



AI, Judges and Judgment: Setting the Scene

Citation

Buckland, Robert. "AI, Judges and Judgment: Setting the Scene." M-RCBG Associate Working Paper Series 2023.220, Harvard University, Cambridge, MA, November 2023.

Published version

<https://www.hks.harvard.edu/centers/mrcbg/publications/awp/awp220>

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37377475>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access License Articles (IOAL), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)



HARVARD Kennedy School

MOSSAVAR-RAHMANI CENTER
for Business and Government

AI, JUDGES AND JUDGEMENT: SETTING THE SCENE

Rt Hon Sir Robert Buckland KBE KC MP
Harvard Kennedy School

November 2023

M-RCBG Associate Working Paper Series | No. 220

The views expressed in the M-RCBG Associate Working Paper Series are those of the author(s) and do not necessarily reflect those of the Mossavar-Rahmani Center for Business & Government or of Harvard University. The papers in this series have not undergone formal review and approval; they are presented to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

AI, JUDGES AND JUDGEMENT: SETTING THE SCENE.

A Paper by RT HON SIR ROBERT BUCKLAND KBE KC MP

Artificial intelligence (AI) in the administration of justice is growing at rapid pace.¹ This is driven by widespread recognition of AI justice's undeniable advantages, despite the risks it presents to the integrity of legal systems.

AI justice may, for example, lower the administrative burden of cases. The Crown Courts in England and Wales ended 2022 with a near-record load of over 60,000 outstanding cases.² AI can dramatically increase court efficiency and reduce backlogs, providing standardised outcomes faster and at lower cost. After all, AI judges do not need to rest. At the same time, AI-driven judicial decision-making could make justice more accessible to the large segments of society that cannot afford human lawyers.³

Proponents also argue algorithms could improve the fairness of judgements because "AI judges strictly follow precedents, restrict improper judicial discretion, prevent personal biases and preferences of individual judges, handle large amounts of information, complete complicated calculative balances, and discover statistical representations of variations of fact patterns and legal factors".⁴ Even where AI tools assist human judges, these tools can push relevant legal provisions through comprehensive data retrieval. This in turn can improve judges' understanding of cases, helping them avoid one-sided access to data and information.⁵

At this point, it is important to clarify the different ways in which AI is being deployed in the courtroom. At a foundation level, AI may be used for auxiliary administrative functions. This includes communication between judicial personnel, allocation of resources and cases, and ensuring the anonymisation of judicial decisions, documents, or data. These activities may ostensibly appear separate from the core of judicial decision-making but carry subtler implications. For instance, the allocation of a case to a specific judge, given their unique expertise or biases, could indirectly influence the outcome. These nuances notwithstanding, the primary objective of these AI-driven tasks remain administrative in nature, aiming to streamline the judicial process rather than directly determine case outcomes.

¹ Adrian Zuckerman, "Artificial Intelligence – Implications for the Legal Profession, Adversarial Process and the Rule of Law", *UK Constitutional Law Association*, March 10, 2020, <https://ukconstitutionallaw.org/2020/03/10/artificial-intelligence-implications-for-the-legal-profession-adversarial-process-and-the-rule-of-law/>.

² Dominic Casciani, "Chronic backlog of serious-crime cases hits courts", *BBC*, February 9, 2023, <https://www.bbc.co.uk/news/uk-64586483>.

³ Zuckerman, "Artificial Intelligence".

⁴ Jinting Deng, "Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System", *Georgetown Law Technology Review* 223, no. 3 (2019): 279.

⁵ Zichun Xu, "Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities", *Applied Artificial Intelligence* 36, no. 1 (2022).

Although I make occasional reference to these applications of AI in the courtroom, they are not the focus of my analysis. I focus instead on i). AI judicial tools that human judges could use to help make their decisions; and ii), fully automated judicial decision-making, in which the court is disembodied and replaced by an AI judge.

Most, if not all, current instances of AI judicial decision-making fall within the former category – that is, AI serving as a judicial aide. Predictive algorithms are used in a variety of judicial settings. Their most well-known applications are in predicting pre-trial flight risk and recidivism.⁶ These algorithms, like COMPAS in the US, assist in establishing whether and what amount to set bail or in informing decisions about sentencing and parole.

The latter category represents a significant shift in moving towards a fully automated system where AI judges adjudicate cases without human intervention. The Estonian Ministry of Justice, for example, is exploring whether to introduce ‘robot judges’ that could adjudicate small claims disputes of less than €7,000.⁷ The development and deployment of these judges is probably inevitable. This makes identifying and addressing the weaknesses embedded within AI justice ever more important.

Before discussing the application of AI in judicial decision-making it is important to describe what I mean by human judgement. An analogy for this is the biblical tale of King Solomon, who, when faced with two women claiming to be the mother of a child, proposed to split the child in two. One of the women begged the King to spare the child and to allow the other woman to have it, whereas the other woman agreed to the child being cut in two. This revealed the true mother as the woman who would rather give the child up than see it harmed.

The Judgment of Solomon was not an agreement to divide an asset down the middle as often happens in divorce financial settlements but was the emotionally intelligent response to the words and deeds of a witness. Solomon, rightly, decided that only the true mother would consent to her child being given away, because the child’s life was precious to her above all things. This was an assessment of credibility based upon a shared understanding of basic human emotions. The “law”, insofar as it existed, was based upon Solomon’s experience and qualities as a leader.

Legal judgments are based upon evidence and rules, but is that all? When I was trained as a part time Crown Court Judge, at the end of my course at the Judicial College, the then Lord Chief Justice of England and Wales, Lord Judge, made some remarks in which he reminded all of us that in our work

⁶ Karen Hao and Jonathan Stray, “Can you make AI fairer than a judge? Play our courtroom algorithm game”, *MIT Technology Review*, October 17, 2019, <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>.

⁷ Eric Niiler, “Can AI Be a Fair Judge in Court? Estonia Thinks So”, *Wired*, March 25, 2019, <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>.

and our judgments, we should not lose sight of our humanity; in other words, our experiences as human beings, as opposed to our training as lawyers. This was a reminder that, although the law is there to be applied, judicial discretion is shaped not just by our legal training and experience, but our experiences as humans.

Knowledge is a generic term that takes different forms. Knowledge can be pure and empirical, to use Kant's distinction in the *Critique of Pure Reason*.⁸ Pure knowledge, which can be acquired before use such as through rules and laws, contrasts with empirical knowledge, which can only be gained from human experience.⁹ The former is absolute, universal even, whereas the latter is confined by our own experience.¹⁰ Judgement, therefore, will be either analytical, based upon hard facts where nothing more need be added, or synthetical, based upon arguments where something extra is then added.¹¹ In other words, the power to make determinative judgement is based on universal concepts, but judgement based upon particular information is reflective judgement.

Practical judgement involves the enactment of principles, demanding that a particular action be taken. A good example of the use of practical judgement is a decision whether to be truthful to another person, when telling the truth will upset their feelings. Reflective judgment applies to moral dilemmas, like conflicts of loyalty. In this context, the aim shouldn't be complete convergence or uniformity, but a variety of different responses within an ethical framework. Each answer or use of judgement will be based upon the situation that presents itself. If we are to effectively utilise AI in judgments, we must first seek an alignment between human thought processes and the AI that might replace them.

It seems to me that discriminative AI could deal well with decisions requiring determinative judgement, requiring the use of data to come to a defined outcome. For instance, a discriminative model can be trained to tell whether a given email is "spam" or "not spam" based on its content. These models are typically straightforward and are trained directly on the task at hand. They excel in making decisions based on features of the data and their strength lies in identifying boundaries or distinctions between categories. However, these models do not offer a critical view of the data or underlying structures; they simply work by recognising patterns and differentiating among them.¹²

It is generative AI and its potential impact on practical judgement that is the most interesting topic in my mind. Generative models capture the underlying structure of the data and can produce new data

⁸ Immanuel Kant, *Critique of Pure Reason*, trans Kemp Smith (London: Macmillan 2007)

⁹ Kant, *Critique of Pure Reason*

¹⁰ Ibid.

¹¹ Ibid.

¹² See: Enrico Santus, Nicolas Christin and Harshimi Jayram "Technology Factsheet Series, Artificial Intelligence" (Harvard Kennedy School 2020)

samples (or decisions) which resemble the original dataset. They understand and learn the underlying patterns and distributions of the input data. How might these generative models align with or even redefine the nuanced role of a judge in court?

Judges: what do they do and what are they for?

The task of modern judges is in many ways an unenviable one. Generations later, we have a deep and complex web of statutes, conventions and regulations that cover our jurisdictions. Why then, shouldn't we introduce technological change to help them navigate this maze and to allow them to use their human qualities in coming to reasoned judgments?

