. Auditability, validation, and testing. *See, generally*, Ass'n for Computing Machinery, Statement on Algorithmic Transparency and Accountability (Jan. 12, 2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.

23)

Harry Surden, *Machine Learning and the Law*, 89 Wash. L. Rev. 87, 94 (2014).

24)

Maxi Scherer, *Artificial Intelligence and Legal Decision-Making: The Wide Open? Study on the Example of International Arbitration*, SSRN.Com/abstract=3392669, at 8 and 36 J. Int'l Arb. 539 (2019).

25)

Amazon's recently launched the Alexa Prize competition (https://developer.amazon. com/alexaprize) requires competitors to create algorithms allowing the computer to have freeform conversations and be able to reference and converse about movies and recent news, all sourced from "live" data. The top performing teams achieved a maximum conversation lasting just over 8 minutes before the machine could no longer continue a "quality" conversation. *See* Jan Pichi et al., *Alquist: The Alexa Prize Socialbot*, https://arxiv.org/abs/1804.06705.

26)

While training an algorithm, successful determinations are stored in the computer's memory along with the information that led to each success.

27)

Caryn Devins et al., *The Lawand Big Data*, 27 Cornell J.L. & Pub. Pol'y 357 (2017).

28)

Judge Edward Weinfeld famously observed: "An arbitration tribunal is not a court of record; its rules of evidence and procedures differ from those of courts of record; its fact-finding process is not equivalent to judicial fact finding." *Williamson, P.A. v. John D. Quinn Constr. Corp.*, 537 F. Supp. 613, 616 (S.D.N.Y. 1982).

29)

For a comprehensive list, see American Arbitration Association, Commercial Arbitration Rules and Mediation Procedures, Preliminary Hearing Procedures P-2 (2013).

30)

Rule R-1(a) provides, in part: "The parties, by written agreement, may vary the procedures set forth in these rules." American Arbitration Association, Commercial Arbitration Rules and Mediation Procedures (2013).

31)

9 U.S.C. §§ 1-16.

32)

AT&T Mobility v. Concepcion, 563 U.S. 333 (2012).

33)

Asarco, LLC v. United Steel, Paper, Forestry, Rubber, Mfg., Energy, 910 F.3d 485, 491 (9th Cir. 2018) involved a collective bargaining agreement with a "no add provision" restricting the ability of the arbitrator to "add to, detract from or alter in any way the provisions" of the agreement. After a hearing, the arbitrator found there was a mutual mistake concerning an amendment resulting in a dispute over the eligibility of certain persons to benefit from the agreement. Citing the mutual mistake, the arbitrator reformed the agreement to provide for the inclusion of the persons who might have otherwise been excluded. The court found that, because the arbitrator based his actions on an interpretation of contract law, the court lacked the ability to overrule what it found was an honest interpretation of contract law.
This case presents the question: What would a computer do if called upon to act as the arbitrator? The complexity of the arbitrator and the court's analysis suggests the computer may well be in over its head. Thus the need for an algorithm designed to have the computer signal the need for human intervention.

34)

Section 10(a) provides: (a) In any of the following cases the United States court in and for the district wherein the award was made may make an order vacating the award upon the application of any party to the arbitration:
(1)where the award was procured by corruption, fraud, or undue means;

(2)where there was evident partiality or corruption in the arbitrators, or either of them;

(3)where the arbitrators were guilty of misconduct in refusing to postpone the hearing, upon sufficient cause shown, or in refusing to hear evidence pertinent and material to the controversy; or of any other misbehavior by which the rights of any party have been prejudiced; or

(4)where the arbitrators exceeded their powers, or so imperfectly executed them that a mutual, final, and definite award upon the subject matter submitted was not made.

35)

State arbitration statutes all provide conditions for vacating an award. Most follow the federal statute but there are exceptions. For example, the New York C.P.L.R. allows vacating upon a showing that the arbitrator failed to follow the procedure set forth in Article 75. *See* N.Y. C.P.L.R. 7511(b)(1)(iv). In addition, a finding by an arbitrator that the claim was barred by a statute of limitations is grounds to vacate. *See* N.Y. C.P.L.R. 7511(b)(2)(iv).

36)

CM South East Texas Houston, LLC v. CareMinders Home Care, Inc., 662 F App'x 701, 705 (11th Cir 2016), *Employers Ins. Of Wausau v. National Union Fire Ins. Co.*, 933 F.2d 1481, 1491 (9th Cir 1991), *Hoffman v. Cargill*, 988 F. Supp. 465, 474 (N.D. Iowa 1997).

37)

Andrea Roth, *Trial by Machine*, 104 Geo. L.J. 1245 (2016). See, in particular, the "man behind the curtain" discussion at pp. 1277-80.

38)

Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. Davis L. Rev. 399, 414 (2017). "No one knows why the system selects a document: once the system is trained, no script can be provided to a human sorter to imitate the system's selection of documents. That is, there is no way to accurately summarize the criteria used. Nevertheless, parties rely on predictive coding in very high-stakes litigation. It is treated as reliable." Curtis Karnow, *The Opinion of Machines*, 19 Colum. Sci. & Tech. L. Rev. 136, 142 (2017). "In general, deep learning models (that is, deep artificial neural networks) are often criticized as of being "black-box" models, whose answers, despite being remarkably accurate, are hard to interpret. There is a major need, in the field of AI, to build explainable models, i.e., models capable of motivating their choices, that is models whose decision processes can be interpreted by a human. The direction in which the field is moving is that of integrating so-called sub-symbolic (or connects) approaches, such as artificial neural networks, with so-called symbolic methodologies, which are built on logic. The former are capable of efficiently and effectively dealing with uncertainty in data and can easily exploit very large data collections, but lack in interpretability. The latter, on the other hand, are designed to deal with knowledge representation and reasoning, and thus show a high expressivity, a high interpretability, but cannot easily handle noisy information and scale to big data. There is a strong belief within the AI community that the combination of such diverse approaches is a necessary step to fill the performance gap in tasks related to reasoning." Przemyslaw Palka & Marco Lippi, *Big Data Analytics, Online Terms of Service and Privacy Policies*, https://ssrn.com/abstract=3347364, at 21-22 (internal citations omitted); *see also* Lisa Getoor & Ben Taskar, *Statistical Relational Learning* (2007); Artur d'Avila Garcez et al., *Neural-Symbolic Learning and Reasoning: Contributions and Challenges*, 2015 AAAI Spring Symposium Series.

39)

Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

40)

Paul Bennett Marrow, *Counsel, Beware*, 81 NYSBA J. 36 (2009).

41)

Chris Guthrie, *Misjudging,* 7 Nev. L. Rev. 420 (2007).

42)

See Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 Cognitive Psychol. 430 (1972) [hereinafter "*Subjective Probability*"]; Daniel Kahneman & Amos Tversky, *Availability: A Heuristic for Judging Frequency and Probability*, 5 Cognitive Psychol. 207 (1973); Daniel Kahneman & Amos Tversky, *Judgment Under Uncertainty: Heuristics and Biases*, 185 Science 1124 (1974).

43)

Subjective Probability, *supra* note 42, at 436-37

44)

Chris Guthrie et al., *Inside the Judicial Mind*, 86 Cornell L. Rev. 777, 792-94 (2001) ("The potentially pernicious effects of anchoring also suggest a source of error in both the civil and criminal justice systems. In civil cases, the influence on judges of misleading anchors, such as litigants' for damage awards, can produce biased damage awards.").

45)

Shai Danziger et al., *Extraneous Factors in Judicial Decisions*, 108 Proc. Nat'l Acad. Sci. 6889 (2011), https://www.pnas.org/content/108/17/6889.

46)

What is Garbage In, Garbage Out (GIGO)?, https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo.

47)

Scherer, *supra* note 24, at 21. Hidden bias is not the only concern. Random errors caused by mistyping and mistranslation are examples of other problems. Andrea Roth, *supra* note 37, at 1275.

48)

Scherer, *supra* note 24, at 559-61.

49)

Jennifer Valentino-DeVries, *Websites Vary Prices, Deals Based on Users' Information*, Wall St. J. (Dec. 24, 2012).

50)

See Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 Wash. L. Rev. 579, 591-92 (2018) (discussing a computer being trained to identify a cat).

51)

Hannah Fry, *Hello World: Being Human in the Age of Algorithms*, 68 (2018).

52)

For an in-depth discussion, see Harry Surden, *The Ethics of Artificial Intelligence in Law: Basic Questions*, (Aug. 26, 2019), www.ssrn.com/abstract=3441303.

53)

Levendowski, *supra* note 50, at 586 nn. 26-27 (2018).

54)

Robert Brauneis & Ellen Goodman, *Algorithmic Transparency for the Smart City*, 20 Yale J.L. & Tech. 103, 117-20 (2018).

55)

Wojciech Samek et al., *Explainable Artificial Intelligence: Understanding, Visualizing, and Interpreting Deep Learning Models* (Aug. 28, 2017), https://arxiv.org/abs/1708.08296.

56)

Adey Zegeye, Design/Ethical Implications of Explainable AI (XAI) (May 7, 2019), https://blogs.commons.georgetown.edu/cctp-607-spring2019/2019/05/07/design-ethical-implications-of-e....

57)

Brent Mittelstadt et al., *Explaining Explanations in AI* (Nov 4, 2018), https://arxiv.org/abs/1811.01439.

58)

"[S]ome computer models (such as expert models) are indeed rule-based, using causal logic and deductive reasoning, since they apply pre-established rules in the algorithm to the observable data. Other AI models, however, have different features. In particular, machine learning models, such as neural networks, often have no pre-defined rules. Deductive, causal reasoning is thus replaced by an inverse approach, because the machine learning program extracts the algorithm from the observable data. Rather than using logic, the AI model calculates probabilities, i.e. the likelihood for any given outcome. [¶] Applying such machine learning processes in the legal decision-making context therefore would mean accepting a departure from the above-mentioned understanding of judicial reasoning according to formalist theories. A decision based on those AI models would not be based on pre-determined legal rules, would not be the result of deductive logic, and would not follow the above-described legal syllogism. While this situation would be a cause for concern for legal formalists, it might be seen as vindicating others who have long criticized formalist theories." Scherer, *supra*note 24, at 567.

