

SOME THOUGHTS ON ARTIFICIAL INTELLIGENCE¹

PROFESSOR DOUG JONES AO²

The advent of artificial intelligence (AI) in the legal profession has, of course, already occurred, and its adoption at all stages of the dispute chain (ie, from its inception to its final disposition) promises to become increasingly widespread. To take international arbitration as an example — the relationship of which with technological innovation has an interesting history — it was found in the 2018 Queen Mary Survey on International Arbitration that 47% of arbitration practitioner respondents already made use of AI-based applications in their arbitration work, with 78% opining that such applications should be used more widely.³

By way of preface, there is no agreed definition of AI. However, it may broadly be defined as a machine that is designed to perform “human” tasks with a high level of proficiency, whether by replicating or simulating the processes by which a human would accomplish the task in question.⁴ There is a certain irony in the much insisted upon unsuitability of Artificial Intelligence to typology and definition on the one hand, and the mission of certain AI-promoting legal theorists to define something as complex as the law with reference to statistical data legible to computers on the other.⁵ In any event, the definitional ambiguity of “AI” means that it will be used in a contextual rather than strictly precise way in this paper.

This paper offers some of the author’s thoughts on AI, and the role that it is playing, and should play, in the legal profession. The paper begins in Part I by providing some background to the nature of AI and its current use in legal contexts. Ever-evolving as AI naturally is, particularly in recent months and years, it has certain essential features that, at least at present, appear to define the limits of what AI is capable of. In Part II, this paper considers certain fundamental problems with the use of AI in legal contexts. In doing so, this paper will propose a hypothetical “perfect AI”, with a view to demonstrating that even the perfection of AI technologies should not ever motivate us to supplant the human element present in the law.

The impression given by this bipartite structure, with considerable emphasis placed on the potentially serious shortcomings that are inevitable in AI systems, may be that the author’s

¹ A previous version of this paper was contributed to the Singapore International Commercial Court 2024 Annual Conference, for the Session entitled “Commercial Courts in the Age of New Technologies: Artificial Intelligence and Linguistics”.

² Independent arbitrator and International Judge of the Singapore International Commercial Court. The substantial contribution to the development of this paper by Peter Taurian, Legal Assistant, Sydney Arbitration Chambers, is gratefully acknowledged.

³ Queen Mary University of London, *2018 International Arbitration Survey: The Evolution of International Arbitration*, 31–3, discussed in Kevin Ongena, “AI Arbitrators... ‘Does Not Compute’”, Chapter Five in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 101, 102 [3]. That figure increased in the 2021 survey, although at a lesser rate than was anticipated, which the authors of the report attributed to ongoing mistrust in the technology: Queen Mary University of London, *2021 International Arbitration Survey: Adapting Arbitration to a Changing World*, 20–2.

⁴ See, eg, John Frank Weaver, “Defining AI in Contracts” (2020) 3(6) *RAIL* 435, 435.

⁵ This irony plays out across two pages of Christy Ng, “AI in the Legal Profession”, Chapter Three in Larry A DiMatteo, Cristina Poncibò and Michael Cannarsa, *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge, 2022) 35, in which the author refers to the “counter-intuitive task” of defining AI (at 36) before announcing that AI will need to “mak[e] legalese machine-readable so that clear rules can be extracted by any AI program for downstream users” (at 37).

perspective of AI is characterised by mistrust or suspicion. On the contrary, AI presents, from this author's perspective, a most exciting opportunity for the legal profession, and it should be embraced with open arms. It is this very enthusiasm that warrants an approach to the discussion of AI that is earnest and serious, and that does not flippantly disregard the profound questions that scientists and philosophers have been asking of AI for decades. Too frequently does discussion of AI in legal contexts focus on whether or not the technology is ripe for use. Invariably, it either already is or soon will be. As such, this paper takes no "side", and offers simply that which its title suggests: some thoughts on artificial intelligence.

I Background

Neither a full explanation of how AI operates nor a detailed elaboration of how lawyers are currently using AI-based tools is within the scope of this paper.⁶ However, it is necessary to say something of both by way of background before evaluating the consequences of using AI.

1 AI in the Legal Profession

There are two categories of AI and its application in the context of the legal disputes:

- a) Predictive AI (particularly "QLP", Quantitative Legal Prediction) — for which the desired output is either a binary (ie, "you will win" / "you will lose", or with more complex variations, such as quantum of damages, length of sentence, etc), or, more likely, a probabilistic prediction as to the likelihood of each outcome. These are naturally oriented towards decision-making, in that they project what is likely to happen in the future, such as would enable an adjudicator or practitioner to take steps to avoid or achieve that outcome — however knowledge of such likelihoods might naturally motivate particular arguments from lawyers or reasoning from judges. For the decision-maker, it may make predictions as to the likelihood of real-life events (most famously the likelihood of recidivism), such as might motivate a judge to sentence or approach a bail hearing in a particular way. For the practitioner, it may make "meta-predictions" as to the likelihood of winning a case before a particular court, on the basis of previous courts' approaches to the topic, such as might motivate the practitioner to proceed with the matter in a particular way (eg, avoiding litigation).
- b) Generative AI (particularly "LLMs", Large Language Models) — for which the desired output is text (known as "natural language" so as to distinguish itself from "programming language"). These are naturally oriented towards the preparation and drafting of legal texts, whether statutes, contracts, submissions or decisions. The extent to which they are used — summarising statutes and cases or written submissions at the less intrusive end of the spectrum; generating entire draft decisions

⁶ The most comprehensive and useful guide for legal audiences is that of Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart, 2020).

or submissions at the more dramatic end — is simply a matter of the efficiency and fidelity of the AI as perceived by the user.