England and Wales: a case study.

The coronavirus pandemic resulted in a quantum leap in digitalisation and remote hearings. Although the technology is not new, the widespread use of remote links - and not just for administrative hearings - has allowed judges to see in-person hearings are not always the best option. Although the debate into whether there is a qualitative difference between evidence in person or evidence via video link continues, there are few studies to suggest remote evidence is not achieving best evidence.

The use and spread of digital means of justice, as seen in England and Wales for Guilty pleas, straightforward traffic offences, and money claims system in civil cases, raises the question of whether fully qualified judges are necessary to adjudicate every case. A court clerk, if properly trained, or a sophisticated algorithmic model, could potentially address the thousands of formulaic matters.

How do we then determine the cases for which we need judges for, and could an artificial intelligence model feasibly provide a wider skill set and world view?

From my own experience as a Crown Court Recorder sitting mainly in the Midlands of England, I can describe the life and work of a judge. It is a lonely calling. Those who become full-time judges will leave their former legal profession behind; their professional relationships changed, permanently. From the moment a judge steps into their chambers and closes the door, a palpable sense of isolation sets in. It is your court, and you call the shots. With that power comes huge responsibility, however, that can sometimes weigh extremely heavily.

In developed "rule of law" jurisdictions, judges are not only independent of other arms of the state but are independent of each other. Their decisions will be theirs and theirs alone, even if now and again judges may discuss issues arising in specific cases with their colleagues. The judge sits at the epicentre of their own universe, with court staff helping them and the parties or their representatives to interact with. When a jury is involved, the judge has the responsibility for their welfare and

guidance as to the law, and it is a job that requires great concentration, energy, and communication skills. Being a judge is so much more than having a knowledge or mastery of the law and procedure.

Human intuition, emotional intelligence, sheer common sense: these are all an essential part of the makeup of a good judge. In other words, judgement isn't merely about the legal or factual judgement that judges must come to, but the myriad of small decisions that, taken together, make sure that the administration of justice runs well. One of the main criticisms of human judgement and its application to judicial decision-making, however, as referred to above, is the existence of bias, either overt or covert. This is a problem that has been intrinsic to judicial decision-making over the generations.

Bias

One of the greatest strengths of AI's application to judicial decision-making might also be its most serious weakness. Many argue that AI, if coded and implemented correctly, could reduce, or even eliminate the biases of human judges.¹³

The 'Hungry Judge' effect – that a hungry judge is a stricter one – is commonly adduced in favour of this argument; one study found that judges issue more lenient decisions after a meal.¹⁴ Subsequent studies have observed how other extraneous variables can influence judges, too. They include the weather,¹⁵ the performance of local sports teams,¹⁶ and a defendant's mugshot.¹⁷ There are then also disparities between judges (and their respective adjudications), which can be considerable. In a case of 'refugee roulette', for instance, one American judge granted asylum to only 5% of Colombian applicants, while another granted it to 88%.¹⁸ It is not far-fetched therefore to suppose that the rigid application of law by an AI judge or judicial assistant would be conducted in a more rational, deliberative, and consistent manner. If anything, as Sunstein notes, these algorithms could help judges identify and attenuate the cognitive biases that sway their decision-making.¹⁹

¹³ See, for example, Monika Zalnieriute, Lyria Bennett Moses, and George Williams, "The Rule of Law and Automation of Government Decision-Making", *Modern Law Review* 82, no. 3 (2019): 425.

¹⁴ Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions", *PNAS* 108, no. 17 (2011): 6889.

¹⁵ Anthony Hayes and Soodeh Saberian, "Temperature and Decisions: Evidence from 207,000 Court Cases", *American Economic Journal* 11, no. 2 (2019): 239.

¹⁶ Ozkan Eren and Naci Mocan, "Emotional Judges and Unlucky Juveniles", *American Economic Journal: Applied Economics* 10, no. 3 (2018): 171.

¹⁷ Jens Ludwig and Sendhil Mullainathan, "Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific", *Chicago Booth Research Paper*, no. 22-15 (2022): 3.

¹⁸ Andrew I. Schoenholtz, Jaya Ramji-Nogales, Phillip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication", *Stanford Law Review* 60, no. 2. (2007): 296.

¹⁹ Cass R. Sunstein, "The Use of Algorithms in Society", preliminary draft (2013): 2.

This is also relevant in light of questions regarding the socio-economic backgrounds of judges and how their upbringing and education shapes their world view.²⁰ JAG Griffith's book *The Politics of the Judiciary* is now nearly half a century old.²¹ There has been a drive for diversity by the Judicial Appointments Commission in the UK but there remains an issue that, in the search for quality, we are still unable to draw from as wide a pool as is desirable. Is this situation, therefore, any better than the inscrutable workings of the black box algorithm?²²

There is merit to both arguments. But they overlook how the application of AI in judicial decision-making can bring with it its own biases, some of them algorithmic.²³ I outline four of them. These are biases that – if left unchecked – may be difficult to detect and far more harmful to the fairness and legitimacy of legal systems than those of human reasoning. They are a mixture of conscious human prejudice, misuse, and exploitation, plus cognitive bias and design and data flaws.

The first and most obvious source of bias is algorithmic. It stems from the selection and use of data. As the Americans would say: 'Garbage In, Garbage Out' (GIGO), which is a term commonly used to describe this problem. The quality of output is determined by the quality of input; if the models are trained on problematic data, the outputs will be problematic, too.²⁴ Biased outputs arise when the training data itself contains indiscriminate skews against groups of people, generally minority and marginalised communities.

Machine learning systems, after all, rely on historical datasets. For example, consider an AI tool being developed to adjudicate asylum cases. The underlying machine learning system will sift through pre-labelled data to identify patterns and correlations and generate predictions or decisions on the individual in question.²⁵ This is an iterative process. The algorithm will use the pre-labelled data – in this case the circumstances of previous asylum cases and their corresponding outcomes – to bolster its performance. But these pre-labelled datasets are generally made in the context of historic decision-making.²⁶ This may unfairly disadvantage certain groups if the datasets themselves contain prejudices

²⁰ See: Mathilde Cohen, "Judicial Diversity in France: The Unspoken and the Unspeakable" *Law and Social Enquiry*, 43, no. 4 (Autumn 2018) 1542-1570 and Edward M Chen "The Judiciary, Diversity and Justice for All" *California Law Review*, Vol. 91 (July 2003)

²¹ J.A.G Griffith, *The Politics of the Judiciary* (London: Fontana 1977)

²² Algorithms that are not interpretable, their inner workings are opaque or difficult for humans to understand. See Will Knight "The Dark Secret at the Heart of AI" *MIT Technology Review* (April 2017)

²³ Algorithmic bias here is defined as the systemic, repeatable errors in a computer system that create unfair outcomes.

²⁴ Mirko Bagaric, Dan Hunter, and Nigel Stobbs, "Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms are Color Blind and Never Tire", *University of Cincinnati Law Review* 88, no. 4 (2020): 1038.

²⁵ Zalnierute, Moses, and Williams, "The Rule of Law and Automation of Government Decision-Making", 426.

²⁶ Hao and Jonathan Stray, "Can you make AI fairer than a judge?"

based on bias human decisions or historical and social inequities. Further, it may well ossify existing bias in a way that a system determined by humans does not do.

What if there is, for instance, an existing judicial bias against Chadian asylum seekers? The system will identify this bias and seek to optimise its performance in light of it. The AI tool will then replicate and perpetuate the bias as it begins to adjudicate new asylum cases. Some Chadians asylum seekers whose applications ought to have been approved will instead be rebuffed. This in turn will generate perverse feedback loops as this tool and others continue to learn and improve from an expanding dataset. Human bias will be embedded – and perhaps amplified – in the code. This algorithmic bias may be much more difficult to identify and correct, particularly if an AI judicial tool has higher barriers to transparency, as I discuss below.

The concerns here are twofold. First, these algorithms contravene the foundations of Western law and criminal procedure, which emphasise individualised suspicion and investigation. Prosecutions should be grounded on the facts related to the individual themselves, not correlations observed in data on the conduct of other people with similar characteristics.²⁷

Second, it is likely that many – if not most – criminal and legal datasets are replete with human prejudices, which then generate biased outputs.²⁸ As Benjamin notes, these datasets are “produced through histories of exclusion and discrimination”.²⁹ Marginalised groups, for instance, appear more often in datasets on arrests and convictions.³⁰ This reflects in part judicial and policing biases towards these marginalised communities. Therefore, an AI judge or judicial assistant may make unjustified conclusions about an individual’s culpability based on their race, even when the consideration of such characteristics in an algorithm is prohibited.