59)

For an in-depth discussion, see *id.* at 562-72.

60)

Id. at 557, worries that the reliance on data composed of previous decisions, there is a risk that computer results will be "conservative" and not be adapted to policy changes over time: "Velocity refers to the frequency of incoming data that needs to be processed. Big Data is often challenging because of the sheer amount and high frequency of the incoming data. In the legal context, such risk is very low. As already pointed out above, in terms of volume, the problem is likely to be of scarcity rather than abundance of data. Therefore, over time, decisions might not be frequent, and when they occur there might have been a change in policy so that the previous data is outdated. These policy changes can be radical and swift at times. To take an example from the international arbitration context, the decision of the Court of Justice of the European Union in Achmea has fundamentally changed the compatibility of investor-state arbitration with European law overnight. This raises the question how AI models which, by definition, are based on information extracted from previous data may deal with those policy changes. It is true that the essence of machine learning is the ability to improve the algorithm over time. Nevertheless, such improvement is always based on past data. Policy changes in case law necessarily require departures from past data, i.e. previous cases. For these reasons, AI models are likely to keep conservative' approaches that are in line with previous cases." Scherer's concern disregards the reality that an arbitrator is rarely, if ever, endowed with the authority needed to fashion a rule needed to address evolving policy considerations.

61)

Samek et al., *supra*note 55; Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, Harvard Law School, Public Law & Legal Theory, Research Paper Series, #18-07 (Nov. 6, 2017), https://ssrn.com/abstract=3064761.

62)

For an example where this type of team approach has had positive results, see Katyanna Quach, *Don't try and beat AI, merge with it says chess champ Gary Kasparov*, The Register (May 10, 2018), https://www.theregister.co.uk/2018/05/10/heres_what_garry_kasparov_an_old_world_chess_champion_thin....

63)

Caryn Devins et al., *The Law and Big Data*, 27 Cornell J.L. & Pub. Pol'y 357, 359-60 (2017).

64)

Scherer, *supra*note 24, at 15 ("Any data-driven AI programs first and foremost require access to data. Machine learning models, which are based on probabilistic inferences, are data hungry: the larger the sample data, the more accurate the model's predictive value.").

65)

Elite Data Science, Overfitting in Machine Learning: What It Is and How to Prevent It (2019), https://elitedatascience.com/overfitting-in-machine-learning.

66)

Confirmation is the process whereby an award is converted into a judgment. Confirmation typically doesn't require the record of the hearing. An application to vacate does require a fairly comprehensive description to the court of the issues and facts formulating the basis for the demand. *See, e.g.*, 9 U.S.C. § 13.

67)

Nev. Rev. Stat. § 38.209 ("'Arbitrator' means an individual appointed to render an award, alone or with others, in a controversy that is subject to an agreement to arbitrate").

68)

Gordon v. United States, 74 U.S. 188, 194 (quoting *Bouvier's Law Dictionary*, title "Arbitrator.").

69)

Grays Harbor County v. Williamson, 96 Wn. 2d 147, 156 (1981).

70)

State ex rel. Cushion v. City of Massillon, 2011 Ohio App. LEXIS 3922; *Konz v. Morgan Stanley Smith Barney,* 2018 U.S. Dist. LEXIS 180069*; Bolick v. Merrill Lynch, Pierce, Fenner & Smith, Inc,* 2006 U.S. Dist. LEXIS 330; *Nieves v. Travelers Cas. Ins. Co. of Am*., 2015 U.S. Dist. LEXIS 95774; *Society of Am. Foresters v. Renewable Natural Resources Found*., 114 Md. App. 224 (1997); *State ex rel. Cushion v. City of Massillon;* 2011 Ohio App. LEXIS 2457*; Hino Motors Mfg. United States v. Naftaly*, 2011 Mich. App. LEXIS 1562.

71)

For purposes of this paper, these include the AAA, JAMS, CPR, NAF, and FINRA.

72)

See, e.g., FINRA Arbitration Rules 12100 (r) and (y). JAMS Comprehensive Arbitration Rules and Procedures, Rule 7 (a) states: "In these Rules, the term 'Arbitrator' shall mean, as the context requires, the Arbitrator or the panel of Arbitrators in a tripartite Arbitration."

73)

Forum, Code of Procedures for Resolving Business-to-Business Disputes, Appendix "A", https://www.adrforum.com/assets/resources/Arbitration/Rules/Forum.B2B_ Rules.v2.3.pdf.

74)

Himabindu Lakkaraju et al., *Identifying Unknown Unknowns in the Open World: Representations and Policies for Guided Exploration* (Oct. 28, 2016), https://arxiv.org/abs/1610.09064.

Document information

Publication

- Journal of International Arbitration

Bibliographic reference

Maxi Scherer, 'Artificial Intelligence and Legal Decision-Making: The Wide Open?', in Maxi Scherer (ed), Journal of International Arbitration, (© Kluwer Law International; Kluwer Law International 2019, Volume 36 Issue 5) pp. 539 - 574

Artificial Intelligence and Legal Decision-Making: The Wide Open?

Maxi Scherer

(*)

The article explores the use of Artificial Intelligence (AI) in arbitral or judicial decision-making from a holistic point of view, exploring the technical aspects of AI, its practical limitations as well as its methodological and theoretical implications for decision-making as a whole. While this article takes the angle of international arbitration, it looks at examples and studies from a wide variety of legal areas and its conclusions are relevant for adjudicatory decision-making more globally. The author assesses existing studies on decision outcome prediction and concludes that the methodology and assumptions employed put into doubt the claim these models might be used for ex ante outcome predictions. The article also discusses whether AI models, which are typically based on information extracted from previous input data, are likely to follow 'conservative' approaches and might not be adapted to deal with important policy changes over time. The article further finds that a blind deferential attitude towards algorithmic objectivity and infallibility is misplaced and that AI models might perpetuate existing biases. It discusses the need for reasoned decisions, which is likely to be an important barrier for AI-based legal decision-making. Finally, looking at existing legal theories on judicial decision-making, the article concludes that the use of AI and its reliance on probabilistic inferences could constitute a significant paradigm shift. In the view of the author, AI will no doubt fundamentally affect the legal profession, including judicial decision-making, but its implications need to be considered carefully.

1 INTRODUCTION

> L'avenir n'est jamais que du présent à mettre en ordre. Tu n'as pas à le prévoir, mais à le permettre. *Antoine de Saint-Exupéry*

The relationship of the future to the present is the topic of de Saint-Exupéry's somewhat mysterious quote. The first sentence states that the future simply is the present in better order or better organized. Concerning the future, the second sentence goes on, the task is not to foresee it, but to allow or enable it. How better
P 540
to enable a more organized future than with the use of technology, such as Artificial Intelligence (AI)? It is trite to underline the importance AI already has in our daily lives. Whether we are aware of it or not, AI is used to filter spam emails, write newspaper articles, provide medical diagnoses, and assess access to credits.

Nevertheless, lawyers typically believe that the impact on their profession will be limited. This ignores that AI already touches many areas of law, including contract analysis, legal research, e-discovery, etc. (1) For instance, computer programs are available to help lawyers to analyse the other side's written submissions and to provide relevant case law that was omitted therein or rendered since. Unsurprisingly, AI in law is a growing business. (2)

In international arbitration, the use of AI has been predicted for a wide variety of tasks, including appointment of arbitrators, legal research, drafting and proofreading of written submissions, translation of documents, case management and document organization, cost estimations, hearing arrangements (such as transcripts or simultaneous foreign language interpretation), and drafting of standard sections of awards (such as procedural history). (3)

This article will not deal with those aspects but instead focus on one of the more controversial areas which is at the core of the arbitral process: the decision-making itself. (4)
P 541
It will explore whether and how AI can be used to help or potentially replace arbitrators in their task to decide the dispute. Importantly, the subject of this article differs from discussions about online arbitration, which generally refers to proceedings for which processes are streamlined thanks to the use of technology, such as electronic filings, but where human arbitrators remain the decision-makers. (5) Also, while this article focusses on arbitral decision-making, it uses examples and studies from a wide variety of legal areas and its conclusions are relevant for judicial decision-making more globally, not only in international arbitration.

When considering AI for arbitral decision-making, some have speculated about the feasibility of 'robot-arbitrators', [(6)] but little research has gone into the potential implications of the use of AI in this area. Authors typically either assert that AI is inevitable in the future, [(7)] or express scepticism, mainly on the assumption that some 'human factor' would be necessary to ensure empathy and emotional justice. [(8)] This article seeks to explore the topic in a more in-depth fashion, assessing the technical aspects of AI and its implications and limitations, as well as addressing the more fundamental impact it may have on human decision-making processes and theories thereof.

Section 2 defines AI and describes its most important features. A good understanding of the technical aspects of AI is necessary to fully assess its implications for legal decision-making. Section 3 analyses existing studies on the use of AI to predict the outcome of legal decisions. It evaluates their method and results, questioning the extent to which those studies point towards a general applicability of AI for *ex ante* outcome prediction. Section 4 considers the inherent limitations of AI models used, based on the so-called four Vs of Big Data – Volume, Variety, Velocity, and Veracity – and examines their consequences for legal decision-making. In particular, this section discusses the need for sufficient non-confidential case data, the requirement of repetitive fact-patterns and binary outcomes, the problem of policy changes over time, and the risks of bias and data diet vulnerability. Section 5 highlights one
P 542
major draw-back of AI decision-making: the difficulty with providing reasoned legal decisions obtained by AI. Section 6 analyses the changes AI decisions would bring for legal theories of judicial decision-making. It shows that AI would change the normative basis for decision-making and thus constitute a significant paradigm shift from a theoretical point of view. Section 7 sets out the conclusions and the main findings of this article.