The above distinction is useful when considering the utility and appropriacy of the AI systems in their respective contexts. What they output is fundamentally different. However, the distinction does not bear substantially on how the systems work, discussed further below. The above also alludes to another important distinction about *when* AI is invoked in the life cycle of, for example, a dispute. Ethical and other considerations will be different when considering a lawyer using an AI system to help summarise or draft pleadings as opposed to a judge using an AI system to help decide a dispute.

It is possible to understand the history of AI's adoption in the legal profession in three stages:⁷ first, tools were developed to streamline and automate certain routine tasks, including by way of workflow automation, e-discovery tools and data extraction tools; secondly, and much more recently, tools were developed to assist lawyers in more substantive ways, including in more interpretive tasks such as contract analysis and due diligence assessments; finally, the most advanced stage involves the use of big data for predictive analysis. The realm of legal argumentation (ie, the evaluation of how legal propositions are argued by lawyers and determined by judges) and its compatibility with AI assistance remains the most popular area of discussion for legal scholarship on AI.⁸

As will be discussed further below, it is important not to caricature the role that AI systems play and are likely to play in dispute resolution contexts. However, it suffices to say that the suggestion that AI might be capable of replacing human judges and arbitrators, and of rendering serious assistance in a wide variety of legal tasks, has absolutely been made and discussed in very serious scholarship.⁹

2 How AI Operates

Fundamentally, the kinds of AI systems being considered in this paper are “artificial neural networks”, a sub-category of “machine learning” AI systems. “Machine learning” refers to the fact that they develop their own means of “joining the dots” between input and desired output by sweeping through vast quantities of data and determining and testing the most optimal ways to spot correlations¹⁰ — these systems are also called “data-based systems”, as opposed to “knowledge-based systems”, as they rely on data at first instance to formulate their

⁷ See Christy Ng, “AI in the Legal Profession”, Chapter Three in Larry A DiMatteo, Cristina Poncibò and Michael Cannarsa, *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge, 2022) 35, 37.

⁸ Trevor Bench-Capon, “Thirty Years of Artificial Intelligence and Law: Editor’s Introduction” (2022) 30 *Artificial Intelligence and Law* 475, 476–7.

⁹ See, eg, Kevin Ongena, “AI Arbitrators... ‘Does Not Compute’”, Chapter Five in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 101, 103 [4]. This suggestion was discussed recently at panels in Australian Arbitration Week, including a session of the ACICA & CI Arb International Arbitration Conference on Monday 9 October 2023 entitled “Arbitration New Frontiers: AI, Sustainable Practices and Other Ideas” and presented by Professor John Swinson, Angelina Gomez, Kiran Sanghera, Schellie Jane Pryce and Daisy Mallett, and an event at Ashurst on 10 October 2023 entitled “Harnessing Generative AI: Large Language Models as Catalysts for Innovation in Arbitration” and presented by Luke Carbon, Lachlan McCalman and Natasha Blycha.

¹⁰ Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 27–8.

methodologies rather than human input per se.¹¹ “Artificial neural networks” are a common way of structuring AI programs that engage in this learning exercise that, notionally, imitates the structure of human neurons.¹²

Before proceeding with any attempt to explain how AI works, it should be noted that “AI explainability” is itself a topic of scholarly interest. As discussed further below, certain parts of AI systems (the “hidden layers”) are designed to operate without human supervision or instruction. These are the layers that “join the dots” between input and output. While it is possible to establish parameters and principles that guide these hidden processes, and it is possible, to a certain extent, to reconstruct what the machine has actually done to arrive at its conclusion, there is a very real sense in which these intermediate processes admit of no elaboration.¹³ For this reason, neural networks have been compared with quantum mechanics — both are conceptually virtually impossible to explain, but “work out” neatly from a mathematical standpoint.¹⁴ Of course, this inexplicability is at the centre of discussions concerning the ethics of AI use.

Fundamentally, and unless AI may be said to have transcended into the realm of being truly “brain-like”, AI is a form of computation, and thus a form of information processing.¹⁵ Information processing entails both an internal treatment of the information that is inputted and a way of outputting that information in a way that is intelligible and useful.¹⁶ The relevant question, therefore, is twofold: first, how does the AI system go from input to output; and secondly, how does the AI system know how to go from input to output?

The structure of artificial neural networks is inspired by the neurons of the human brain: particularly, the idea of the neuron as a fundamental neurological “atom”, and the notion of neurons forming a chain, wherein the information input into one neuron is the information output by its antecedent in the chain.¹⁷ Correspondingly, the “layers”, as they are called, of artificial neural networks only receive data from the layer immediately beneath it, and only transmit data to that immediately above it.¹⁸ The first layer is called the input layer, as it receives data from an external input (eg, a prompt in ChatGPT). The final layer is called the output layer, as it transmits data not to another layer but externally to the user. The intermediate layers of the artificial neural network are called the “hidden layers”, because, as foreshadowed above, their processes are not prescribed by (and are partially invisible to) the

¹¹ See Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 4–5 [8].

¹² Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 28.

¹³ See Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 1; Nathan Colaner, “Is Explainable Artificial Intelligence Intrinsically Valuable?” (2022) 37 *AI & Society* 231, 231.

¹⁴ Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 3.

¹⁵ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 4 [8].

¹⁶ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 4 [8].

¹⁷ Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart, 2020) 31.