The COMPAS system, which is used by American judges to make decisions on granting bail and sentencing, illustrates this clearly. One investigation found that it generated false negatives for white people and false positives for people of colour.³¹ In other words, the system tended to recommend granting bail to white people who later re-offended. And it tended to recommend denying bail to people of colour, many of whom did not later re-offend. The investigation concluded that white defendants who did not re-offend were half as likely to have been designated high-risk by COMPAS as defendants of colour. Stanford researchers illustrated how AI has an anti-Muslim bias problem, too.

²⁷ James E. Baker, Laurie N. Hobart, and Matthew G. Mittelsteadt, “AI for Judges: A Framework”, *Center for Security and Emerging Technology*, CSET Policy Brief (2021): 29.

²⁸ Hao, Karen, and Jonathan Stray. “Can you make AI fairer than a judge?”

²⁹ Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. New York, NY: Polity, 2019.

³⁰ Hao, Karen, and Jonathan Stray. “Can you make AI fairer than a judge?”

³¹ Julia Angwin et al., “Machine Bias”, *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

GPT-3 churned out large quantities of Islamophobic content creatively and consistently because the text on which it was trained was biased towards Muslims.³²

These problems require a seemingly simple solution: train algorithms on better, less offensive data. But it is not straightforward. Training AI is already a time and human-intensive process. Vetting large swathes of content is costly and impractical. Schmelzer calls this difficult process of collecting, cleaning, and preparing data the ‘Achilles’ Heel of AI’.³³ Programmers can intervene directly in the pre-labelling process, seeking to remove possible bias. But this is too time-consuming on a large scale, and prejudice can creep into this process, too. Instead, many researchers propose post hoc solutions to reduce these biases.³⁴ They can frontload text prompts, for example, with short positive phrases. But this is not a general-purpose solution. And it again provides an opening for bias to arise through the selection of phrases and the selective application of front-loading.

There are also questions of which data to use in the first place, especially in common law systems. Do we add additional weight to landmark judgments? Should we include the entire corpus of English tort law when developing an AI tool to assist in English tort cases? This seems ridiculous, but some important precedents – like *Mouse’s Case* (1608) – or analogous cases may date back several decades or centuries. How then do we determine the composition of our dataset? These are questions that must be carefully considered.

They also highlight another problem of extending AI’s role in the courtroom: a disembodied court deprives us of the opportunity to construct precedent-setting legal arguments and judgments. The ability to revise the law in this way by shaping its gradual evolution is a key component of common law systems.³⁵ Judges must strike a careful balance between stability and change – between respect for precedent and adapting the law to unforeseen circumstances. This role of a judge is an essential one that an AI courtroom renders obsolete, however. AI adjudication promotes legal stasis and impedes the natural fluidity of the law.

There is then also algorithmic bias through coding. There are not too many lawyers or legal scholars who know their bytes from their bits. Nor is there an abundance of coders well-versed in the intricacies of the law. Invoking C.P. Snow’s *The Two Cultures* – in which he highlights the divide

³² Abubakar Abid, Maheen Farooqi, and James Zou, “Persistent Anti-Muslim Bias in Large Language Models”, *AIES ’21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (2021): 299.

³³ Ron Schmelzer, “The Achilles’ Heel of AI”, *Forbes*, March 17, 2019, <https://www.forbes.com/sites/cognitiveworld/2019/03/07/the-achilles-heel-of-ai/>.

³⁴ Abid, Farooqi, and Zou, “Persistent Anti-Muslim Bias in Large Language Models”, 304.

³⁵ Andrew C. Michaels, “Artificial Intelligence, Legal Change, and Separation of Powers”, *University of Cincinnati Law Review* 88, no. 4 (2020): 1083.

and mutual distrust and animosity between the sciences and humanities – is apt, if not a little crude.³⁶ Law and AI, of course, are not bitter adversaries. But these two fields do remain deeply divided and isolated from one another. There is, with exceptions, an absence of mutual intelligibility. This can hinder serious examination of the difficulties of coding law or the deeper jurisprudential questions that must be confronted in that process. Without a stronger link connecting these two disciplines, accurately and comprehensively translating law into code will be an insurmountable task.

Even then, there is the question of whether laws can be reliably converted into software in the first place. All laws, including the most basic, contain subtleties that require assumptions when programmers encode them.³⁷ This leaves a wide opening for diverging interpretations (or indeed misinterpretations) of laws, which can in turn produce disparate enforcement outcomes. In one Australian study, for example, researchers found that programmers automating the enforcement of speed limits wrote code that issued very different numbers of tickets for the same data.³⁸

Can there be a single, correct interpretation, as Dworkin suggests?³⁹ Perhaps where the law is simple and clear. But what about more complex laws? Would a single interpretation even be desirable? Surely a degree of flexibility here is needed, particularly in common law countries, where interpretations evolve over time. Judges of course can also interpret the same rule in markedly different ways. But this is carried out in public, and their interpretations are subject to review through an appeals process. It is also an important component of the judicial system. As Hart notes, the doctrine of strong discretion in judicial interpretation is “a necessary by-product of the inherent indeterminacy of social guidance”.⁴⁰

If a programmer does not interpret the law in an appropriate way, or does not accurately translate its ambiguities into code, this may be problematic to identify and fix. Given AI’s scalability, the program can adversely affect citizens at a scale and speed that is difficult to correct.

³⁶ C.P. Snow, *The Two Cultures and the Scientific Revolution: The Rede Lecture, 1959* (New York, NY: Cambridge University Press, 1961).

³⁷ The Hon. Justice Melissa Perry, “iDecide: the Legal Implications of Automated Decision-making” (lecture, University of Cambridge, September 15, 2014), <https://www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20140915>.

³⁸ Lisa A. Shay, Woodrow Hartzog, John Nelson, and Gregory Conti, “Do robots dream of electric laws? An experiment in the law as algorithm”, in *Robot Law*, ed. by Ryan Calo, A. Michael Froomkin, and Ian Kerr (Cheltenham: Edward Elgar Publishing, 2016), 276.

³⁹ Ronald Dworkin, “No Right Answer”, *New York University Law Review* 53, no. 1 (1978): 1.

⁴⁰ H.L. A. Hart, *The Concept of Law* (Oxford: Clarendon Press, 1994), 123.

Such was the case in Australia's Robodebt scandal. The Online Compliance Intervention scheme was a method of automated debt assessment and recovery that overstepped its bounds.⁴¹⁴² It wrongly targeted and harmed many Australians, which took years to uncover fully. And it was recently the subject of a Royal Commission. The legal assumptions on which the code underlying AI's present use in the courts is based are therefore not nearly as examinable, intelligible, or reviewable as those of human judges. If the role of AI in judicial decision-making is to be expanded, then these are challenges that need to be first addressed.

Coding deeper legal questions and principles – as well as broader social and ethical norms – is an even more challenging task. Those thorny, indeterminate issues I raise above on the nature of judgement and what constitutes good exercise of it, for example, would need to be settled.⁴³ Value judgements would need to be formalised, too. A comprehensive, automated legal system would need all this to give structure and foundation to the code lest the algorithm's outputs deviate from accepted norms. Outside of a program these questions and principles can exist in greater abstraction. But to distil them into code and achieve consensus seems an overwhelming social and jurisprudential task that requires consideration of political and economic concerns, too.

Already lawyers and programmers are grappling with this issue in relation to conflicting notions of fairness. What is fairness? Should it be of the procedural kind, which can overlook sociohistorical and policy inequities? Or should we prioritise distributive or representational fairness?⁴⁴ The quandary that bedevils coders and lawyers alike is that these competing notions come at the expense of one another. There are inherent trade-offs with which decision-makers must engage, and maximising one kind of fairness can mean sacrificing another. In determining the COMPAS algorithm's high-risk threshold score for reoffending, for example, programmers faced a difficult choice. They could either favour procedural fairness, treating individuals with the same risk score in the same way. Or they could prioritise distributive fairness, keeping the error rates comparable between groups.⁴⁵ They ultimately leant towards the former, but one can make an equally strong argument for the latter.

⁴¹ Caroline Gans-Combe, "Automated Justice: Issues, Benefits and Risks in the Use of Artificial Intelligence and Its Algorithms in Access to Justice and Law Enforcement", in *Ethics, Integrity and Policymaking: The Value of the Case Study*, ed. by Dónal O'Mathúna, Ron Iphofen (Berlin: Springer, 2022), 176.