2 FEATURES OF ARTIFICIAL INTELLIGENCE MODELS

Lawyers often lack basic understanding of artificial intelligence. [(9)] AI-savvy lawyers are said to be as rare as vegan butchers. [(10)] Without becoming computer-scientists, it is important for lawyers to understand the basic features of artificial intelligence. Only with a good understanding of AI is it possible to assess its potential implications on the legal profession and legal thinking. The aim of this section is therefore to provide some important technical background information on AI.

Artificial intelligence can be defined as 'making a machine behave in ways that would be called intelligent if a human were so behaving'. [(11)] This was indeed the definition proposed by John McCarthy, a late computer scientist and arguably the one who coined the term 'AI' in 1956. Other similar definitions exist. For instance, the *Oxford Dictionary* defines artificial intelligence as the '[t]heory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages'. [(12)]

These definitions show human intelligence as a bench-mark for AI. The term 'intelligence' in itself is not straightforward to define and has caused philosophers, psychologists, cognitive scientists, and other researchers to disagree. [(13)] At a basic
P 543
level, one can describe intelligence as 'the ability to learn, understand, and make judgments or have opinions that are based on reason'. [(14)] This ability distinguishes human beings from other forms of non-intelligent or less intelligent life. [(15)]

In early stages of AI-research, computer scientists tried to develop programs that mimic human intelligence by seeking to understand human cognitive processes and replicate them. [(16)] For instance, computer scientists tried to understand the processes involved in learning a language and thus develop an algorithm – a sequence of precise instructions – that would enable computers to learn a language. Results were poor, particularly with complex tasks, such as language-learning. [(17)]

To a lesser extent, similar models are still used today. They are called expert systems or rules-based programs. [(18)] These systems are based on a set of rules, generally in the form of 'if-then' instructions (e.g. if the light turns red, then stop), also called the knowledge base. They make use of logical inferences, based on the rules contained in the knowledge base. There are several reasons those programs are not as powerful as other models further described below. Most importantly, they are laborious because the knowledge base needs to be created manually by defining the rules and coding the program accordingly. [(19)] Moreover, the use of *ex ante* rules, such as 'if/then' principles, are often unsuitable to describe accurately complex and dynamic realities. [(20)]

Different models were thus developed. The quantum leap in AI-research occurred with the so-called 'dataquake', the emergence of huge amounts of data. [(21)] This surge of data was due to the

combination of increased computer processor speed (which is said to double every twelve-eighteen months according to the so-called Moore's Law) [(22)] and decreased data storage costs (which is said to follow a
P 544
similar pace according to the so-called Kryder's Law). [(23)] The emergence of 'Big Data' allowed a significant shift in the development of artificial intelligence. Rather than developing complex algorithms for cognitive processes, AI is being used to 'learn' from existing data.

Machine learning refers to a subfield of AI-research concerned with computer programs that learn from experience and improve their performance over time. [(24)] The reference to 'learning' does not refer to cognitive processes thought to be involved in human learning; rather it refers to the functional sense of learning: the ability to change behaviour through experience over time. [(25)] The process of machine learning has achieved surprising results in many areas. [(26)] To continue with the previous example of language-learning, computer translation programs are remarkably accurate these days. Contrary to the earlier attempts described above, no programmer needs to code an algorithm for translation; rather, computer models, such as neural networks, use massive amounts of available data to 'learn' the relevant features and continually improve with immediate online feed-back through user clicks. Boden notes that 'many networks have the uncanny property of self-organization from a random start'. [(27)]

Machine learning at its core relies on the inference of hidden factors or patterns from observed data. [(28)] Using large amounts of sample data and with sufficient computing power, the computer extracts the necessary algorithms, rather than those algorithms being coded into the machine. In many areas, defining the algorithm in the form of precise *ex ante* instructions proves difficult. [(29)] For instance, humans might easily recognize which email is spam, but cannot provide precise and exhaustive instructions for this classification task. However, if the program is given a large set of sample data in which emails are labelled as 'spam' or 'not spam', the program will be able to detect the necessary classification algorithm. It does so by recognizing repeat patterns for spam emails and infers that future emails with the same features should also be classified as spam.

The search for hidden patterns is illustrated by the term 'data mining'. The analogy is that one has to work through tons of earth from the mine to find precious material. [(30)] In the AI context, the program weeds through large amounts of data with the aim to find an accurate model. Once the hidden model is
P 545
detected, this can be used to predict future cases (e.g. classify a future email as spam or not), which is of particular importance in the legal context, as further discussed below. [(31)]

The ability of pattern-recognition relies on statistics and probability calculations. [(32)] In simple terms, the computer program calculates, for each factor or combination of factors it observes, the probability to lead to a certain outcome. For instance, if the words 'sex' and 'Viagra' are in an email, the probability for it to be spam is high. Probabilistic theories, such as Bayesian networks, are the source of success of machine learning AI. [(33)] The learning programs resemble a general template with modifiable parameters, with the aim to adapt the parameters of the model on the basis of the information extracted from the sample data. As Alpaydin puts it, in AI '[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 consequence, AI models are able to produce 'intelligent' outcomes which, if performed by humans, are thought to involve high-level cognitive processes (e.g. understanding emails in order to classify them as spam). [(35)] However, this result is achieved without anything that resembles 'intelligent' human-cognitive processes but is merely based on probabilistic models. As one author describes 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 implications for legal decision-making that arise as a result of this shift from early models that focus on human-like processes, to statistical or probabilistic models that achieve human-like results without 'intelligent' processes, is discussed in greater detail below.

AI-researchers distinguish several types of machine learning, depending on the degree of human input. Supervised learning requires human interaction: the programmer trains the program by defining a set of desired outcomes (e.g. classification into spam/no-spam) for a range of input. [(37)] This means that the data of the training set must be adequately labelled (e.g. emails identified as spam or not)
P 546
and some form of human feed-back is required (e.g. when the program wrongly classifies an email). To the contrary, unsupervised learning requires no, or virtually no, human interference. There are no pre-established assumptions or pre-defined outputs; rather, the program detects co-occurring

features which will engender the expectation that they will co-occur in the future. (38) This is the case, for instance, with many modern language translation programs discussed above.

Importantly, there is not one single AI system, but a variety of different models. (39) For the purpose of the current study, the differences between the two approaches described above are important. On the one hand, expert models are *rule-based* and use *logic* as the normative principle. They may also be described as using a *forward* approach, because they apply pre-established rules to the observable data. The method is *causal*, deducing the outcome from the pre-established, fixed rules coded in the algorithm. On the other hand, machine learning models, such as neural networks, have often no pre-defined rules but use *pattern-recognition* and are built on *probabilistic methods* as the normative principle. They may also be described as using an *inverse* approach, because they extract the algorithm from observable data. The method is *predictive*, calculating the likelihood for any given outcome based on the extracted, and steadily improving, algorithm.

3 LEGAL DECISION-MAKING AND AI: THE USE OF QUANTITATIVE PREDICTION

The idea that AI-driven programs could predict the outcome of legal decision-making seems counter-intuitive to most lawyers. Lawyers instinctively believe that legal decision-making requires cognitive processes – such as understanding the parties' legal submissions and determining the right outcome through reasoning – which cannot be achieved by computer programs. However, as discussed in the previous section, computer models are able to achieve 'intelligent' results, which, if performed by humans, are believed to require high-level cognitive processes.

Several studies may lend support to the thesis that computer programs are better than humans in predicting the outcome of legal decision-making. (40) For instance, an early study showed that computer programs excelled over human
P 547
experts in predicting the votes of individual US Supreme Court justices in upcoming decisions for the 2002 term. The computer model achieved a correct prediction rate of 75%, whereas the human expert group, composed of eminent lawyers and law professors, correctly guessed only 59.1% of votes. (41)

The basic explanation for this – apparently triumphant – AI-success is that human brains suffer 'hardware' limitations which computer programs surpass easily. (42) In coming years, it is expected that computers available at the consumer level will reach storage capacity of several petabytes. Fifty petabytes are sufficient to store the information content of the 'entire written works of mankind from the beginning of recorded history in all languages'. (43) Accordingly, computers can simply stock amounts of data and draw from that data – or experience – much more quickly and efficiently than humans ever will. (44)

This section discusses two recent studies on the prediction of legal decision-making, looking at their methodology and results. Section 3.1 analyses a study conducted in 2016, which relates to decisions of the European Court of Human Rights, and section 3.2 looks at a study from 2017 predicting US Supreme Court decisions.

3.1 Predicting decisions of the European Court of Human Rights

The study conducted by a group of researchers in 2016 (45) focussed on decisions by the European Court of Human Rights (hereafter the 'ECtHR') rendered in the English language about three provisions of the European Convention on Human Rights (hereafter the 'Convention'), (46) 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. Those provisions were chosen because they provided the highest number of decisions under the Convention and thus sufficient data on
P 548
which to base a study. (47) For each of those provisions, the study selected an equal number of decisions in which the ECtHR found a violation and in which it found none. This resulted in a total dataset of 584 decisions: 250 for Article 3, 80 for Article 6, and 254 for Article 8. (48)

The methodology used in the study focussed on the textual information contained in the decisions, using natural language processing and machine learning. (49) The study input was the text found in the decisions, following the usual structure of decisions of the ECtHR including sections on the procedure, factual background, and legal arguments. (50) Not included in the input were the operative sections of the decisions where the Court announces the outcome of the case. (51) The output target was a binary classification task as to whether or not the ECtHR found a violation of the underlying provision of the Convention. (52) The model was trained and tested on a 10% subset of the dataset. (53)

As a result, the model obtained an overall accuracy to predict the outcome of the Court's decision in 79% of all cases. (54) The decision sections with the best predictive value were those setting out the factual circumstances and procedural background (76% and 73%, respectively), whereas the legal reasoning section had a lesser outcome prediction value (62%). (55) The study also set out the most frequently used words for various topics, indicating their relative predictive weight for a violation or non-violation. For instance, the most frequently used words with a high prediction value included under Article 3 of the Convention: 'injury', 'damage', 'Ukraine', 'course', 'region', 'effective', 'prison', 'well', 'ill treatment', 'force', and 'beaten' (56) ; under Article 6 of the Convention: 'appeal', 'execution', 'limit', 'copy', 'employee', 'January', and 'fine' (57) ; and under Article 8 of the Convention: 'son', 'body', 'result', 'Russian', 'department', 'attack', and 'died'. (58)

The authors of the study claim that their work may lead the way to predicting *ex ante* the outcome of future ECtHR cases. They state that:

> [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 therefore
> P 549
> stand as a (crude) proxy for, applications lodged with the Court as well as for briefs submitted by parties in pending cases. (59)

The authors further see in the above-mentioned results a confirmation of legal realist theories according to which judges are primarily responsive to non-legal, rather than to legal, reasons when deciding cases. (60) They conclude that 'the information regarding the factual background of the case as this is formulated by the Court in the relevant subsections of its judgment is the most important part 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) The conclusion on the validation of legal realists' theories will be discussed in detail in section 7 below. This section provides some comments on the methodology and results, as well as the claim that the study leads the way to *ex ante* outcome prediction.