¹⁸ Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 29.

user and the initial human programmer.¹⁹ When there are multiple such hidden layers, the AI is described as “deep” (hence “deep AI” or “deep learning”).²⁰

What happens in those intermediate layers is essentially the realm of algebra and probability. In each layer, the input value is subject to an “activation function” before it is passed on to the next layer.²¹ For example, in basic neural networks, the activation function may consist in two operations, or “parameters”: weight (multiplication) and bias (addition). Massive deep learning neural networks will contain billions or trillions of parameters.²² It is worth considering how this mathematical process might be applied to words, such as in large language models. Large language models develop their own vocabulary using numbers, wherein natural language words, or portions of words, are assigned their own numeral. Common words will have a numeral entirely to themselves. Less common words may not, and may instead need to be constituted through a combination of smaller “syllables” known as “tokens”. ChatGPT, for example, has a vocabulary of 50,257 tokens. Although one need not go so far as to say that the AI understands the meaning of individual words,²³ it is capable of “learning” semantic relationships between words (eg, synonyms) — called “embeddings” — such that a consideration of semantics may factor into what is otherwise just raw correlation.²⁴ There is also the ability of large language models to appreciate that certain words are more important than others in particular contexts. This is called “attention”: “self-attention” mechanisms enable these models to attend selectively to the key elements of an input; “cross-attention” mechanisms enable these models to alter what they generate in light of the most important elements of the input.²⁵

The second question is possibly more difficult: how does the AI *know* what steps to take to transition from input to output.

The answer is “machine learning”. Machine learning requires that an AI system begin with totally randomised parameters, that it will itself iteratively rewrite in order to process information in a useful manner in respect of certain kinds of data. It does this rewriting in advance of its use, in what is known as the “training” or “pre-training” stage. During this stage, the machine is fed data from which it is to detect and graph patterns.²⁶ In the case of large

¹⁹ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 8 [13].

²⁰ Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 34; Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 9 [14].

²¹ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 7 [13].

²² Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 1.

²³ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 12–13 [30].

²⁴ Gerhard Paaß and Sven Giesselbach, *Foundation Models for Natural Language Processing: Pre-trained Language Models Integrating Media* (Springer, 2023) 2.

²⁵ See, eg, Kejian Liu et al, “An Effective Personality-Based Model for Short Text Sentiment Classification Using BiLSTM and Self-Attention” (2023) 12 *Electronics* 3274, 2.

²⁶ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 6 [11].

language models, this data takes the form of “corpora” of natural language available online²⁷ — which may be the entire contents of Wikipedia or Google, and which may be subject to certain discrimina (eg, excluding material from certain online publications). During training, the AI makes use of “backpropagation”, which is a technique that enables machine-learning to correct its mistakes.²⁸ First, there is the “feedforward stage”, in which the machine proceeds from input to output according to its current parameters. Then, the machine compares its actual output with its intended output, and calculates the degree of “error” as between these two outputs. Lastly, the machine engages in some calculus operations (Leibnitz’s chain rule) to modify the current parameters in accordance with the degree of error. This process proceeds (propagates) from the “back” (output layer) to the “front” (input layer), and so rewrites the parameters in inverse order. A new set of working parameters have now been reached, and may themselves be tested in the same way. This also entails that the inclusion of more datasets motivates the machine to reappraise and optimise its approach. For this reason, machine learning relies on large datasets through which to pre-train and reach operable status.²⁹ How much pre-training is best is typically an open question governed by practical concerns.³⁰

In machine learning, there is a trichotomy between supervised, unsupervised and reinforcement learning.³¹ Imagine that one wants to create an AI that recognises inferior court decisions in which the obiter dicta of the superior court was not followed:

- a) In supervised learning, the datasets on which the machine trains are “labelled”. This means that a human will give the AI decisions that match the prompt *and* label for the AI the relevant piece of information (ie, the portion of the judgment that demonstrates a departure from an obiter dictum). The machine learns to correlate the two, such that when given a new decision that it has never before been exposed to, it knows what to look for. This is not to be confused with the phenomenon of “self-supervised learning”, wherein the machine tests itself by, for example, selectively hiding from itself certain words and trying to predict what words are missing. It may then reveal the hidden word to see whether it was successful.³² Properly understood, this is still a form of “unsupervised learning” (see below), as it occurs without human assistance.
- b) In unsupervised learning, the datasets on which the machine trains are unlabelled. This means that a human will give the AI decisions that match the prompt and the machine will itself determine the common denominator.³³ Ideally, this common

²⁷ Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 1.

²⁸ Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart, 2020) 32.

²⁹ Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 1–2.

³⁰ Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 31.

³¹ See generally Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart, 2020) 29; Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 30; Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 6–7 [12].

³² Melanie Mitchell and David C Krakauer, “The Debate over Understanding in AI’s Large Language Models” (2013) 120(13) *PNAS* e2215907120, 1.

³³ A provocative example is an artificial neural network designed to read minds: it absorbs data of brain activity in people who are looking at particular images, and by determining the correlation between the brain activity and the physical image, can accurately project what people are looking at by scanning

denominator will be that relevant portion of the judgment — however, in these “hidden steps”, the machine may instead choose any number of items to be common denominators, including ones with a spurious connection with the topic at hand, such as personal details of the parties or judge.³⁴ When these spurious correlations are nested within hidden layers of a deep learning AI, the fault may become even more obscure and irremediable; the variation of even a few trivial and ultimately inconsequential details may be enough to derail the neural network.³⁵ AI systems designed to use this learning method have the advantage of generally being able to analyse datasets at vastly superior speeds to supervised learning techniques, and therefore are able to achieve optimal accuracy more efficiently.³⁶