⁴² The Online Compliance Intervention scheme, implemented by the Australian government, used an automated system to match income data from the Australian Tax Office with income reported to Centrelink (social welfare service) to detect overpayments. However, this system was flawed: i). the system annualised income data, assuming consistent earnings throughout the year. This method failed to account for intermittent or irregular income, common among welfare recipients, leading to inaccurate debt calculations. ii). The system placed the onus on individuals to prove they didn't owe money, this approach was burdensome for vulnerable people. iii). debt notices without a clear explanation or understanding of how the debts were calculated caused distress and confusion.

⁴³ Baker, Hobart, and Mittelsteadt, "AI for Judges", 44.

⁴⁴ Hao and Stray, "Can you make AI fairer than a judge?".

⁴⁵ Ibid.

Judges do make the same decisions daily. But again, this process is carried out in public, is more easily reviewable and contestable, and builds in more flexibility. Encoding these principles and questions instead deprives them of much of their nuance.

I do not wish to suggest that accurately and adequately translating the law and its foundations into code is as an intractable task. Judicious, innovative programming can minimise some of these trade-offs; Yang and Dobbie, for instance, show how two statistical solutions can reduce algorithmic biases such that some degree of both procedural and distributive fairness is retained.⁴⁶ Other design difficulties – like overfitting, the mismatch of training data to an algorithm’s use, or statistical bias – can also be fixed with better statistical modelling and data.⁴⁷ But there are still serious challenges turning law into code that must be overcome.

The third bias arises through the intentional misuse of an AI judge or judicial tool. In countries with weaker judicial systems, we should be concerned that the application of AI in the courtroom could be co-opted for political expediency or some other gain. Already we see this in the policing and surveillance practices of certain countries; China, for instance, has deployed advanced facial recognition technology to track Uyghur Muslims.⁴⁸ By creating, implementing, and controlling access to the code, states could manipulate the algorithms and their outputs to influence sensitive political cases. In other words, they could determine the outcome of legal proceedings. With the help of AI judges and judicial tools, they could, for example, undermine a corruption case involving party officials. Or they could target and impose harsher sentences on minority groups and political dissidents.

In many authoritarian countries, where the state has already co-opted the judicial system, such practices are commonplace.⁴⁹ But algorithms in this case could make these processes more efficient. And they could operate under a guise of impartiality that seemingly distances the AI and its decisions from political intervention and the state. In countries with more robust judicial systems, meanwhile, we should still be wary that the algorithms could be influenced discreetly to favour political incumbents. Where trust in AI is high or public understanding and scrutiny of AI’s use in courts low, these concerns are amplified.

⁴⁶ Crystal S. Yang and Will Dobbie, “Equal Protection Under Algorithms: A New Statistical and Legal Framework”, *Michigan Law Review* 119, no. 2 (2020): 291.

⁴⁷ Baker, Hobart, and Mittelsteadt, “AI for Judges”, 44.

⁴⁸ Paul Mozur, “One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority”, *New York Times*, April 14, 2019, <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>.

⁴⁹ Tamir Moustafa and Tom Ginsburg, “Introduction: The Function of Courts in Authoritarian Politics”, in *Rule By Law: The Politics of Courts in Authoritarian Countries*, ed. by Tamir Moustafa and Tom Ginsburg (Cambridge: Cambridge University Press, 2008), 7.

There is scope for misuse even within the application of AI to more technical components of the judicial system. North Macedonia embraced the digital transformation of its court case management system in 2010 with the introduction of an automated tool.⁵⁰ Although not AI, the Automated Court Case Management Information System (ACCMIS) was designed to improve the efficiency of this process by replacing the manual distribution of cases. ACCMIS's appeal was also grounded in its randomised allocation of cases to judges, which would prevent judicial interventions in favour of the government. The system itself, however, was later subject to manipulation.⁵¹ One court president was found guilty of – and many others were said to have been – selectively and manually distributing cases of political significance. Judges with close connections to the governing party were assigned cases involving high officials who were to be tried for serious crimes.

Bias through misuse is not limited to political expediencies. Expanding the scope of AI in judicial decision-making can create an opportunity for corporate entities and individuals to acquire strategic legal advantages, too. Firms can treat political and social institutions, including the law, as firm (or 'institutional') resources.⁵² These resources can be exploited by savvy legal entrepreneurs, working – in this instance – in tandem with programmers. Those with good knowledge of the algorithms or even the resources to replicate them could tailor their arguments or frame the case in such a way that raises their odds for a favourable outcome.

Should the introduction of AI into the courtroom not eliminate any opportunities to acquire competitive legal advantages? Uncertainty, after all, spawns entrepreneurship.⁵³ And it is legal flexibilities, which judicial algorithms should reduce, that generate legal uncertainties. In other words, legal parties should be able to compete on a level playing field. This overlooks, however, how uncertainties over the nature of the algorithms themselves – if they remain opaque – would emerge instead.⁵⁴ These uncertainties would encourage informational and resource-based asymmetries to arise that favour larger, more technologically inclined firms.

This relates to the final avenue through which bias can materialise: from the interaction of human judges and the judicial system with AI itself. This interaction could distort judicial decision-making or in several ways. First, concerns over jurisdictional appeal may incentivise the deployment of more

⁵⁰ Michal Škop et al., "AlGovrithms 2.0: The State of Play", *Open Data Kosovo* (2021): 22.

⁵¹ Ibid, 23.

⁵² William D. Oberman, "Strategy and Tactic Choice in an Institutional Resource Context", in *Corporate Political Agency*, ed. by Barry M. Mitnick (New York, NY: SAGE Publications, 1993), 215.

⁵³ Justin W. Evans and Anthony L. Gabel, "Legal Competitive Advantage and Legal Entrepreneurship: A Preliminary International Framework", *North Carolina Journal of International Law and Commercial Regulation* 39, no. 2 (2014): 350.

⁵⁴ Zalnieriute, Moses, and Williams, "The Rule of Law and Automation of Government Decision-Making", 438.

firm-friendly algorithms in the courtroom. Such concerns are already present.⁵⁵ But it is much simpler (and less conspicuous) to alter code in such a way that favours businesses that governments may seek to attract. The right, subtle changes, which would undermine the integrity of the judicial system, could make a jurisdiction a more appealing market; it could expand opportunities for firms to extract legal advantage and competitive value.

Second, AI analysis of court data, which is becoming increasingly common,⁵⁶ could lead to more forum shopping within a jurisdiction. AI tools – like Lex Machina and Solomonic – that assess the risks and probabilities associated with particular judges or courts already exist.⁵⁷ These ‘predictive justice’ tools can in turn provide desirability scores for different venues, helping determine in which court plaintiffs should lodge their lawsuits.⁵⁸ The US – where forum shopping is more common – in particular provides fertile ground for AI-driven predictive justice, although there are statutory constraints that limit where a lawsuit may be brought.⁵⁹ Wider, more accurate use of such tools could foster an environment in which unwelcome pressure is placed on courts and judges in a way that is not in the interests of justice.

There are other features of AI-human interaction in the courtroom that need to be examined for bias. Is there a tendency, for example, for judges to unduly accept or override the recommendations of existing AI judicial tools? We want AI neither to be wasted on intransigent judges nor encourage unwavering deference and judicial conformism. This is a possible bias we must therefore evaluate before expanding the judicial role of AI further.

Transparency, Interpretability, and Public Opinion

These biases can all be addressed to varying degrees. But this will not matter if the transparency and interpretability of the algorithms are low, and if the public is sceptical of this technology. In fact, many of these biases are amplified by poor public perception and understanding of AI judicial decision-makers. A lack of transparency and interpretability, for example, weakens the ability of

⁵⁵ Panikos Georgallis, João Albino-Pimentel, and Nina Kondratenko, “Jurisdiction shopping and foreign location choice: The role of market and nonmarket experience in the European solar energy industry”, *Journal of International Business Studies* 52 (2020): 853.

⁵⁶ Rob Chesnut, “‘Judge Shopping’ Will Only Get Easier Without Rule Changes”, *Bloomberg*, August 21, 2019, <https://news.bloomberglaw.com/us-law-week/judge-shopping-will-only-get-easier-without-rule-changes>.

⁵⁷ Amanda Chaboryk, Adam Sanitt, David Wilkins, “New Digital Case Law Service to Further Transform UK Litigation”, *Oxford Business Law Blog*, July 27, 2022, <https://blogs.law.ox.ac.uk/business-law-blog/blog/2022/07/new-digital-case-law-service-further-transform-uk-litigation>.

⁵⁸ John Facciola and Nicholas Mignanelli, “Artificial Intelligence and the Courts: Materials for Judges”, *American Association for the Advancement of Science* (2022): 12.