First, it remains somewhat unclear which parts of the ECtHR decisions were included in the study's input. As indicated above, the operative part of the decision in which the Court announces the outcome of the case, is obviously not included, (62) otherwise the prediction-task would be moot. Less clear is whether the part of the legal section containing the Court's reasoning is included or not. The study indicates that the aim was to 'ensure that the models do not use information pertaining to the outcome of the case' but this caveat seems to apply only to the operative sections of the decisions. (63) The law section is said to be included (64) and this typically includes the Court's legal reasoning, as indicated in the study. (65)

If the Court's legal reasoning is indeed included in the data input, the study's overall prediction results are all but surprising. Any trained lawyer – and probably most non-lawyers – would be able to guess, in virtually 100% of the cases, the outcome as to whether the Court finds a violation or not, after having been given the Court's reasoning. The study's overall prediction rate of 79% is therefore to be interpreted in this context. Moreover, the inclusion of the Court's legal reasoning significantly undermines the study's claim to lead the way towards possible *ex ante* outcome prediction. The Court's reasoning is precisely not available *ex ante* and therefore cannot be included in the prediction of future cases.
P 550

Second, one may query whether the factual background part in the Court's decision does not already contain 'hints' concerning the decision's outcome. The study acknowledges the 'possibility that the formulation by the Court may be tailor-made to fit a specific preferred outcome'. (66) Without suggesting any form of bias or lack of neutrality on the part of the ECtHR judges, the facts described in the judgment may be a selection of those facts that will be relevant for the decision's legal reasoning and outcome, leaving aside other non-pertinent facts pleaded by the parties. Therefore, one may express doubts as to the study's assumption that 'the text extracted from published judgments of the Court bears a sufficient number of similarities with, and can therefore stand as a (crude) proxy for, applications lodged with the Court as well as for briefs submitted by parties in pending cases'. (67)

Third, the most frequently used words for various topics with a high prediction value set out in the study would have to be used in any *ex ante* prediction model. This seems problematic for a number of reasons. Some of the words – such as ‘result’, ‘employee’, ‘region’, ‘copy’, or ‘department’ – seem random and it is hard to see how they would be able to predict *ex ante* the outcome of future cases. Others are very case-specific and would be problematic if used for future predictions, including words such as ‘Ukraine’, ‘January’, or ‘Russian’. Using these words for future outcome prediction might lead to facts relating to those countries or dates being determinative on the outcome. Implications of possible text-based prediction tools are further discussed below. (68)

Overall, while the result of the study, obtaining 79% accuracy to predict the outcome of the ECtHR decisions, seems impressive at first sight, a closer analysis of the methodology and assumptions employed puts into doubt the claims for possible *ex ante* outcome predictions.

3.2 Predicting decisions of the US Supreme Court

Another group of researchers focussed on the prediction of US Supreme Court decisions and published their final results in 2017. (69) Their study drew from previous work on US Supreme Court predictions, (70) but was innovative in several

P 551

aspects. First, the study’s goal was to obtain a model that would generally and consistently be applicable to all US Supreme Court decisions over time, not only in a given year or for a given composition of the Court with justices. (71) Second, the study also applied the principle that ‘all information required for the model to produce an estimate should be knowable prior to the date of the decision’. (72) As has been discussed in the previous section, this is to ensure that the model can be used for *ex ante* outcome prediction.

In order to achieve these aims, the study input included US Supreme Court decisions from almost two centuries, from 1816 to 2015. This resulted in input data of more than 28,000 case outcomes and more than 240,000 individual justices’ votes. (73) Rather than relying on the textual information contained in the decisions themselves, as was the case for the ECtHR study, this study labelled the data relating to each decision, using certain features. (74) First, some features relate to the specific case at hand, such as the identity of the parties, the issues at stake or the timing of the decision to be rendered. Second, other features draw information from the lower court’s decision which is to be examined. This includes, among others, the identity of the courts of origin (i.e. which circuit), the lower court’s disposition and directions, as well as which lower courts are in disagreement over the issue at stake. Third, another category of features focusses on the Supreme Court’s composition, such as the identity of the justices, and their previous rate of reversal votes or dissents, as well as their political preferences. Fourth, a final set of features relates to the procedure before the US Supreme Court, such as the manner in which the Court took jurisdiction and the reasons for granting certiorari, (75) whether or not an oral argument was scheduled and, if so, the time between the argument and the decision.

P 552

The study output target was two-fold: predicting the outcome of the decisions and predicting each justice’s votes. (76) For the outcome of the decisions, the classification task was binary, as to whether the Supreme Court reversed or affirmed the lower court’s decision. (77) There are some (albeit few) cases in which the Supreme Court does not review a lower court’s decision, but rather decides a dispute as the original court of jurisdiction. (78) The study excluded those cases from the decision outcome prediction because they do not fall into a binary classification task. (79)

Using machine learning, the researchers trained the model on a sample from the dataset, and then applied the obtained model to the remaining, out-of sample, data. (80) Overall, the model predicted the votes of individual justices with 71.9% accuracy, and the outcome of the decisions with 70.2% accuracy. (81) While there was fluctuation in any given year or decade, the study claims that the model delivered ‘stable performance’ over time. (82) The study also claims that the model ‘significantly outperforms’ possible baseline comparison models. (83)

Testing the study’s methodology and results against its aim to provide a general model for *ex ante* outcome prediction, the study contains some important limitations.

First, while the study applies the principle that ‘all information required for the model to produce an estimate should be knowable prior to the date of the decision’, (84) some of the input data features are available only shortly before the decision is rendered. For instance, whether or not an oral argument is scheduled and, if so, the time between the argument and the decision, is information typically available only at a late stage of the proceedings. (85) This significantly limits the use of those features for *ex ante* outcome prediction.

Second, a majority of the input-data labels are specific to appellate or Supreme courts tasked with the review of lower courts’ decisions. As detailed above, many features used in the study are related to the lower court’s decision to be examined

P 553

(e.g. which circuit, the lower court's disposition and directions) as well as the Supreme Court justice's handling of previous decisions from lower courts (e.g. reversal rates). Few of the input features are original to the dispute, such as the identity of the parties, the issues at stake or procedural aspects before the decision is rendered. Accordingly, it is questionable whether the methodology or model may equally apply and provide successful results for cases where the court originally decides a dispute, rather than reviewing a lower court's decision.

Third, and somewhat relatedly, the decision outcome prediction only applies for the binary classification tasks as to whether the Supreme Court reverses or affirms the lower court's decision. As mentioned above, cases in which the Supreme Court decides a dispute as the original court of jurisdiction are excluded from the study. The study notes that this is so because 'the Court and its members may take technically nuanced positions or the Court's decision might otherwise result in a complex outcome that does not map onto a binary outcome'. (86) The very same may be said about most instances in which a court originally decides a dispute, rather than reviewing another court's decision. In those cases, the court will have to decide technically complex and nuanced matters of facts and law which are difficult to classify into a binary model. The issue of binary-tasks for AI models are further discussed below. (87) At this stage, suffice it to note that the study's methodology is not easily transposable to lower courts' decisions whose task is to originally decide a dispute rather than reviewing another court's previous decisions.

Fourth, one might also note that decisions of Supreme courts generally, and of the US Supreme Court in particular, are often highly political. US Supreme Court's justices are indeed appointed considering their political orientation, among other things. (88) The points of law on which the US Supreme Court renders decisions are often those on which lawyers from different sides of the political spectrum come out differently, the possibility of gun control being one example. (89) To the contrary, lower courts' decision are typically more fact-driven and less legally principled. Some of the features used (e.g. the judge's political orientation) are therefore less likely to be outcome-determinative, or at least the relation between the feature and the outcome is not going to be as straight-forward.
P 554

Overall, the above-mentioned studies therefore have important inherent limitations as to their general applicability for *ex ante* outcome prediction. Nevertheless, they spark the questions as to whether AI-driven and machine learning based outcome prediction tools might not be a useful addition to human decision-making. Max Radin wrote in 1925 about judicial decision-making that the judge's 'business is prophecy, and if prophecy were certain, there would not be much credit in prophesying'. (90) If AI models could prophesize or help with predictions, should they not replace, or at a minimum be taken into account by, human decision-makers? The following sections of this article aim at helping to provide an answer to this question.

4 LIMITATIONS ON LEGAL DECISION-MAKING WITH AI: THE FOUR 'V'S OF BIG DATA

Data specialists often refer to the four Vs of Big Data – Volume, Variety, Velocity, and Veracity – as the cornerstones of data-driven projects. (91) The four Vs describe challenges to Big Data use. They also help to assess data-driven AI programs such as those described in the previous section, and their use in the legal sector. This section looks at the four Vs in turn and discusses inherent limitations of data-driven models for legal decision-making with AI.

4.1 Volume: Need for sufficient non-confidential case data

Any data-driven AI programs first and foremost require access to data. Machine learning models, which are based on probabilistic inferences, are data-hungry: the larger the sample data, the more accurate the model's predictive value. In the legal sector, the volume of data required leads to a possible two-fold limitation of AI programs.