- c) Reinforcement learning is a form of semi-supervised learning, in which, as the machine begins to formulate strategies at finding correlations between datasets, a human programmer reviews the machine’s progress and provides feedback that the machine may take into account in future training. Therefore, if a machine outputs 30 decisions, only 28 of which actually depart from the relevant obiter, a human programmer can expose the two anomalies to the machine, which can motivate it to excise certain faulty correlations from its calculations. A subset of this model is the “generative adversarial network”, whereby that human review task is delegated instead to another neural network (the “discriminator”).³⁷

In summary, the following reflects the basic life cycle of an artificial neural network, such as is likely to be used in certain legal contexts:

1. The “skeleton” of the neural network is constructed, with input, output, and hidden intermediate layers.
2. The parameters that govern the operations at each layer are randomised.
3. Data is fed into the neural network, from which the neural network is “trained” to detect correlations and improve the viability of its parameters.
4. Once training is complete, data may be inputted at will, and the output will be calculated according to those parameters that were most successful in the training phase.

their brains, even if the image that they are looking at was never seen by the AI in the first instance: Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, 2016) 35.

³⁴ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 76–7 [7].

³⁵ Rembrandt Devillé, Nico Sergeysse and Catherine Middag, “Basic Concepts of AI for Legal Scholars”, Chapter One in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 1, 9–10 [15].

³⁶ See, eg, Qiantong Xu et al, “Self-Training and Pre-Training are Complementary for Speech Recognition” [2022] *IEEE International Conference on Acoustics, Speech and Signal Processing* 3030.

³⁷ Julija Kalpokiene and Ignas Kalpokas, “Creative Encounters of a Posthuman Kind: Anthropocentric Law, Artificial Intelligence, and Art” (2023) 72 *Technology in Society* 102197, 4.

II The Problem

The introduction to the issue of AI's adoption in the legal profession above provides a necessary background to any evaluation of its feasibility or desirability. However, having laid down this groundwork, it is prudent to pause and consider what exactly is being contemplated in allowing AI to make its way into the toolkit of legal practitioners. Too often is the resistance against AI couched simply in terms of technical problems with AI tools — bias;³⁸ inexplicability of results (the “black box” problem); “hallucinated” or spurious facts — design flaws that are interesting and serious, but that may eventually become insignificant as the technology is refined. However, obsessing over the technical details of AI can obscure (and has obscured) an earnest approach the overarching consideration: what are the implications of using AI to complete legal tasks?

To every limitation on the capabilities of AI that one may raise may be adduced the counterargument that AI can simply develop and eventually acquire the relevant capability. The point of this paper is not to refute this. The point is rather to suggest that certain capabilities lie, almost definitionally, beyond the scope of AI, such that even the most charitable view of AI should be incapable of sheeting home certain functions to it. And, correspondingly, the suggestion is that there is more to the world of the law, legal reasoning and adjudication of disputes than those who suggest that AI may soon be a viable replacement to humans would currently admit to.³⁹ This Part of this paper advances this argument in three parts: first, the current methods used to formulate AI systems, even if refined so as to minimise the undesirable “design flaws” highlighted above, are incapable of being applied to legal reasoning; secondly, even if we were to hypothesise the existence of AI systems that *perfectly* imitate or replicate the human mind, there would continue to be problems with implementing them in our legal order; and thirdly, these issues will manifest themselves even if the use of AI is overseen with caution by humans, in what we may term a “responsible” manner.

1 AI as Ill-Suited for Legal Reasoning

Principal among the objections to the deployment of AI in the legal profession might be the view summarised by Matthias Van Der Haegen as follows: “[I]aw can simply not be reduced to an equilibration and calculation of variables and factors with predetermined weights”.⁴⁰ Impressive and complex as machine learning-based AI systems are, the calibration of numerical variables with predetermined weights essentially defines how they operate. Unless

³⁸ Bias may creep into an AI system at various junctures: at the training stage, in the collection of data for analysis, or in the very architecture of the algorithm: Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 86–7 [26].

³⁹ A similar caveat is carved out by Michele Taruffo, “Judicial Decisions and Artificial Intelligence”, Chapter in Giovanni Sartor and Karl Branting (eds), *Judicial Applications of Artificial Intelligence* (Springer, 1998) 207, 213.

⁴⁰ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [20].

it should be thought that the law *is* capable of being so reduced, there is a fundamental disjunct between the methodology of AI-based “reasoning” and legal reasoning.

Even if it is thought possible to reduce certain legal propositions to such formal logic, there are certain incontrovertible features that they will invariably possess that seem unamenable to AI-assisted reasoning. For example, legal propositions can be amended or overruled, often changing the entire landscape of an area of the law overnight, so to speak⁴¹ — scholars have referred to the Court of Justice of the European Union’s decision in 2018 in *Achmea* as an example of such an occurrence.⁴² Ultimately, this issue relates to the idea articulated above that the law cannot be considered as consisting entirely in the sum of legislation and judicial decisions that have been put in writing and made available on the internet. The possibility for the law to move beyond that which has already been said, whether through a dramatic and sudden overruling of a legal proposition, or a more natural (but no less important) judicial development, is immanent. Algorithmic legal prediction is inherently backward-looking, and designed to replicate and conform with the most popular historical statements of the law. It should be taken as self-evident that the “correct” or “just” solution to a legal problem is by no means always the historical or popular answer to that problem. If AI is widely adopted, it may have serious consequences insofar as it contributes to the “sclerosis and ossification” of the law.⁴³ To quote Kevin Ongenae, a researcher in dispute resolution: “If ‘law is a living thing’, AI arbitrators might kill it”.⁴⁴

Legal rules are also by their nature ambiguous and open-ended⁴⁵ — an AI’s ability to replicate an inquiry into, for example, causation, must either be capable of the flexibility with which a human judge would approach the subject, or alternatively make such shortcuts as necessary to reconstitute the open-ended rule in a more rigid, logical structure.⁴⁶ More

⁴¹ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [20]. Van Der Haegen’s counterargument (at 84–5 [22]) that such reversals are rare in no way minimises their fatality to the effectiveness of AI tools, and in any case fails to acknowledge that even subsequent decisions that agree and consolidate existing areas of the law can fundamentally change the way in which certain legal principles are perceived.