⁵⁹ Alex Botoman, “Divisional Judge-Shopping”, *Columbia Human Rights Law Review* 49, no. 2 (2018): 299.

judges and litigators to assess how an algorithm reached its verdict; what evidence it considered and how it weighted it; and to what extent this evidence reflects the circumstances of the case.⁶⁰ This in turn raises due process concerns that could undermine the fairness of the judgement and the perceived legitimacy of the judicial system.

Burrell's analysis of algorithmic opacity helps us illustrate these concerns.⁶¹ She identifies three variants of it: intentional, arising from corporate or state secrecy; technical, due to the complex nature of AI; and intrinsic, stemming from the fundamental differences in human and algorithmic cognition. Intentional opacity often emerges from proprietary or legal privacy concerns. Governments may outsource the construction or operation of these algorithms to private companies. These "private, profit-maximising entities, operating under minimal transparency obligations",⁶² understandably seek to restrict the public dissemination of their trade secrets. They will insist on contractual terms that prevent further disclosure of the algorithms by the government.

Such is the case with the COMPAS system, which was built by Northpointe. Information on its methods and datasets used in training continues to be withheld, despite public concerns over its algorithmic bias and violations of due process.⁶³ These concerns were raised directly in *Loomis vs. State*.⁶⁴ COMPAS had classified a man as at high risk of reoffending, and he was sentenced to six years. The man appealed the ruling on the basis that the judge had considered the output of the algorithm whose underlying methodologies were unexaminable. The Wisconsin Supreme Court ultimately found that the use of closed-source software did not violate the defendant's due process rights.⁶⁵ But it still raised concerns over the broader transparency issues associated with proprietary algorithms. One justice noted in her concurring opinion how "making a record, including a record explaining consideration of the evidence-based tools and the limitations and strengths thereof, is part of the long-standing, basic requirement that a circuit court explain its exercise of discretion at sentencing".⁶⁶

Existing intellectual property rights clearly serve as a major barrier to algorithmic transparency. They limit the auditability of AI, which prevents the external assessment of this technology needed to help

⁶⁰ Danielle Leah Kehl and Samuel Ari Kessler, "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing", Responsive Communities Initiative, Berkman Klein Center for Internet & Society, *Harvard Law School* (2017): 32.

⁶¹ Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, 3 (2016).

⁶² Maayan Perel & Niva Elkin-Koren, "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement", *Florida Law Review* 69, no. 1 (2017): 181.

⁶³ Zalnieriute, Moses, and Williams, "The Rule of Law and Automation of Government Decision-Making", 449.

⁶⁴ *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

⁶⁵ *Ibid.*

⁶⁶ *Ibid.*

guard against the biases described above.⁶⁷ There is an accountability deficit. Instead, confidentiality provisions ensure the private sector vendors that develop this software acquire near-monopolistic control on information about the technological, ethical, social, and political implications of AI justice, with no responsibility to facilitate public discussion. Concerns overdue process – despite the Wisconsin Supreme Court’s findings – persist. Questions about what kinds of public information and access democratic governance requires remain, too.⁶⁸

The argument for open-source software in AI judicial decision-making is thus a compelling one. We ought to know what algorithms judiciaries use, how they work, and what their effects on those whose lives it influences are. And yet there are countervailing arguments that complicate this matter. Mandating transparency may make it more difficult for governments to attract the most capable private sector vendors.

There is also opacity arising from technical illiteracy. For those working with AI in the courtroom, the algorithms and their outputs can be confusing and interpreted incorrectly.⁶⁹ A limited understanding of the inner workings of the algorithms and their outputs can undermine a judge’s decision-making when this technology is consulted. Judges may use the algorithm’s output in a way that it is not meant to be used or infer something they are not meant to infer. They may not, for instance, recognise that these systems capture correlations and not necessarily causal relationships.⁷⁰ This illiteracy applies to the public, too. Most individuals – including those affected by adverse decisions – will not possess the specialist skills required to understand the technology or challenge a decision it makes or influences.⁷¹ And so even if the underlying algorithms are open-source, these technical barriers to transparency will persist. Those without these specialist skills can of course take counsel from those with it, just like people now who have limited knowledge of the law. But such counsel is unlikely to be as readily accessible – or even available if confidentiality provisions continue to obscure the underlying algorithms.

Third, Burrell highlights intrinsic opacity, which arises from a fundamental mismatch between how humans and algorithms understand the world. This is where interpretability – or ‘explainability’ – comes into play. AI’s machinations have been described as a ‘black box’ because they can be

⁶⁷ Guido Noto La Diega, “Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection and Freedom of Information”, *JIPITEC* 9, no. 3 (2018): 9.

⁶⁸ Hannah Bloch-Webba, “Transparency’s AI Problem”, *Knight First Amendment Institute at Columbia University*, <https://knightcolumbia.org/content/transparencys-ai-problem>.

⁶⁹ United Nations Institute for Disarmament Research (UNIDR). “Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence.” *UNIDR*, 2018, p. 4.

⁷⁰ Zalnieriute, Moses, and Williams, “The Rule of Law and Automation of Government Decision-Making”, 450.

⁷¹ *Ibid*.

challenging to explain or articulate.⁷² Because humans reason differently to AI, even experts cannot reliably interpret the interactions among algorithms and data. The inner-workings of these algorithms – and crucially, how they arrived at specific results – can be mysterious; deep neural networks, which rely on an ever-multiplying set of hidden neural layers, are particularly afflicted with this incomprehensibility.⁷³ This makes it difficult to “[obtain] human-intelligible and human-actionable information about the operation of autonomous systems”.⁷⁴ And this in turn makes auditing the algorithm more challenging and the cost of identifying and then remedying biases much higher.

Many programmers are directing their efforts to developing ‘explainable AI’ (or XAI).⁷⁵ These algorithms seek to translate machine learning inferences into language accessible to people. They help characterise model fairness, accuracy, and outcomes, which should build trust and improve confidence in these algorithms.^{76 77} But there is also a trade-off here between accuracy and explainability. Those who use these algorithms prefer “transparent, interpretable models not only for predictive decision-making but also for after-the-fact auditing and forensic purposes”.⁷⁸ And yet programmers tend to create more accurate algorithms with increasingly complex, black-box models, including deep learning models.

⁷² Bloch-Webba, “Transparency’s AI Problem”.

⁷³ M. Beatrice Fazi, “Beyond Human: Deep Learning, Explainability and Representation”, *Theory, Culture & Society* 38, no. 7 (2021): 56.

⁷⁴ Berkman Klein Center Working Group on AI Interpretability. “AI: Transparency and Explainability.” *Berkman Klein Center for Internet & Society at Harvard University*. <https://cyber.harvard.edu/projects/ai-transparency-and-explainability>.

⁷⁵ Zalnieriute, Moses, and Williams, “The Rule of Law and Automation of Government Decision-Making”, 450.

⁷⁶ LIME (Local Interpretable Model-agnostic Explanations) represents advancement in explainable AI, providing an insight into the decision-making of complex models through the creation of simpler, interpretable models that approximate the original's predictions. However, LIME is limited by its reliance on a method called ‘perturbation strategy’. This involves making small changes to the input data and observing how these changes affect the AI's predictions. The problem is that this approach can sometimes lead to inconsistent results. Imagine if changing a small detail sometimes significantly alters the AI's decision and other times it doesn't; this inconsistency can make it hard to trust the explanations provided by LIME, especially in legal situations where consistency is crucial. See Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier.” *arXiv.org*, 2016. <https://arxiv.org/abs/1602.04938>.

⁷⁷ Similarly, SHAP (SHapley Additive exPlanations) gives a detailed breakdown of how much each feature (like evidence or legal arguments) influences the AI's decision. While SHAP's method is more consistent, it can be complex and slow, especially with large and complex cases. This makes it difficult to use SHAP in situations where quick decisions are needed. Also, for people who aren't experts in AI, SHAP's detailed explanations can sometimes lack the intuitive clarity needed for non-technical stakeholders in the judicial system. See Scott M. Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems*, 1970.

https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

⁷⁸ Innocent Kamwa, Subhransu Samantary, and Geza Joos, “On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors”, *IEEE Transactions on Smart Grid* 3, no. 1 (2012): 152.