First, case data is not always easily accessible. In certain areas of law, decisions are confidential and thus not available to non-parties. Confidentially can be based on protecting the affected parties' rights or the underlying transactions. For instance, international commercial arbitration awards are generally not published and the constitution of a database to establish an AI model would therefore prove
P 555
difficult. (92) However, this is not to say that AI models in international commercial arbitration are impossible. Initiatives exist to publish commercial awards on a regular basis, typically in a redacted format. (93) In any event, even without publishing confidential awards, institutions could collect them and make them available for the purpose of building AI models.

Second, when case data is accessible, a large sample size is important. While there is no hard rule of a required sample size, the more data, the more accurate the extracted model. Accordingly, areas of law with large numbers of decisions on a given topic will be more suitable for AI models. In international investment arbitration, although there are no reliable statistics on how many awards are rendered per year, on the basis that around sixty new cases are initiated per year, (94) the number of arbitral awards should similarly only be in the double-digits, (95) which does not make for a particularly sample size.

4.2 Variety: Requirement of repetitive patterns with binary outcomes

In addition to the necessary data volume, there is also a question about the variety of the input data. In data-research terminology, variety of data refers to the fact that data comes from different sources and may be structured (e.g. a file containing names, phone numbers, addresses) or unstructured (photos, videos, social media feeds). (96) In the legal context, the variety question is likely to be framed in a different manner. The variety will not so much come from different sources or formats – since the input data is likely to be limited to previous decisions – but
P 556
rather from the content dealt with in those decisions. For AI-driven decision-making two variety questions come to mind.

The first question relates to the data input and to what extent AI-based decision-making models require repetitive fact patterns or, conversely, whether they would be able to deal with topics that are complex and non-repetitive. In the above-mentioned study on US Supreme Court decisions, the computer program was developed for decisions spanning over almost two hundred years and dealing with a large variety of issues. (97) Nevertheless, the more outliers or non-repetitive issues, the more difficulties the AI model will face. In international arbitration, therefore, AI programs are more likely to apply to international investment arbitration (which typically raises a number of well-known issues) than in international commercial arbitration (which deals with diverse and often unique issues).

The second question relates to the model output. The legal prediction studies discussed above all use a binary classification as the output task. In the case of the ECtHR, the binary classification was whether or not a violation of the relevant provision of the Convention occurred, and in case of the US Supreme Court decision, the binary classification task was whether or not the Court affirmed the lower court's decision. As already noted above, this raises the question whether those, or other similar models, could be built for more diverse, non-binary tasks. (98)

One might be tempted to reply that any legal decision could be subdivided into a multitude of binary classification tasks, such as whether (1) the tribunal has jurisdiction: yes/no; (2) the parties validity entered into a contract: yes/no; (3) one party breached the contract: yes/no etc. Lord Hoffman has famously described a standard of proof issue using a binary analogy:

> 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 happened. If he does discharge it, a value of 1 is returned and the fact is treated as having happened. (99)

However, while it is true that many legal questions of fact or law can be reduced to a 0/1 or yes/no binary task, the problem is that there will be a multitude of such binary tasks in each case, and determining all of them will be case-specific. For an AI model to be able to extract the required patterns and algorithms from the input data, having one clear output question facilitates the model-building process. This
P 557
is why, in the study on US Supreme Court decisions, the research group specifically excluded those decisions in which the Supreme Court was the court of original jurisdiction, which did not correspond to a simple binary classification task. (100)

4.3 Velocity: Problem of policy changes over time

Velocity refers to the frequency of incoming data that needs to be processed. Big Data is often challenging because of the sheer amount and high frequency of the incoming data. In the legal context, such risk is very low. As already pointed out above, in terms of volume, the problem is likely to be of scarcity rather than abundance of data. [(101)] Therefore, over time, decisions might not be frequent, and when they occur there might have been a change in policy so that the previous data is outdated. These policy changes can be radical and swift at times. To take an example from the international arbitration context, the decision of the Court of Justice of the European Union in *Achmea* has fundamentally changed the compatibility of investor-state arbitration with European law overnight. [(102)]

This raises the question how AI models which, by definition, are based on information extracted from previous data may deal with those policy changes. It is true that the essence of machine learning is the ability to improve the algorithm over time. Nevertheless, such improvement is always based on past data. Policy changes in case law necessarily require departures from past data, i.e. previous cases. For these reasons, AI models are likely to keep 'conservative' approaches that are in line with previous cases.

4.4 Veracity:Risk of bias and data diet vulnerability

Finally, veracity relates to the accuracy and trustworthiness of the data used. In the AI context, the question is whether there are any hidden data vulnerabilities which might affect the model's accuracy. The robustness and trustworthiness of AI are recurrent topics in the discussion on AI. [(103)]

As a starting point, one might assume that AI models have the advantage of algorithmic objectivity and infallibility over humans who inevitably make mistakes and are influenced by subjective, non-rational factors. Research in the
P 558
area of psychology, cognitive science, and economy has shown that humans often fail to act rationally. [(104)] Most famously, Nobel-prize winner Daniel Kahneman and Amos Tversky have studied heuristics and cognitive biases in human choices. [(105)] Their studies provide multiple examples in which heuristics (i.e. cognitive short-cuts for otherwise intractable problems) and biases (i.e. factors which appear to be irrelevant to the merit of our choices but affect them nonetheless) appear in human day-to-day decisions. [(106)]

Applying this research in the legal sector, a group of Israeli and US researchers have shed some light on the importance of extraneous factors in judicial decision-making. [(107)] Looking at more than 1,100 decisions rendered over ten months by Israeli judges in relation to 40% of the country's parole applications, [(108)] the study showed that the majority of applications are rejected on average, [(109)] but the probability of a favourable decision is significantly higher directly after the judge's daily food breaks. [(110)] While not falling into the generalization of the well-known saying that 'justice is what the judge had for breakfast', the results 'suggest that judicial decisions can be influenced by whether the judge took a break to eat'. [(111)] This research provides an empirical example about how human decision-making is affected by extraneous factors, such as food breaks, which ought to be irrelevant to the merit of the case. [(112)]
P 559

Some authors have therefore concluded that AI-based decision-making would be superior to human decision-making on the basis that computers would be immune to cognitive biases or undue influence of extraneous factors. [(113)] However, a blind deferential attitude towards algorithmic objectivity and infallibility is misplaced. AI research over the past years has highlighted the risks of misbehaving or biased algorithms. Important studies discuss bias concerns in computer systems used for a variety of tasks, such as flight listings, credit scores, or on-line advertisements. [(114)] Referring to a 'scored society', some have argued that hidden and unregulated algorithms produce authoritative scores of individuals that mediate access to opportunities. [(115)] As other authors put it, 'procedural consistency is not equivalent to objectivity'. [(116)]

Any data-based computer models are only as good as the input data. Vulnerability in the data diet has negative consequences on the extracted model. In particular, the underlying data which was used to train the algorithm might have been 'infected' with human biases. The machine learning algorithm will be based on those biases and possibly even exaggerate them by holding them as 'true' for its future decisions or outcome predictions.

For instance, in the area of investment arbitration, concerns have been voiced that arbitral tribunals are inherently and unduly investor-friendly. [(117)] I do not discuss here whether this criticism is well-founded, [(118)] but rather assume for the purpose of the present demonstration that such human bias exists. In this case, an AI model based on investment arbitration data would be likely to perpetuate such (alleged)
P 560
favour given to investors. The model would likely predict favourable outcomes for investors against States in a disproportionate number of cases.

Even without going as far as pointing towards human biases in the underlying data, the model might extract patterns from the data and extrapolate them in a way that might lead to systemic mistakes. For instance, studies have shown that the use of algorithms in criminal risk assessment in the United States has led to racially biased outcomes. (119) The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system is widely used in the United States to assess the recidivism risks for defendants. Under this system, studies found that '[b]lack defendants were … twice as likely as white defendants to be misclassified as a higher risk of violent recidivism', whereas 'white violent recidivists were 63% more likely to have been misclassified as a low risk of violent recidivism, compared with black violent recidivists'. (120) Whether this racial bias in the computer program was based on existing human biases in the training data remains unclear. It might also have resulted from the fact that the algorithm wrongly classified black defendants at the higher recidivist rate because this racial group is overrepresented in certain kinds of crimes. The computer model might have extrapolated from this pattern the wrong assumption of a higher recidivist risk.

The occurrence of systemic errors based on hidden patterns in the underlying data is a serious risk. As discussed above, in the study on ECtHR decisions, words with high predictive value include 'Ukraine' or 'Russian'. (121) Presumably, this was the case because a significant number of ECtHR cases are directed and decided against these countries. (122) Statistics show that a number of countries receive the most applications and condemnations. (123) A computer program modelled on data containing a higher proportion of condemnations of a given country might extrapolate a higher risk of a violation committed by this country in the future and its outcome predictions might thus be biased against this country.
P 561

It is therefore important to consider whether and how systemic mistakes in algorithms might be addressed. In systems where the algorithm is coded by a human programmer, the mistake will often be in the design of the algorithms itself. It can be changed once the mistake is detected. (124) To the contrary, in machine learning systems the algorithm is extracted from the data in the sample set, as described above. (125) Mistakes will thus usually result from the input data and are more difficult to detect and fix. Hiding sensitive elements in the input date, such as ethnic background or geographical origin, could be considered in helping to prevent issues. However, even if those sensitive features are hidden, algorithms might nevertheless implicitly re-construct them from proxy variables. (126)

Moreover, as discussed above, the aim of machine learning is that the computer programs learn from experience and improve their performance over time. (127) The algorithm is therefore influenced not only by the original training dataset, but also by the use and continued data-input over time. Users therefore have a certain 'power' to change the algorithms. The swearing habit and other unacceptable behaviour of the AI-chatbot Tay, following interaction with its Twitter users, is a salient example. (128) One could also imagine that users in the legal context attempt to unduly influence or game the algorithms to obtain favourable results. For instance, if it were transparent that certain words, or cluster of words, such as in the study of the ECtHR decisions, led to a positive case prediction, the targeted use of those words in a party's legal submissions might lead to an inappropriate influence of the outcome.