⁴² Kevin Ongenae, “AI Arbitrators... ‘Does Not Compute’”, Chapter Five in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 101, 109 [19], citing Maxi Scherer, “Artificial Intelligence and Legal Decision-Making: The Wide Open?” (2019) 36(5) *Journal of International Arbitration* 539, 557.

⁴³ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 86 [24].

⁴⁴ Kevin Ongenae, “AI Arbitrators... ‘Does Not Compute’”, Chapter Five in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 101, 108 [18].

⁴⁵ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [20]; Bart Verheij, “Artificial Intelligence as Law: Presidential Address to the Seventeenth International Conference on Artificial Intelligence and Law” (2020) 28 *Artificial Intelligence and Law* 181, 187.

⁴⁶ Although this is not to deny that certain areas of the law are, naturally, better suited for codification than others: Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [20]. See further Bart Verheij, “The Toulmin Argument Model in Artificial Intelligence: Or: How Semi-Formal, Defeasible Argumentation Schemes Creep into Logic”, Chapter 11 in Iyad Rahwan and Guillermo R Simari (eds), *Argumentation in Artificial Intelligence* (Springer, 2009) 219, 219–20.

ambiguous yet are the series of fundamental principles — of fairness, reasonableness, commonsense, etc — that might be said to undergird the law.⁴⁷

It is prudent also to be cautious about the term “prediction” as it is used in the context of predictive AI tools. Human observers naturally describe an AI tool as “predicting” the outcome of a case or “guessing” what a user might want based on their question, but this is only because we intuitively understand a temporal and causal relationship between the facts of the case and the legal outcome of the case, or between a request for information and the granting of that information. From the perspective of the AI, however, it is not so much that it, say, concludes that an offender will (in the future, conjecturally) likely reoffend, as that it concludes that an offender is (in the present, as a matter of fact) likely to reoffend.⁴⁸ It is also worth noting that “prediction” in more scientific contexts, such as “weather prediction”, carries with it a normative quality — ie, the prediction makes a claim to being the “correct” (or to being likely the “correct”) answer.⁴⁹ That same connotation should not be uncritically applied to legal contexts, in which a prediction as to how, say, a court would likely resolve a claim should possess no such normative quality. This is why there is something jarring about the use of the word, “accuracy”, in the context of AI prediction or AI decision-making. What does it even mean to be “accurate” at determining the answer to legal questions?⁵⁰

Unless the “accuracy” of those AI systems discussed above, such as the quantitative legal prediction tool, were to reach 100%,⁵¹ the allowance of a margin of error of even 1% or 2% can be dangerous and pernicious. In response to this, it is often said that bias is a feature also of human judges, who, though they may cloak their decisions with legitimate factual and legal arguments, may ultimately arrive at their decision due to, for example, prejudicial views about a party. This relates to the so-called “hungry judge effect”, which, while it has successfully entered the *zeitgeist*, has by no means received unanimous support among social scientists and legal academics.⁵² However, it is submitted that there is a fundamental difference between the mathematical *fact* of a 1–2% bias, and the *potential* of bias in any judge. For an offender who is sentenced according to the former metric, their sentencing *is*, as a matter of fact, afflicted by bias. For an offender who is sentenced by the latter (the judge), there is the *potential* that the judge may, in their conscious and subconscious reasoning processes, be affected by bias. I suggest that this is not a trivial or a cynical distinction: in the former case, society is deliberately subjecting one of its members to a system that *will*, as a matter of fact, exercise a known degree of prejudice. This inevitability is what makes this idea

⁴⁷ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [20].

⁴⁸ A similar point is made by Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [19].

⁴⁹ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 83 [19].

⁵⁰ Simon Deakin and Christopher Markou, “From Rule of Law to Legal Singularity”, Chapter One in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Hart, 2020) 1, 7.

⁵¹ A hypothetical scenario discussed below.

⁵² See especially Konstantin Chatziathanasiou, “Beware the Lure of Narratives: ‘Hungry Judges’ Should Not Motivate the Use of ‘Artificial Intelligence’ in Law” (2022) 23 *German Law Journal* 452, launching a powerful criticism of the original study and its misuse (or uncritical use) subsequently by lawyers and legal theorists.

repugnant to any concept of justice. Nor should bias simply be seen as merely a technical issue that goes to the accuracy or effectiveness of an AI tool. Fairness is a fundamental feature of the law — a tool that vitiates the fairness of legal process might rightly be said to vitiate and frustrate the very reason why that tool was deployed in the first place.⁵³

2 Hypothesising a Perfect AI

However, let us turn now to the first hypothetical scenario: notwithstanding the impossibility of making this kind of assessment, imagine that AI systems attain 100% accuracy, and become capable of producing predictions or answers to legal questions that are the same as those that a human might produce. While this is a fanciful hypothetical, it should be said that the accuracy, particularly of tools that predict the outcomes of judicial decisions, is astounding.

It has been suggested that, should AI tools be capable of operating at such a perfect level of accuracy, the method by which they achieve such an accurate result should be irrelevant. For instance, Van Der Haegen argues that:⁵⁴

A valid comparison between the process of algorithmic and human legal prediction can only be made by comparing their outputs. If a human assessment of the outcome of a legal dispute is indistinguishable from an algorithmic prediction, the manner in which they came about is irrelevant.