It is worth noting that human judges, too, can be black boxes; their decisions are not always transparent. It may not be clear, for example, how a judge chooses between competing conceptualisations of fairness. But their decisions are more easily contestable and reviewable, and they are carried out in public. In other words, greater public accountability minimises the adverse effects of opaque judgements. It is unsurprising, therefore, that the House of Lords argued “it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual’s life, unless it can generate a full and satisfactory explanation for the decisions it will take”.⁷⁹

Algorithmic opacity thus produces obscure justice, which is obviously objectionable. It facilitates – rather than mitigates – the biases that undermine the integrity and fairness of the judicial system. And it makes detecting these biases much more difficult. As the Senior President of Tribunals noted, “when justice slips out of sight...the prospect of arbitrary, incompetent or unlawful conduct raises its head”.⁸⁰

This relates to the popular perception of AI judges and judicial tools, as opaque algorithms will erode public trust in them. Trust in the judicial system undergirds the system’s very foundations; people obey the law in part because they perceive it as legitimate and procedurally just.⁸¹ Expanding AI’s role in the courtroom could weaken this trust and people’s legal compliance if they believe AI adjudications to be less than fair than human ones. They may even reject robot judging altogether. One study illustrates that people do share this intuition: they view human judges as fairer than their AI counterparts.⁸² In other words, there is a perceived ‘human-AI fairness gap’.

The degree of aversion to AI in the courtroom seems to depend on how the technology is used – that is, on the nature and extent of judicial reliance on AI at different stages of adjudication.⁸³ Using algorithms during information acquisition is generally perceived as fairer; trust in AI-led information analysis, decision selection, and decision implementation is lower.

This is not surprising. There are concerns over the accuracy, thoroughness, and reliability of AI verdicts, much of which stems from the absence of interpretable decisions; these concerns bear less on

⁷⁹ House of Lords Artificial Intelligence Select Committee, “AI in the UK: Ready, Willing and Able”, *Report Session 2017-2019* (2019): 128.

⁸⁰ Sir Ernest Ryder, “Securing Open Justice” (lecture, Max Plank Institute Luxembourg for Procedural Law & Saarland University, February 1, 2018), 5.

⁸¹ Tom R. Tyler, *Why people obey the law* (Princeton, NJ: Princeton University Press, 2006), 19-29.

⁸² Benjamin Minhao Chen, Alexander Stremitzer, and Kevin Tobia, “Having Your Day in Robot Court”, *Harvard Journal of Law and Technology* 36, no. 1 (2022): 127.

⁸³ Dovelé Barysè and Roeë Sarel, “Algorithms in the court: does it matter which part of the judicial decision-making is automated?”, *Artificial Intelligence and Law* (forthcoming) (2023): 2.

the information acquisition stage, which is seen to require less of the ‘softer’ human skillset.⁸⁴ In an extreme case, in which the court’s “multiagency structure...collapses into one procession operation”,⁸⁵ the court is disembodied entirely, limiting an individual’s participation in the judicial process. Relevant information is inputted, and the AI judge reaches a verdict, depriving people of the ability – or even the right – to be heard and understood. This then undermines confidence in the judicial process itself, shaking the foundations of perceived procedural justness.

If a court is fully automated, poor public perception of the technology would also certainly increase the number of appeals to a human judge. This would reduce many of the efficiency gains that make AI adjudications so appealing. And so, as Campbell notes, we must ask whether “AI courts can enable public participation, give participants a sense of being fairly heard...[and] vindicate the legitimacy not just of the courts, but of the governmental systems within which they reside”.⁸⁶

Evidence suggests that providing individuals with an opportunity to speak and be heard – a form of ‘algorithmic offsetting’ – offsets much of this human-AI fairness gap.⁸⁷ Introducing interpretable decisions shrinks the gap, too, bolstering the perceived accuracy and conscientiousness of the decision-making process⁸⁸. Embedding these offsetting approaches within a ‘technological due process’ would further enhance public confidence in algorithmic decision-makers. This process would give individuals the “right to inspect, correct, and dispute inaccurate data and to know the sources (furnishers) of the data”.⁸⁹ It would also ensure algorithms remain publicly accessible. Transparency, interpretability, and public perception are thus closely entwined; transparency and interpretability can help build trust in AI adjudications, just as they do across judicial systems now.

And yet some public scepticism of AI’s extension in the courtroom is useful in ensuring algorithms remain accountable and fair. The performance of models can be overhyped and exaggerated, amplifying a common misconception that these algorithms always surpass human-level reasoning.⁹⁰ This is dangerous while AI judicial decision-makers remain uninterpretable, untransparent, and susceptible to the algorithmic and other biases described above. Unwitting belief in their capacity to deliver the correct verdict could thwart serious examination of their underlying shortcomings. There is

⁸⁴ Chen, Stremitzer, and Tobia. “Having Your Day in Robot Court.” 127.

⁸⁵ James Kwan, James Ng, and Brigitte Kiu, “The use of artificial intelligence in international arbitration: Where are we right now?”, *International Arbitration Law Review* 22, no. 1 (2019): 20.

⁸⁶ Ray Worthy Campbell, “Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning”, *Colorado Technology Law Journal* 18, no 2. (2020): 341.

⁸⁷ Chen, Stremitzer, and Tobia. “Having Your Day in Robot Court.” 128.

⁸⁸ Models could, for example, be evaluated by their data source, selection criteria, pre-processing regimen, weighting, feature importance, and accuracy metrics (such as precision, recall, F1 score).

⁸⁹ Danielle Keats Citron and Frank Pasquale, “The Scored Society: Due Process for Automated Predictions”, *Washington Law Review* 89, no. 1 (2014): 1.

⁹⁰ Celeste Kidd and Abeba Birhane, “How AI can distort human beliefs”, *Science* 380, no. 6651 (2023): 1222.

a fine balance to strike – once AI justice is inevitably and gradually improved – in boosting popular confidence in these tools while tempering unrealistic expectations of their capabilities.

I want to turn to an extrinsic threat to the integrity of judicial decision-making, and that is the question of “deepfakes”.⁹¹

The potential impact of ‘deepfakes’ to sow doubt and confusion is a topic familiar to most of the legal profession and wider public. From the reasonably sophisticated fake image of the Pope with his papal puffer jacket, which went viral earlier this year,⁹² to more primitive ‘shallow fakes’ such as a 2019 video of an apparently intoxicated Nancy Pelosi, audio visual fakes are becoming increasingly prevalent and ever more convincing.⁹³ One only needs to consider the power of audio deepfakes or look at software such as Which Face is Real, which creates lifelike portraits of people who never existed, to see how AI could be used to fabricate evidence or undermine a case.⁹⁴

The first deepfakes were created on single neural networks but now are far more likely to have been assembled through Generative Adversarial Networks, which are both more sophisticated and readily available online.⁹⁵ Such rapid technological advancement has led to a great deal of discussion amongst scholars, both to establish the scale of the challenge posed and to consider how judges should mitigate against them. Scholars such as Delfino argue that the existing mechanisms for the governing the admissibility of evidence are inadequate and as such a greater role should be given to the judge to act as a ‘gatekeeper’ on information.⁹⁶ Venema and Geradts by contrast, propose enhancing jurors’ synthetic media literacy to ensure they make the most informed judgements.⁹⁷ Both perspectives and others, are of merit, and I include the disagreement that exists in the literature to highlight the complexity of the issue at hand. My prognosis for a way forward will appear in the next paper.

⁹¹ The term ‘Deepfake’ is a portmanteau of ‘deep learning’ and ‘fake’ first coined by Reddit users in 2017, see - Douglas Harris, “Deepfakes: False Pornography is Here and the Law Cannot Protect You”, *Duke Law and Technology Review*, (2019) 99-100

⁹² Joel Goble. “I Thought I Was Immune to Being Fooled. Then I Saw the Pope in a Coat.” *The Guardian*, March 27, 2023.

⁹³ Simon Illery. “Fake Photos of Pope in a Puffer Jacket Go Viral, Highlighting the Power and Peril of AI.” *CBS*, March 28, 2023; Blair Guild and Elyse Samuels. “Pelosi Videos Manipulated to Make Her Appear Drunk Are Being Shared on Social Media.” *Washington Post*, May 23, 2019.

<https://www.youtube.com/watch?v=sDOo5nDJwgA>. The doctored video of then Speaker Pelosi is a ‘shallow fake’ as it uses existing footage.

⁹⁴ Jevin West and Carl Bergstorm. 2019. https://www.whichfaceisreal.com/results.php?r=1&p=1&i1=image-2019-02-17_123752.jpeg&i2=35757.jpeg.

⁹⁵ Riana Pfefferkorn. “Deepfakes’ in the Courtroom.” *B.U. Pub. Int. L.J.* 29 (2020): 245.

⁹⁶ Rebecca A Delfino. “Deepfakes on Trial: A Call to Expand the Trial Judges’ Gatekeeping Role to Protect Legal Proceedings from Technological Fakery.” *Hastings Law Journal* 74, no. 2 (February 2023): 293-348.