Overall, this section has shown that a number of system-inherent limitations exist in the use of AI programs for legal decision-making. These limitations need to be carefully considered before promoting the use of AI in this context. Moreover, other more fundamental and wide-reaching concerns exist and are discussed in the next sections.
P 562

5 BLACK BOX OF LEGAL DECISION-MAKING WITH AI: NEED FOR REASONED DECISIONS

Providing a reasoned decision that outlines the premises on which it is based constitutes one of the fundamental features of legal decision-making. Schematically, one can distinguish several objectives for providing reasons in legal decisions. First, reasons help the losing party to understand why it lost and make the decision more acceptable (legitimacy objective). Second, reasons also allow the parties to the dispute, and if the decision is published, third parties in similar situations, to adapt their behaviour in the future (incentive objective). Third, reasons further allow other decision-makers to follow the same rationale or explain their departure therefrom (consistency objective). While one might discuss whether there is market for unreasoned decisions (e.g. in certain instances, parties might be interested in 'quick-and-dirty' unreasoned decisions), legal decisions must provide reasons unless the parties have provided otherwise.

AI programs will have significant issues in providing reasoned legal decisions and meeting those rationales. (129) Indeed, not only in the legal sector, but more broadly, the inability to explain results obtained with AI programs has raised concerns. [(130)] For example, disturbing results were obtained from an AI program able to guess a person's sexual orientation from publicly posted profile pictures. [(131)] The accuracy rates are troubling (83% for women and 91% for men) but what is even more alarming are the researchers' difficulties in determining the bases on which the AI program obtained those results. [(132)] This highlights the general problem for AI research of the so-called explainability or interpretability of its results. [(133)]

This difficulty is due to the features of certain AI models. Expert models or decision-trees follow pre-established rules, as detailed above. [(134)] It is therefore possible to identify the causes that led to a given result on the basis of those

P 563

rules and thus make the model explainable. [(135)] To the contrary, as also explained above, other machine learning models, such as neural networks, often have no pre-defined rules but use pattern-recognition to extract the required algorithm. [(136)] These systems may use hidden units which correspond to hidden attributes not directly observed. [(137)] As a consequence, the process by which those AI models obtain results is 'black-boxed' and not easily explainable. [(138)]

AI research tries to deal with those issues and develop Explainable Artificial Intelligence, also called XAI. [(139)] One possible route is the use of counterfactual scenarios. The model selects alternative samples with different features, compares the different outcomes under each and is therefore able to identify how and why they differ. [(140)] For instance, the model will be able to detect that the outcome in a given case would have been different, had feature X been absent or feature Y been added. In other words, the model for the actual decision-making is accompanied by another model, the purpose of which is to provide an explanation. [(141)]

The difficulty with providing reasoned legal decisions obtained by AI is two-fold. First, it may be difficult to identify the actual factors that have led to a certain outcome prediction in case of black-boxed models. Second, even if certain factors are identifiable as causes for a given outcome prediction, these factors might not prove a useful explanation. For instance, in the above-mentioned study on ECtHR decisions, certain words, or cluster of words, were identified with a high predictive value. [(142)] However,

P 564

the information that words such as 'injury', 'Ukraine', 'copy', or 'January' have contributed to the outcome prediction falls short of an explanation which is deemed sufficient for a legally reasoned decision.

It is important to distinguish here between causal attribution, which is the process of extracting a causal chain and displaying it to a person, and causal explanation, which includes the social process of transferring knowledge between the explainer and the explainee with the goal that the explainee has the information needed to understand the causes of the event. [(143)] The latter not only requires AI to identify causes, but also to provide contextual explanation. Miller has shown that useful AI explanation must therefore take into account the human addressee. [(144)] This means, among other things, that explanation selection is important: typically, only a small subset of all possible causes are useful as an explanation for any given individual. [(145)] For instance, drawing from the ECtHR study results, the fact that an event happened in 'January' might be a cause for the decision, but less useful an explanation than that it constituted 'ill treatment'.

An explanation is also generally presented relative to the explainer's beliefs about the explainee's beliefs. [(146)] Dworkin has emphasized the importance of the shared context of law. In his major work, *Law's Empire*, he developed a theory of law as an interpretive practice that occurred in a community of interpreters. [(147)] Borrowing from the hermeneutical tradition, Dworkin claims that an understanding of a social practice, like law, requires turning to the meaning it has for participants. The meaning of law can therefore only be retrieved from within a shared context. [(148)] These contextual elements are likely to pose problems for AI-based legal explanation or reasoning.

Moreover, social scientists have tested the value of probabilistic explanations. [(149)] Overall, the use of statistical or probabilistic relationships are not as satisfying as causal explanations. For instance, if a student received a 50/100 in an exam and asks about the reasons for such score, the teacher's explanation that a majority of the class received the same score is unlikely to satisfy the student's request. Adding why most students received this score might be felt as an improvement, but not as much as explaining what this particular student did to receive his or her result. [(150)]

This example illustrates the difficulties for explanations or reasons in AI decision-making, which are, as detailed above, typically based on statistical or

P 565

probabilistic models. [(151)] Providing an 'explanation', say, that the likelihood of a claim to be dismissed is 86%, will not satisfy the losing party. It does not meet any of the objectives for legal reasoning outlined at the outset of this section. First, the legitimacy objective is not met, because statistical information is unlikely to help the losing party to understand why it lost and make the decision more acceptable. Second, the incentive objective fails because statistical information also does not allow parties or third parties to adapt their behaviour in the future. Finally, the consistency objective is not satisfied because other decision-makers have no information as to why they should follow the same rationale or depart therefrom.

The need for reasoned decisions is therefore likely to be an important barrier for AI-based legal decision-making. The impact of the probabilistic nature of AI models, however, raises even more fundamental questions as to the overall paradigm of decision-making, as discussed in the next section.

6 PARADIGM-SHIFT IN LEGAL DECISION MAKING: PROBABILISTIC INFERENCE INSTEAD OF DEDUCTIVE REASONING AND LOGIC?

Evaluating whether AI would be able to contribute to legal decision-making invariably raises the question how humans make legal decisions. As early as 1963, Lawlor speculated that computers would one day become able to analyse and predict judicial decisions, but noted that reliable prediction would depend on a 'scientific' understanding of the ways the law and the facts impact the judges' decision. (152) Even today, such 'scientific' understanding of judicial decision-making is lacking and is a debated topic amongst legal philosophers and theorists.

Theories of judicial decision-making abound, but a fundamental distinction exists between those that postulate the use of logic by ways of deductive reasoning on the basis of abstract, pre-determined legal rules (regrouped in the category of 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 shows that the use of AI in legal decision-making does not fit easily in either category. AI models would elevate probabilistic inferences to be the basis for legal decision-making and, as this section shows, this would constitute a sharp paradigm shift.
P 566

6.1 Legal formalism and the use of deductive reasoning and logic

Legal formalism, in its purest form, posits that law is, and should be, an entirely self-contained system, in which judges never face choices or questions of interpretation that would be resolvable through extra-legal considerations. (153) Rather, 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)

A judicial decision is thus the product of a seemingly mechanical or mathematical application of pre-established legal principles or rules to the proven facts using means of logic. (155) The underlying idea can be expressed in the simple formula 'R + F = C' or 'rule plus facts yields conclusion'. (156) More specifically, the legal syllogism will consist of a major premise in the form of the pre-established rule (e.g. 'if P then Q') and a minor premise seeking to establish that the required condition stipulated in the major premise (P) occurred in fact. If such condition is met, by means of a deductive reasoning, or subsumption, the judge concludes that the legal consequence (Q) is to be applied in the case at hand as a matter of logic. (157)

Today, it is rare to find 'pure' formalists, but the main idea of legal decision-making as based on deductive reasoning and logic remains influential. In his seminal work *The Concept of Law*, Hart introduced an important distinction between clear cases, for which the simple deductive reasoning applies, and hard cases, for which extra-legal moral and political consideration may come into play. (158) Drawing on the philosophy of Wittgenstein, Hart emphasizes the indeterminacy of natural language and the open texture of law, for instance, through the use of general standards, such as 'good faith'. (159)

Even in their more nuanced forms, legal formalist theories still point to deductive, logical, rule-based reasoning as the guarantee for the objectivity, impartiality, and neutrality of law. MacCormick wrote in 1994:
P 567

> 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)

AI processes, if applied in the legal context, would potentially run counter to this understanding of legal decision-making. As described above in section 2, some computer models (such as expert models) are indeed rule-based, using causal logic and deductive reasoning, since they apply pre-established rules in the algorithm to the observable data. Other AI models, however, have different features. In particular, machine learning models, such as neural networks, often have no pre-defined rules. Deductive, causal reasoning is thus replaced by an inverse approach, because the machine learning program extracts the algorithm from the observable data. Rather than using logic, the AI model calculates probabilities, i.e. the likelihood for any given outcome. (161)

Applying such machine learning processes in the legal decision-making context therefore would mean accepting a departure from the above-mentioned understanding of judicial reasoning according to formalist theories. A decision based on those AI models would *not* be based on pre-determined legal rules, would *not* be the result of deductive logic, and would *not* follow the above-described legal syllogism. While this situation would be a cause for concern for legal formalists, it might be seen as vindicating others who have long criticized formalist theories.