It should be said that there appears to have been a surreptitious shifting of the goal posts in respect of AI's capacities. The promise of artificial *intelligence* is that it should somehow transcend the confines of traditional computing such as enables it actually to *understand* (*intellegere*) the inputs and outputs in which it is dealing. An AI that is capable of cognition, contemplation and reflection would appear to be well on its way to attaining a kind of sentience or consciousness, which, as opposed to computation, is the unique quality that we value in the human mind. To say that AI cannot *really* understand, to any meaningful extent, that which it generates violates this promise — it should not be open to promoters of AI then to say that true intelligence was never truly relevant in the first place, or to say that our understanding of intelligence is clearly outdated and must be reappraised so as to include whatever AI is doing. That reasoning is circular.

Notwithstanding this, it is suggested that a wilful ignorance of the process by which AI brings about its decision leads to a dangerous oversimplification of the issue.

If AI simply brings about an end product through “mimicry”, it is a valid question to ask what it is mimicking. As discussed above, AI operates by analysing the massive quantities of data available online. That provides a hard limit on the capabilities of an AI; such that if the dataset should be thought incomplete, the AI will similarly be incomplete. What, then, is missing from the data available online? In the legal context, it is suggested that the law, and

⁵³ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 86 [26].

⁵⁴ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 84 [21].

attendant concepts of justice, reasonableness, proportionality, etc, does not reside online within the text of the various “sources” of law. What is online is merely the reflected image of certain humans’ attempts at distilling and applying the law in various contexts. More broadly, if humanity and emotion should be thought at all relevant to the law, we might ask whether the text available on the internet is capable of capturing the totality of human experience. An AI that is weened solely on this might be capable of mimicking *something* about human experience or the law, but not those things in themselves. It is by no means obvious that AI comes “close enough”, because we have never had occasion to consider this question in practice. All humans have human experiences — indeed, that is all we have, and every second while conscious only contributes to this.

This should not be seen as a denigration of the power of mimicry in learning: it is essential for human learning as well. However, there is a reason why humans are capable of becoming proficient in some topic without sifting through and taking stock of millions of examples. The human capacity for abstraction far exceeds that of any machine.⁵⁵

Consider, for example, the so-called “Infinite Monkey” theorem, that holds that, given an infinite number of monkeys bashing away aimlessly at a typewriter, it is almost certain⁵⁶ that one will type out the full works of William Shakespeare. Would we truly say that those works, as produced by the monkey, possess the same aesthetic value as the original, in which the author *intended* that his collocation of words produce a certain effect? Would the answer be different if we were to grant that the monkey that did so knew how to type individual letters, and could appreciate that, for example, a “z” ordinarily does not follow an “f” — ie, if it had some sense of how letters interact with one another? Intuitively, we might say that the work of the monkeys, as it involved no creativity or skill, is less easily deemed a work of art or creativity than the original work of Shakespeare. But that is by no means *the* correct answer. For instance, New Critics or followers of Roland Barthes might hold that the intent of the author (ie, Shakespeare or the monkey) should be irrelevant to our evaluation of the final work.⁵⁷

Wherever one lands in respect of the above questions, it should be clear that a work of art, intended primarily for aesthetic evaluation and emotional responses from an audience, is a fundamentally different thing to, for example, a judge’s order. The distinguishing characteristic is the fact that the latter is not a statement of or comment on the world, but a normative pronouncement as to what should occur. It is, to borrow the terminology of linguistics, a “speech act”: an expression that not only describes, but performs a function as well. Speech acts are common features of daily life. For example, when one says “sorry”, the point is not just to convey the utterance from speaker to listener, but to effect an apology, with a demonstration of regret and a presumptive request for forgiveness. The idea that an act may only have force or value if accompanied by an attendant intention should be a familiar concept to lawyers: *actus reus non facit reum nisi mens sit rea*. Accordingly, the word “sorry” only succeeds as a speech act if the speaker is truly sorry while saying it. Consider, for instance, that someone has wronged this author, and emailed “sorry” by way of apology; and consider

⁵⁵ See Christopher Markou and Simon Deakin, “*Ex Machina Lex: Exploring the Limits of Legal Computability*”, Chapter Two in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Hart, 2020) 31, 52.

⁵⁶ In the mathematical sense, referring to what is essentially a certainty when dealing with an infinite data set.

⁵⁷ See generally, on the topic of evaluating creativity in the context of AI, Julija Kalpokiene and Ignas Kalpokas, “Creative Encounters of a Posthuman Kind: Anthropocentric Law, Artificial Intelligence, and Art” (2023) 72 *Technology in Society* 102197, 3.

that this author accepts that apology. If someone were subsequently to reveal that the email was only sent by mistake — an autocorrection or a “pocket text” — the whole force of the apology would instantly be nullified.

Legal pronouncements are speech acts of a somewhat more significant nature, in that the function that they perform is to justify the deployment of coercive state power in punishing an individual or resolving a dispute. What gives the order of a court its force is not simply the coherence of the order’s expression or the mode according to which it is formatted: it is the fact that it reflects the intention of society through its proxy, the judge of the court.⁵⁸

Consider, similarly, the scenario whereby parties adduce arguments with the knowledge that one party’s arguments are more intuitive and conform more to a conventional view of the law, whereas the other party’s arguments, while still having the potential to prevail, are less intuitive and require a greater leap on the part of the decision-maker. It might rightly be said of judges deciding such a dispute that 99% of them would prefer the more intuitive argument. However, this *post hoc* statistical analysis of judgments is quite distinct from the idea that one argument is, *a priori*, 99% more likely to be correct. In such cases, it was the fact that the judge *could have* made the alternate decision, but chose not to, that gives credence and validity to the ultimate decision, and to the entire process surrounding it.⁵⁹ It is quite clear, by contrast, that if an AI were to have an element of “randomness” to it — as some do, via what is known as a “temperature” setting — such that it opts for the less-intuitive answer in every hundredth case, that would not be considered a meaningful and just decision, precisely for the reasons discussed above in relation to “speech acts”. This is why creating an “accurate” AI that mimics, but does not fully replicate, human reasoning can only take us so far.