⁹⁷ Agnes E. Venema and Zeno J. Geradts. “Digital Forensics, Deepfakes and the Legal Process.” *SciTech Lawyer* 16, no. 4 (2020).

In my view, the two central challenges deepfake technology poses for judges are firstly the authentication of evidence and secondly the maintenance of trust in the trial process. For centuries, the authentication of evidence was ‘seeing is believing’ for judges and jurors alike. New forms of evidence such as film and DNA were commonly used in trials by the latter half of the last century. Their use posed new challenges in a courtroom scenario, illustrated by the controversy over the use of film as evidence during the Nuremberg Trials.⁹⁸ However, obtaining consensus on the veracity of film and of DNA was ultimately established. Even though their manipulation as evidence sources was of course possible, close technical analysis or expert advice ensured that tampering is recognisable.

In theory, the same could also be true for deepfakes. Although judges will often no longer be able to tell the veracity of evidence with the naked eye, synthetic media specialists should be able to verify content.⁹⁹ The issue with this optimistic perspective, is as Ajder puts it, “as the technology for generating synthetic media and deepfakes increases and becomes more accessible, the number of human experts who could rule with authority on whether a piece of media is real or not, has not.”¹⁰⁰

To fill this gap of technical expertise, could we not train an AI system to identify manipulated content? As appealing as this option seems, the same potential issues of transparency and interpretability would persist and in addition, as Dana Rao Adobe’s legal, security and policy lead highlights, detection developers would be locked in a perennial race with the developing deepfakes.¹⁰¹

If judges are unable to trust the evidence placed in front of them, there is a danger they become overly cynical. Indeed, as Chesney and Citron argue, deepfakes will allow liars to cast doubt on real footage or audio when the resources are not there to expose their fraud.¹⁰² Such actions, coupled with wider public scepticism of digital media, could lead to what they describe as “the liars’ dividend,” i.e., as the public (judges included) are more aware of the dangers of deepfakes, the more sceptical they become of all digital content, including unaltered content.¹⁰³ Bennett goes further arguing that as deepfake technology improves, it has the potential to make even witnesses vulnerable to deepfakes due to the

⁹⁸ Telford Taylor. *The Anatomy of the Nuremberg Trials: A Personal Memoir*. New York, 1992, 146-149.

⁹⁹ Agnieszka McPeak. "The Threat of Deepfakes in Litigation Raising the Authentication Bar to Combat Falsehood." *Vanderbilt Journal of Entertainment and Technology Law* 23 (2021): 433.

¹⁰⁰ Henry Ajder. "Courts and Lawyers struggle with growing prevalence of deepfakes." In *American Bar Association Journal*, by Matt Reynolds, June 2020

¹⁰¹ Dana Rao. "Creating a 'lie detector' for deepfakes." Interview by David Pogue. CBS, January 29, 2023. Audio, 1.28. <https://www.youtube.com/watch?v=Bf7AS1HBeaM>.

¹⁰² Bobby Chesney and Danielle Citron. "Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107 (2019): 1776-1785.

¹⁰³ Chesney and Citron. "Deepfakes."

inherent suggestibility flaw present in people's memories, with complex deepfakes prompting witnesses to recall falsehoods.¹⁰⁴

Such problems, which would have been pure speculation less than a decade ago, are already posing practical challenges to judges across jurisdictions. In a British custody case in 2019, a woman alleged that her husband was violent and presented a recording of him making threats to her in order to substantiate her claim.¹⁰⁵ After analysing the recording's metadata, it was revealed that the recording was in fact a 'cheapfake' and the woman had manipulated a recording to make it sound like her husband, using readily-available software and having followed online tutorials for instruction.¹⁰⁶

In the United States vs Doolin the defendant, Joshua Christopher Doolin, on trial for his part in the January 6 riot at the US Capitol, sought to profit from the 'liars' dividend.¹⁰⁷ Arguing that open-source video footage of the riots was inadmissible, as due to "recent technological advances, relying on open-source media with no evidence of a chain of custody" did not meet the threshold of evidence set out by the prosecution.¹⁰⁸ Doolin cited the aforementioned video of Nancy Pelosi amongst others, as grounds for the court to deny the Government their proposed motion to authenticate the videos "until they can support the circumstantial evidence they claim to possess."¹⁰⁹ The government responded by arguing circumstantial evidence presented a prima facie basis to believe the open-source footage was genuine.¹¹⁰ There was no disputing that deepfake technology was real, but rather Doolin's argument went "to the weight of the video evidence, not its admissibility."¹¹¹

Like the earlier British case, Doolin's attempt to benefit from the uncertainty created by deepfakes was unsuccessful, being sentenced to 18 months in prison with a further 36 months of supervised release.¹¹² However, both cases illustrate challenges that judges will face from deepfake technology in the coming years whilst hinting at how they might be overcome.

¹⁰⁴ Mark W. Bennett. "Unspringing the Witness Memory and Demeanour Trap: What Every Judge and Juror Needs to Know about Cognitive Psychology and Witness Credibility." *American University Law Review* (2015).

¹⁰⁵ Reynolds. "Courts and Lawyers struggle with growing prevalence of deepfakes."

¹⁰⁶ Ibid. 'Cheapfake' refers to a less sophisticated doctoring of audio-visual content.

¹⁰⁷ "Scheduling Order at 1, United States v. Doolin, No. 21-cr-00447 (D.D.C. September 2, 2022), ECF No. 151."

¹⁰⁸ "Def.'s Resp. to United States Mot. in Lim. Regarding Authentication of Certain Video Evid., United States v. Doolin, No. 21-cr-00447 (D.D.C. August 5, 2022), ECF No. 135." – threshold of evidence being 'Federal Rule of Evidence 901a' – "Federal Rules of Evidence" (Washington: U.S. Government Publishing Office 2021)

¹⁰⁹ Ibid. at 2-3.

¹¹⁰ Matthew Ferraro and Brent Gurney. "The Other Side Says Your Evidence is a Deepfake. What Now?" *Law360* 2 (Wilmer Hale December 21, 2023).

¹¹¹ Ferraro and Gurney, "The Other Side Says Your Evidence is a Deepfake" and Govt's Reply to Def's Resp. to United States Mot. in Lim. Regarding Authentication of Certain Video Evid. at 1, United States v. Doolin, No. 21-cr-00447 (D.D.C. Aug. 12, 2022), ECF No. 140

¹¹² United States Department of Justice. 2023. "Florida Man Sentenced on Felony and Misdemeanour Charges for Actions During Jan 6. Capitol Breach." Departmental Press Release, August 16, 2023.

Conclusion: In this first paper, I have attempted to outline the importance of human judgement as part of judicial decision-making, the role of the modern judge, and the principal intrinsic and extrinsic risks that affect AI processes. In my next paper, I aim to set out a way forward for the creation of a justice framework within which AI may operate that can maintain public trust and accountability.

Acknowledgments

I would like to thank both the Senior Fellows at the Harvard Kennedy School for their insight and support and my Harvard research assistants Max Murphy, Marcus Ray and Michael Bryan, along with my Senior Parliamentary Researcher, Will Siebert.

Bibliography

Abid, Abubakar, Maheen Farooqi, and James Zou. "Persistent Anti-Muslim Bias in Large Language Models". *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (2021)*: 298-306.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias". *ProPublica*, May 23, 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Bagaric, Mirko, Dan Hunter, and Nigel Stobbs. "Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms are Color Blind and Never Tire". *University of Cincinnati Law Review* 88, no. 4 (2020), 1037-1081.

Baker, James E., Laurie N. Hobart, and Matthew G. Mittelsteadt. "AI for Judges: A Framework". *Center for Security and Emerging Technology*, CSET Policy Brief (2021).

Barysè, Dovilè, and Roeë Sarel. "Algorithms in the court: does it matter which part of the judicial decision-making is automated?". *Artificial Intelligence and Law* (forthcoming) (2023).

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. New York, NY: Polity, 2019.

Bennett, Mark W. "Unspringing the Witness Memory and Demeanour Trap: What Every Judge and Juror Needs to Know about Cognitive Psychology and Witness Credibility". *American University Law Review* (2015)

Berkman Klein Center Working Group on AI Interpretability. "AI: Transparency and Explainability". *Berkman Klein Center for Internet & Society at Harvard University*. <https://cyber.harvard.edu/projects/ai-transparency-and-explainability>.

Bloch-Webba, Hannah. "Transparency's AI Problem". *Knight First Amendment Institute at Columbia University*. <https://knightcolumbia.org/content/transparencys-ai-problem>.