6.2 Legal realism and the importance of extra-legal factors

Legal formalism has attracted important criticism over time. In the first half of the twentieth century, legal realists attacked the fundamental postulates of formalists theories. (162) Even though realist theories vary significantly, they have some commonalities. Llewelyn and others attacked the idea that the law was a mechanical application of pre-determined rules by the judge by means of logic and deductive reasoning. (163) Accepting that legal certainty was a myth, realists,
P 568
such as Frank, developed what they called rule scepticism and drew attention to the fact that rules do not play a determinative part in legal decision-making. (164) Rather, judges decide cases based on extraneous non-legal factors or their 'hunches' and then *ex post* provide their decision with a seemingly logical rule-deferring coating. (165) Unmasking the hypocrisy and double-standard of judicial decision-making, realists argue that logic and rule-deference is only a facade and ignores the social interests involved. This thought was later developed by the movement of critical legal theory, emphasizing the political significance of the law as a means of empowerment and emancipation. (166) Rather than being a mechanical and supposedly neutral application of rules, law does not contain a 'right answer' but corresponds to competing normative visions. (167)

Even before the legal realist movement became well-known, Justice Oliver Wendell Holmes described decision-making in similar ways. In 1897, in his seminal work, *The Path of Law*, he criticized what he called the 'fallacy of logic':

> 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 insisted that law was imminently a matter of prediction, emphasizing the importance of statistics for the future of the law. He described his work as a study on prediction and more precisely 'the prediction of the incidence of the public force through the instrumentality of the courts'. (169) He thus argued that 'a legal duty so called is nothing but a prediction that if a man [or woman] does or omits certain things, he [or she] will be made to suffer in this or that way by judgment of the court; and so of a legal right'. (170) In order to make correct predictions, he surmised on the use of statistics for future lawyers' generations, noting that '[f]or the rational study of the law the black-letter man [or woman] may be the man [or woman] of the present, but the man [or woman] of the future
P 569
is [one] of statistics and the master of economics', adding that '[t]he number of our predictions when generalized and reduced to a system is not unmanageably large'. (171)

Holmes's emphasis in 1897 on prediction and statistics in legal decision-making, in lieu of logic, shines today in new light when considering the implications of AI. As discussed above, predictions based on statistics or probabilities are precisely features used in AI machine learning models. (172) Moreover, the importance of extraneous non-legal factors, as argued by the legal realists, is confirmed by the predictive AI studies, cited above. (173) In the ECtHR study, the part of the judgments with the highest predictive value is not the legal section but the section relating to the factual background. [(174)] Also, the US Supreme Court study included in the computer model extra-legal elements such as the judges' political preferences. [(175)]

Are we therefore to conclude, as some have argued, [(176)] that AI would vindicate the legal realists' theories? And that the possible use of machine learning models in legal decision-making would be in line with what human judges have always done? Would therefore, in essence, the debate between formalists and realists eventually be won by the latter? These conclusions, however, ignore an important point: the central place of probabilities as a normative basis for AI machine learning. As discussed in the next section, this goes well beyond legal realist theories.

6.3 Use of probabilistic inferences: Towards legal determinism?

When discussing legal theories on judicial decision-making, an important distinction needs to be drawn between their descriptive aspect (i.e. how judges *do* effectively reason and make decisions) and their prescriptive or normative aspect (i.e. how they *should* reason and make decisions). [(177)]

Legal formalism contains both a descriptive and normative element. Formalists *describe* the process by which judges apply the law as a matter of logic, deduction, and legal syllogism. [(178)] They also argue that the self-contained nature of the law, the neutrality of legal thinking untouched by extraneous non-legal factors is, normatively, how it *should*
P 570
be. This is in order to keep the law clear of politics or morality [(179)] and provide for a 'modern paradigm of legal rationality under the "rule of law"'. [(180)]

Legal realism, to the contrary, is first and foremost concerned with descriptive aspects. Holmes, Frank and others trace the *reality* of judicial decision-making – hence the name of the movement. They highlight the influence of extraneous non-legal factors, criticizing the formalistic, automatic, mathematical rule-application approach as utopian and far from the real world. However, they do not go as far as arguing that judges *should* take into account extraneous non-legal factors. To use the Israeli parole study, mentioned above, [(181)] as an illustration: while it might be a matter of fact that judges are influenced by extraneous factors such as food breaks, no one seriously argues that this is a good thing and should be the normative basis for judicial activity.

Normative aspects are not entirely foreign, though, to other theories, such as the critical legal theory movement, for instance. Unger and others have stressed the political significance of law and the social interests involved. Bringing out the normative aspects, law is taken as a means for effective radical social transformation. [(182)]

When looking at AI models, the foregoing leads to a number of observations. AI models would not only decide based on probabilities as a matter of fact, but would also be their normative basis. As mentioned above, a decision based on machine learning AI models would *not* be based on pre-determined legal rules, would *not* be the result of deductive logic, and would *not* follow the above-described legal syllogism. [(183)] This would be true on a descriptive level (i.e. how these models do effectively decide) and, importantly, also on a normative level (i.e. how these models should decide). Replacing logical, deductive and rule-based reasoning by probabilistic inferences as the normative framework of judicial decision-making would therefore not only constitute a departure from legal formalism, but would also go well beyond legal realists' theories.

Indeed, realists accept that judges, after having made their decision based on a variety of factors, including non-legal, political, and moral considerations, do render their decision coated in a format that seeks to comply with logic, using a rule-based deductive reasoning. [(184)] What realists criticize is the hypocrisies of such a facade, but they accept that such facade or format exists. AI-based decision-
P 571
making would take away such format. AI decisions would not be rendered making reference to deductive or causal reasoning based on legal rules. The problems related to this lack of reasoning have already been highlighted in section 5 above.

More fundamentally, however, the absence of a logical framework in judicial decision-making has implications that go beyond the descriptive or normative aspects discussed. Hart has distinguished three levels in judicial reasoning: (1) the processes or habits of thought by which judges actually reach their decision (descriptive psychology); (2) recommendations concerning the processes to be followed (prescriptive judicial technology); and (3) the standards by which judicial decisions are to be appraised. [(185)] It is at the third level that the absence of logic, at a minimum, causes concern because it undermines the assessment or justification of the decision. Or as Hart puts it:

> the issue is not one regarding the manner in which judges do, or should, come to their decisions; rather, it concerns the standards they respect in justifying decisions, however reached. The presence or absence of logic in the appraisal

> of decisions may be a reality whether the decisions are reached by calculation or by an intuitive leap. (186)

In addition, to the extent legal theories emphasize the political significance of law, as well as the fact that decision-makers have discretion to 'fill in' general standards, such as 'good faith', (187) the question arises how these political or moral considerations would be managed in an AI model. Who or what would be in a position to influence those political or moral considerations? In a traditional computer model, one might point towards the programmer. However, as described in section 2, in advanced AI models, the algorithm is not coded by a programmer but extracted from the observable data. Therefore, the only basis for the decision, even on morally or politically sensitive issues, will be past data. As already pointed out above, AI models are thus likely to take a conservative approach, even in a machine learning context of ever-improving algorithms. (188)

Using statistics or probabilities as the normative framework for judicial decision-making seems also problematic for other reasons. So far, probabilities or statistics are not an accepted legal basis for decisions. (189) English and other common law lawyers will be familiar with the term 'balance of probabilities' which sets out a standard of proof. (190) Importantly, however, this applies only to the establishment of facts. For instance, in *Miller v. Minister of Pensions*, the UK Supreme Court (then House of Lords) elaborated the balance of probabilities concept, stating that if 'the evidence is such that the tribunal can
P 572
say "We think it is more probable than not", the burden is discharged, but if the probabilities are equal, then it is not'. (191) Once the facts are established using this method, probabilities have no room in judicial decision-making. For instance, one cannot grant a claim merely on the basis that there is an 80% chance that the established facts constitute a violation of the contract.

The previous example illustrates well the concrete issues with probabilistic bases for decision-making. What threshold would be appropriate above which a claim is deemed granted? Would anything above 50% be sufficient? Or would one require a higher threshold of, say, 80%? Even with such a higher threshold, though, one consciously accepts that there is a 20% likelihood that the case is decided wrongly.

In this context, it is also worth remembering the issue of data diet vulnerability and resulting bias risks, discussed above. (192) For instance, one might consider a situation where State X has been repeatedly found in violation of a substantive investment protection mechanism found in investment treaties. Does this influence the likelihood that State X will lose a future investment claim brought by another investor?

In sum, using a probabilistic analysis as the normative basis for decision-making is not only an important paradigm shift from a theoretical point of view, it also raises important concrete questions. This new approach could be called legal determinism since it determines future outcome on probabilistic calculations based on past data. As shown in this article, it has a number of implications for judicial decision-making which need to be considered carefully.

7 CONCLUSION

The aim of this article is to explore the use of AI in arbitral or judicial decision-making. Having assessed the technical aspects of AI and its implications and limitations, as well as the more fundamental impact it may have on human decision-making processes and theories thereof, the main findings and conclusions of this study are as follows:

−Existing studies on decision outcome prediction, while obtaining spectacular accuracy rates of 70–80%, contain important limitations. An analysis of the methodology and assumptions employed puts into doubt the claim these models might pave the way for *ex ante* outcome predictions. Among other things, it is questionable whether the models may equally apply and provide successful results for cases where the court originally decides a dispute, rather than reviewing a lower court's decision. (193)

P 573
−The technical features of AI imply certain requirements for its use in judicial decision-making, at least as of today. This includes, for instance, the need for sufficient non-confidential case data (194) and, possibly, the requirement of repetitive fact-patterns and binary outcomes. (195) Given that AI models are typically based on information extracted from previous input data – even in ever-improving machine learning algorithms – they are likely to follow 'conservative' approaches and might not be adapted to deal with important policy changes over time. (196) Also, a blind deferential attitude towards algorithmic objectivity and infallibility is misplaced. Any data-based computer models are only as good as the input data, and there is therefore a risk that they perpetuate existing biases. (197)

−The need for reasoned decisions is likely to be an important barrier for AI-based legal decision-making. At least at the current technological level, it may be difficult to identify the factors that have led to a certain outcome prediction in case of black-boxed models. Moreover, even if certain factors are identifiable as causes for a given outcome prediction, these factors might not prove a useful explanation for human addressees in a given context. (198)

−The use of AI does not fit easily in legal theories on judicial decision-making. AI models elevate probabilistic inferences to be the normative basis for legal decision-making. This not only constitutes a paradigm shift from a theoretical point of view, but also raises important questions as to whether and how the outcome of future decisions should be determined on probabilistic calculations based on past data.