A well-known problem with complex machine learning-based AI systems is the so-called “black box” problem, whereby the processes that the system undertakes in the “hidden layers” between input and output are invisible and inexplicable even to the human computer scientists that created the AI in the first place.

It is often presented by way of rebuttal to this criticism that humans possess the same kind of black box problem, in that the myriad of conscious and subconscious factors that lead to decisions are invisible, even to the human that is making them.⁶⁰ The similarity between this human “black box” and the black box of AI systems is, however, illusory. Inspiration, spontaneity and inexplicability in the context of human decision-making is something with which all humans are intimately and innately familiar. This familiarity is not taught or learnt, but *experienced* by human minds from the very beginning of our lives. And whereas subconscious processes are, as the name suggests, imperceptible to our conscious experience; the *fact* that subconscious processes work an effect on conscious decision-making, and the way in which they do so, is itself a phenomenon of consciousness that all have experienced. This is all to say that the inexplicability present in a human decision does not vitiate the quality of that decision — it is an inseparable, essential feature of that decision, that is readily appreciated

⁵⁸ See Frank Pasquale, “Foreword” in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Hart, 2020) v, vi.

⁵⁹ See Michele Taruffo, “Judicial Decisions and Artificial Intelligence”, Chapter in Giovanni Sartor and Karl Branting (eds), *Judicial Applications of Artificial Intelligence* (Springer, 1998) 207, 208.

⁶⁰ See, eg, Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 94 [39].

by all who might be affected by it. By contrast, there is simply no such intuitive and innate sympathy for machine insolubility.

Nor is it true that the law *only* interests itself in the post hoc justification that judges provide for their decision: ie, what we term “reasoning” or “the judgment”. Imagine if all difficult decisions were made by the roll of a die, with judges writing their judgments after the fact so as conform with whichever party happens to have won. While, in practice, there would be no way of peering into a judge’s chambers and exposing this method, it is clearly and essentially repugnant to our concept of justice. Obversely, consider if a judge were to arrive at a decision by legitimately considering the parties’ claims, but then task an AI with reverse engineering some reasoning for the relevant outcome. This would be a perverse *hysteron proteron* that violates the base purpose of giving reasons. While this is not to say that reasoning necessarily always precedes the decision-making process in the mind of the judge, a complete disjunct between the “intelligence” that conceived of the answer to a question and the “intelligence” that joined the dots to that answer from the original facts is undesirable.⁶¹

Finally, what of the possibility of a neural network achieving its goal of entirely recreating a human mind? (It must be stressed that this is beyond contemplation with today’s technology, and indeed does not truly reflect the end goal that AI developers are currently working towards). With an AI of this kind, not only would the “outcome”, for example, a judgment, be correct, but it would be arrived at using the same steps as a human mind would take to reach that outcome. By contrast to the hypothetical discussed above, there is no mere mimicry of human reasoning, but rather a perfect recreation of it.

For such an AI, that is structurally and functionally identical to a human brain, no obvious⁶² criticism can be mounted at the ability of the AI to understand legal propositions, to exercise “good judgement” or “common sense” in arriving at an answer, or to make sufficient use of emotion in doing so. Its abilities would, *prima facie*, be identical to those of any human. That being the case, why should the “shell”, fleshy or mechanical, in which this brain is situated have any bearing on our trust of its abilities?

The answer is that, in this author’s opinion, it simply *will* have a bearing (ie, that people will mistrust the decisions of even such a machine when compared to those of a human), and that this is a matter of human intuition, perhaps bred particularly of the evolutionarily-instilled value that we place on human life. While this intuition may, to a certain extent, be labelled irrational or unjustifiably anthropocentric, it is something of a paradox to criticise anthropocentrism in the law. The law is, by almost any metric, designed to regulate and give structure to the conduct of humans in human society. It is, by definition, anthropocentric; and as was alluded to above, its “humanness” is an essential aspect of it. While this, it should go without saying, does not preclude the use of non-human tools, including AI, in administering justice and giving effect to the law, the complete abdication of decision-making power even to a perfectly “human” non-human machine is jarring. Finally, while we speak about such very interesting and worthwhile topics as the legal status of AI, whether it might be thought of as a legal person, capable of exercising rights and having those rights vindicated by the law, it

⁶¹ Michele Taruffo, “Judicial Decisions and Artificial Intelligence”, Chapter in Giovanni Sartor and Karl Branting (eds), *Judicial Applications of Artificial Intelligence* (Springer, 1998) 207, 210.

⁶² Only “obvious” because this presumes a physicalist view of the human brain; which, while intuitive and persuasive to many, has by no means been deemed a scientific inevitability by philosophers. As was argued above, the fact that there exist such debates among philosophers should in itself give pause to those who would seek to recreate a human mind.

should be borne in mind that these questions arise as a *coincidence* of the development of AI. AI is not being researched and developed for the purposes of elevating the status and improving the treatment of machines in society. It is being developed to serve as a tool for use by humans. Regardless of one's views about this idea — it is easy to allude to similarities between this mindset and that of colonialism, or of the exploitation of non-human animals — this simply describes the reality about the development of AI. In this context, to refer to an unwarranted anthropocentrism in connection with AI in the law is simply misguided.