These conclusions, however, should not detract from the most obvious point: AI will fundamentally affect the legal profession and legal activities, including judicial decision-making. It is therefore important to study further how best to use AI, even with the limitations, barriers, and issues highlighted in this article. In international arbitration, which is under constant criticism for being too expensive and time-consuming, the claim by some AI developers that computers 'can do the work that took lawyers 360,000 hours' must be taken seriously. Future research is necessary to explore ways human decision-makers and AI can best be combined to obtain the most efficient results. Coming back to the quotation from Antoine de Saint-Exupéry in the introduction, we may not be able to foresee what the future of AI models looks like, but we can enable that future by carefully considering the implications of judicial decision-making with AI.
P 573

References

*)

Professor of Law, Queen Mary University of London; Special Counsel, Wilmer Cutler Pickering Hale and Dorr LLP. The author wishes to thank Jose Alvarez (NYU), Pierre Brunet (Université de Paris I Panthéon-Sorbonne), Jacques de Werra (Université de Genève), Susan Franck (American University, Washington College of Law), Mark Kantor (via OGEMID), Elizabeth Whitsitt (University of Calgary), for having provided most helpful comments and thoughts on an earlier version of the article. Email: maxi.scherer@wilmerhale.com.

1)

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).

2)

For instance, the part of the US legal service market in relation to new technologies in law is estimated to grow to USD 55 billion (from USD 12 billion in 2017), while at the same time traditional law firm services are estimated to fall to USD 265 billion (from USD 300 billion in 2017). *See* 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/2018conf/ma... (accessed 9 May 2019).

3)

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 New Black – 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).

4)

See also Maxi Scherer, *International Arbitration 3.0 – HowArtificial Intelligence Will Change Dispute Resolution*, Austrian Y.B. Int'l Arb. 503 (2019). For studies on human arbitral decision-making, *see* in particular Susan D. Franck et al., *Inside the Arbitrator's Mind*, 66 Emory L.J. 1115 (2017).

5)

See e.g. Amy J. Schmitz, *Building on OArb 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. For a discussion on the use of arbitration for data disputes, such as those arising out the European General Data Protection Regulation (2016/679) (GDPR), *see* Jacques de Werra, *Using Data Arbitration and Data ADR for Solving Transnational Data Disputes: Lessons from Recent European Regulations?*, Am. Rev. Int'l Arb. (on file with author, forthcoming).

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. 3; 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. 3.

8)

Soares, *supra* n. 3, 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).

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 9 May 2019).

12)

Oxford Living Dictionaries, https://en.oxforddictionaries.com/definition/artificial_intelligence (accessed 9 May 2019).

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 May 2019).

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 May 2019); 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).

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.

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 9 May 2019).

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)

See infra s. 3.

32)

Boden, *supra* n. 18, at 39–40.

33)

Alpaydin, *supra* n. 18, at 63–64, 82–84.

34)

Ibid., at xii.

35)

One early example of 'intelligent' machine behaviour was the IBM 'Deep Blue' computer beating the chess champion Gary Kasparov. On this experiment, which took place already 20 years ago, *see* 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).

38)

Boden, *supra* n. 18, at 40.

39)

For more details, *see ibid.*

40)

For some of the earlier studies, *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 Decisionmaking*, 104 Colum. L. Rev. 1150 (2004). Generally on forecasting, *see* Philip E. Tetlock, *Expert Political Judgment: How Good 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 9 May 2019).

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 9 May 2019).

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.

55)

Ibid. In addition to the decision sections, the study also created certain topics, which overall had a higher prediction result than the decision sections and which, combined, led to the overall result of 79%.

56)

Ibid., at 13, table 3.

57)

Ibid., at 14, table 4.

58)

Ibid., at 15, table 5.

59)

Ibid., at 2.

60)

Ibid., at 12.

61)

Ibid., at 16.

62)

Ibid., at 8.

63)

Ibid.

64)

Ibid., at 8, 10.

65)

Ibid., at 5.

66)

Ibid.

67)

Ibid., at 2.

68)

See infra ss 4.2 and 5.

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 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, table 1.

74)

For a full list of the features, *see ibid.*, at 4–6. Many of the features used were not originally labelled but taken from the Supreme Court Database (SCDB) established by Harold Spaeth for the use of empirical studies. Harold J. Spaeth et al., *Supreme Court Database* (Version 2016, Legacy Release v01 (SCDB Legacy 01)), supremecourtdatabase.org (accessed 9 May 2019).

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 6–8.

81)

Ibid., at 8–9.

82)

Ibid., at 9.

83)

Ibid., at 15.

84)

Ibid., at 3.

85)

The study notes that 'in practice, the predictions for a case may evolve as new information about the case is acquired prior to the decision being rendered'. *Ibid.*, at 5.

86)

Ibid.

87)

See infra s. 4.

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 How Judges 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 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 9 May 2019).

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: Review of Developments in 2016* 1 (May 2017); UNCTAD, *Investor-State Dispute Settlement: Review of Developments in 2017* 1 (June 2018); UNCTAD, *New ISDS Numbers: Takeaways on Last Year's 71 Known Treaty-Based Cases* (13 Mar. 2019), https://investmentpolicyhubold.unctad.org/News/Hub/Home/1609 (accessed 9 May 2019).

95)

This takes into account that, on the one hand, some disputes will settle without any award being rendered, and on the other hand, some disputes will rise to multiple partial awards.

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).

97)

See supra s. 3.2.

98)

See supra s. 3.2.

99)

In re B [2008] UKHL 35.

100)

See supra s. 3.2.

101)

See supra s. 4.1.

102)

Case C-284/16 *Slovak Republic v. Achmea B.V.* (CJEU, 6 Mar. 2018).

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 9 May 2019).

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-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)

Parole is a permanent release of a prisoner who agrees to certain conditions before the completion of the maximum sentence period.

109)

Ibid., at 6889 (64.2% of the applications in the sample were rejected).

110)

Ibid., at 6890 (the probability of parole being granted spikes at approximately 0.65 at the beginning of the session after each food break and declines to nearly 0 at the end of each session).

111)

Ibid. More specifically, the study concludes that judges when making repeat rulings show a tendency to rule in favour of the status quo (i.e. reject the parole application for liberation) and that this tendency can be overcome, for instance, by taking a food break. *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 May 2019).

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).

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., *How We Analyzed the COMPAS Recidivism Algorithm*, ProPublica (23 May 2016), www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed 9 May 2019).

120)

Jeff Larson et al., *supra* n. 19, at 2.

121)

See supra s. 3.1.

122)

In the above-mentioned study on ECtHR decisions, the research group specifically selected the same number of violation and non-violation decisions in order not to pre-influence the data model in one way or another. However, no such precaution was taken when it comes to other criteria, such as geographical origin of parties. *See supra* s. 3.1.

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 and Nissenbaum describe a flight routing system sponsored by a US airline which systematically presented this airline on the first page. *See* Friedman & Nissenbaum, *supra* n. 114, at 331.

125)

See supra s. 2.

126)

Simon DeDeo, *Wrong Side of the Tracks: Big Data and Protected Categories* (2015), https://arxiv.org/pdf/1412.4643v2.pdf (accessed 9 May 2019) (for instance, income might be inferred from proxy variables such as postal codes).

127)

See supra s. 2.

128)

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 9 May 2019) (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).

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 9 May 2019).

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 (accessed 9 May 2019) (defining interpretability as the ability for humans to understand operations either through introspection or through a produced explanation).

134)

See supra s. 2.

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)

See supra s. 2.

137)

Alpaydin, *supra* n. 18, 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 9 May 2019); George Nott, *'Explainable Artificial Intelligence': Cracking Open the Black Box of AI*, Computer World (10 Apr. 2017), www.computerworld.com.au/article/617359/ (accessed 9 May 2019).

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), 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)

See supra s. 3.

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).

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)

Example provided by Miller, *supra* n. 143, para. 4.5.2.

151)

See supra s. 2.

152)

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

153)

See e.g. Hans Kelsen, *Reine Rechtslehre* 478 (2d ed., Deuticke 1960).

154)

Max Weber, *Wirtschaft und Gesellschaft (Economy and Society)* 657–58 (Tübingen 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.

160)

MacCormick, *supra* n. 156, at ix–x.

161)

See supra s. 2.

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).

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)

See supra s. 2.

173)

See supra s. 3.

174)

Aletras et al., *supra* n. 45, at 10. *See supra* s. 3.1.

175)

For a full list of the features, *see* Katz, Bommarito & Blackman, *supra* n. 69, at 4–6.

176)

Aletras et al., *supra* n. 45, at 16 (who argued that their study results 'back … basic legal realist intuitions'). *See supra* s. 3.1.

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 supra s. 6.1.

179)

See e.g. Kelsen, *supra* n. 153, at 478 ('What is here chiefly important is to liberate law from the associate which has traditionally been made for it – its association with morals.').

180)

MacCormick, *supra* n. 156, at ix–x.

181)

See supra s. 4.4.

182)

See e.g. Unger, *supra* n. 167.

183)

See supra s. 6.1.

184)

See e.g. Holmes *supra* n. 168, at 465–66 ('The training of lawyers is a training in logic. The processes of analogy, discrimination, and deduction are those in which they are most at home. The language of judicial decision is mainly the language of logic. And the logical method and form flatter that longing for certainty and for repose which is in every human mind.').

185)

Hart, *supra* n. 177, at 105.

186)

Ibid., at 105. *See also* Richard A. Wasserstrom, *The Judicial Decision* (Stanford University Press 1961).

187)

See supra s. 6.1.

188)

See supra s. 4.3.

189)

See e.g. the discussion in the US Supreme Court case of *McCleskey v. Kemp*, 481 U.S. 279, 287 et seq. (1987).

190)

See e.g. Emily Sherwin, *A Comparative Viewof Standards of Proof*, 50 Am. J. Comp. L. 243 (2002).

191)

House of Lords, [1947] 2 All E.R. 372 (opinion delivered by Lord Denning).

192)

See supra s. 4.4.

193)

See supra s. 3.

194)

See supra s. 4.1.

195)

See supra s. 4.2.

196)

See supra s. 4.3.

197)

See supra s. 4.4.

198)

See supra s. 5.