3 The Extent of AI “Interference”

Much of the above has considered the implications of an AI making decisions in lieu of a human judge, a possibility that is seriously being considered by legal scholarship. However, it is important to bear in mind that there is a spectrum of use being contemplated for AI in the legal profession. On one extreme end of the spectrum, there is the concept of AI making decisions without any human input whatsoever. On the other end, however, there are trivial uses of AI that are already commonplace and invisibly integrated into our daily lives, such as a word processor that suggests altering one's written expression for clarity or brevity: if such a technology were used to assist in the drafting of, say, a sentence in a party's submission, it is unlikely seriously to impinge on the quality of the justice that is ultimately done. While these are extreme examples, they illustrate that the relevant question is not *whether* the tasks of judges and lawyers should be devolved to AI, but *to what extent* they can be devolved before becoming incompatible with our understanding of justice and law.

Work has been done by scholars and institutions to delineate the limits of what is and is not permissible in terms of the use of AI by legal practitioners and decision-makers. For instance, the Silicon Valley Arbitration and Mediation Center (SVAMC) has published draft “Guidelines on the Use of Artificial Intelligence in Arbitration”, of which a key provision is that the arbitrator must not abdicate their decision-making function to an AI.⁶³ It has also been suggested, for example, that AI-based legal predictions should occupy a *sui generis* position in the taxonomy of factors that contribute to the making of a reasoned legal decision — ie, that they not be considered statements of law or statements of fact — which, if properly borne in mind by all parties, appears to be a necessary way of treating AI assistance.⁶⁴

However, whereas the utility of formal safeguards such as this is self-evident, it should not be thought that imposing these such rules and guidelines resolves all issues about the use of AI in dispute resolution proceedings. An image of AI as simply another tool in the arsenal of lawyers, that can be put to use from time to time without consequence, is an illusory one. Indeed, as has often been proclaimed, AI will not just be like any other technology; it promises

⁶³ The draft guidelines of 31 August 2023 are accessible at <https://the Arbitration.org/wp-content/uploads/2023/08/SVAMC-AI-Guidelines-CONSULTATION-DRAFT-31-August-2023-1.pdf>. As of the time of writing, the SVAMC is still accepting comments on the Guidelines.

⁶⁴ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 85 [23].

to infuse all aspects of our personal and professional lives.⁶⁵ And so, as a consequence, while the nature of the law is certainly influencing the ways in which AI is being developed, the development of AI is influencing also the nature of the law.

In other words, it is one thing to comment on whether the use of AI has been “responsible”; which is to say that it is possible to hold its users to account according to some societally-agreed standard. It is quite another thing to comment on whether the use of AI has real consequences for the way in which proceedings are conducted.⁶⁶

For instance, humans possess well-understood biases in favour of quantitative data, and data produced by mathematical models, by means of such phenomena as automation bias and the so-called McNamara fallacy.⁶⁷ As such, even if a judge approaches an AI-generated figure, such as a calculation that an accused has a 70% chance of reoffending, with the circumspection that is required of them in any relevant legislation or guidelines, the very existence of such a statistic works invisible and irreversible impacts on the psychology of that judge. It should also be noted that AI-generated text is frequently described as sounding very persuasive or seeming correct, even when it generates objectively incorrect answers. The reason for this is simply the model by which it generates such text, namely by producing the most *likely* string of words. In this way, it is essentially able to withstand initial human scrutiny by means of deception.

Similarly, consider the scenario, which is already now becoming commonplace, whereby a lawyer or judge leaves it to a generative AI tool to produce a legal document on the basis of certain prompts. Provided that the lawyer or judge personally and cautiously reviews the text that is generated, it is difficult to see how there could be any irresponsibility on the part of that lawyer or judge. However, this obfuscates the influence that the generative AI must obviously have had on the process. Undoubtedly, if the lawyer takes it upon themselves to produce a document “from scratch”, the process of researching and collecting cases, reading them, schematising and writing out a legal argument would shape to a considerable extent the final product. The generation of text and the review of that text occurs simultaneously. Employing AI-generation turns this into a dichotomous, seriatim process. Of course, practical concerns, such as efficiency, often necessitate that there be some delegation of responsibility from, for example, lawyer to paralegal, or judge to associate. But this is a reality of which the entire legal profession is aware and with which it has been familiar for centuries. We can understand the process by which a senior lawyer might explain a task to a junior lawyer, review their work and take questions while the task is being completed, and finally be able to discuss the end product, which includes what we might term a “meta-discussion” whereby the junior lawyer explains how they arrived at certain conclusions. While technological innovation may enable a very sophisticated user interface (UI) as between an AI system and its user, it is

⁶⁵ Jean-Marc van Gyseghem, “Fundamental Rights and the Use of Artificial Intelligence in Court”, Chapter 14 in Larry A DiMatteo, Cristina Poncibò and Michael Cannarsa, *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge, 2022) 257, 258.

⁶⁶ See Michele Taruffo, “Judicial Decisions and Artificial Intelligence”, Chapter in Giovanni Sartor and Karl Branting (eds), *Judicial Applications of Artificial Intelligence* (Springer, 1998) 207, making a similar plea: at 218.

⁶⁷ Matthias Van Der Haegen, “Quantitative Legal Prediction: The Future of Dispute Resolution?” Chapter Four in Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia, 2021) 73, 85 [24].

difficult to see how this could rise to the level of nuance expressed in human interactions such as those discussed in the example above.