Bolt, Robert. *A Man for All Seasons*. London: Vintage 1963

Botoman, Alex. "Divisional Judge-Shopping". *Columbia Human Rights Law Review* 49, no.2 (2018): 297-344.

Brehm, Katie, Momori Kirabayashi, Clara Langevin, Bernardo Rivera Munozcano, Katsumi Sekizawa, and Jiayi Zhu. "The Future of AI in The Brazilian Judicial System". *The National Council of Justice Institute For Technology and Society of Rio De Janeiro*.

Burrell, Jenna. "How the machine 'thinks': Understanding opacity in machine learning algorithms". *Big Data & Society* 3 (2016): 1-12.

Campbell, Ray Worthy. "Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning". *Colorado Technology Law Journal* 18, no 2. (2020): 323-350.

Casciani, Dominic. "Chronic backlog of serious-crime cases hits courts". *BBC*. February 9, 2023. <https://www.bbc.co.uk/news/uk-64586483>.

Chaboryk, Amanda, Adam Sanitt, David Wilkins. "New Digital Case Law Service to Further Transform UK Litigation". *Oxford Business Law Blog*. July 27, 2022. <https://blogs.law.ox.ac.uk/business-law-blog/blog/2022/07/new-digital-case-law-service-further-transform-uk-litigation>.

Chen, Benjamin Minhao, Alexander Stremitzer, and Kevin Tobia. "Having Your Day in Robot Court". *Harvard Journal of Law and Technology* 36, no. 1 (2022), 127-169.

Chen, Edward M. "The Judiciary, Diversity and Justice for All". *California Law Review*, Vol.91 (July 2003)

Chesney, Bobby and Danielle Citron. "Deepfakes: A Looming Challenge for Privacy, Democracy and National Security". *California Law Review*, 107. (2019): 1776-1785

Chesnut, Rob. "'Judge Shopping' Will Only Get Easier Without Rule Changes". *Bloomberg*. August 21, 2019. <https://news.bloomberglaw.com/us-law-week/judge-shopping-will-only-get-easier-without-rule-changes>.

Citron, Danielle Keats, and Frank Pasquale. "The Scored Society: Due Process for Automated Predictions". *Washington Law Review* 89, no. 1 (2014).

Cohen, Mathilde. "Judicial Diversity in France: The Unspoken and the Unspeakable". *Law and Social Enquiry*. 43, no.4 (Autumn 2018): 1542-1570

Danzinger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. "Extraneous factors in judicial decisions". *PNAS* 108, no. 17 (2011): 6889-6892.

Delfino, Rebecca A. "Deepfakes on Trial: A Call to Expand the Trial Judges' Gatekeeping Role to Protect Legal Proceedings from Technological Fakery". *Hastings Law Journal* 74, no.2 (February 2023): 293-348

Deng, Jinting. "Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System". *Georgetown Law Technology Review* 223, no. 3 (2019): 223-280.

Dworkin, Ronald. "No Right Answer". *New York University Law Review* 53, no. 1 (1978): 1-32.

Ellery, Simon. "Fake photos of Pope Francis in a puffer jacket go viral, highlighting the power and peril of AI." *CBS*, 28 March, 2023.

<https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai/>

Eren, Ozkan, and Naci Mocan. "Emotional Judges and Unlucky Juveniles". *American Economic Journal: Applied Economics* 10, no. 3 (2018): 171-205.

eu-LISA and EUROJUST. "Artificial Intelligence Supporting Cross-Border Cooperation in Criminal Justice". *Publications Office of the European Union* (2022).

Evans, Justin W., and Anthony L. Gabel. "Legal Competitive Advantage and Legal Entrepreneurship: A Preliminary International Framework". *North Carolina Journal of International Law and Commercial Regulation* 39, no. 2 (2014): 333-422.

Facciola, John, and Nicholas Mignanelli. "Artificial Intelligence and the Courts: Materials for Judges". *American Association for the Advancement of Science* (2022).

Fazi, M. Beatrice. "Beyond Human: Deep Learning, Explainability and Representation". *Theory, Culture & Society* 38, no. 7 (2021): 55-75.

"Federal Rules of Evidence" (Washington: U.S. Government Publishing Office 2021)

Ferraro, Matthew and Brent Gurney. "The Other Side Says Your Evidence is a Deepfake. What Now?" *Law360*, 2. (Wilmer Hale: 21 December 2023)

Gans-Combe, Caroline. "Automated Justice: Issues, Benefits and Risks in the Use of Artificial Intelligence and Its Algorithms in Access to Justice and Law Enforcement". In *Ethics, Integrity and Policymaking: The Value of the Case Study*, edited by Dónal O'Mathúna, Ron Iphofen, 174-194. Berlin: Springer, 2022.

Georgallis, Panikos, João Albino-Pimentel, and Nina Kondratenko. "Jurisdiction shopping and foreign location choice: The role of market and nonmarket experience in the European solar energy industry". *Journal of International Business Studies* 52 (2020): 853-877.

Golby, Joel. "I thought I was immune to being fooled. Then I saw the Pope in a coat." *The Guardian*, 27 March, 2023.

<https://www.theguardian.com/commentisfree/2023/mar/27/pope-coat-ai-image-baby-boomers>

Guild, Blair and Elyse Samuels. "Pelosi videos manipulated to make her appear drunk are being shared on social media." *Washington Post*. Uploaded 23 May, 2019

Hao, Karen, and Jonathan Stray. "Can you make AI fairer than a judge? Play our courtroom algorithm game". *MIT Technology Review*. October 17, 2019.
<https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>.

Hart, H. L. A. *The Concept of Law*. Oxford: Clarendon Press, 1994.

Harris, Douglas. "Deepfakes: False Pornography is Here and the Law Cannot Protect You." *Duke Law and Technology Review*. (2019): 99-100

Hayes, Anthony, and Soodeh Saberian. "Temperature and Decisions: Evidence from 207,000 Court Cases". *American Economic Journal* 11, no. 2 (2019): 238-265.

House of Lords Artificial Intelligence Select Committee. "AI in the UK: Ready, Willing and Able". *Report Session 2017-2019* (2019).

Kamwa, Innocent, Subhransu Samantary, and Geza Joos. "On the accuracy versus transparency trade-off of data-mining models for fast -response PMU-based catastrophe predictors". *IEEE Transactions on Smart Grid* 3, no. 1 (2012): 152-161.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Kemp Smith. London: Macmillan 2007

Kehl, Danielle Leah, and Samuel Ari Kessler. "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing". *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School* (2017).

Kidd, Celeste, and Abeba Birhane. "How AI can distort human beliefs". *Science* 380, no.6651 (2023): 1222-1223.

Knight, Will "The Dark Secret at the Heart of AI" *MIT Technology Review* (April 2017)

Kwan, James, James Ng, and Brigitte Kiu. "The use of artificial intelligence in international arbitration: Where are we right now?". *International Arbitration Law Review* 22, no. 1 (2019): 19-26.

La Diega, Guido Noto. "Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection and Freedom of Information". *JIPITEC* 9, no. 3 (2018).

Ludwig, Jens, and Sendhil Mullainathan. "Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific". *Chicago Booth Research Paper*, no. 22-15 (2022).

Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, 1970.

McPeak, Agnieszka. "The Threat of Deepfakes in Litigation Raising the Authentication Bar to Combat Falsehood." 23 *Vanderbilt Journal of Entertainment and Technology Law* (2021): 433

Michaels, Andrew C. "Artificial Intelligence, Legal Change, and Separation of Powers". *University of Cincinnati Law Review* 88, no. 4 (2020): 1083-1103.

Moustafa, Tamir, and Tom Ginsburg. "Introduction: The Function of Courts in Authoritarian Politics". In *Rule By Law: The Politics of Courts in Authoritarian Countries*, edited by Tamir Moustafa and Tom Ginsburg, 1-22. Cambridge: Cambridge University Press, 2008.

Mozur, Paul. "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority". *New York Times*. April 14, 2019.
<https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>.

Niiler, Eric. "Can AI Be a Fair Judge in Court? Estonia Thinks So". *Wired*. March 25, 2019. <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>.

Oberman, William D. "Strategy and Tactic Choice in an Institutional Resource Context". In *Corporate Political Agency*, edited by Barry M. Mitnick. New York, NY: SAGE Publications, 1993.

Perel, Maayan, and Niva Elkin-Koren. "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement". *Florida Law Review* 69, no. 1 (2017): 181-221.

Perry, The Hon. Justice Melissa. "iDecide: the Legal Implications of Automated Decision-making" (lecture, University of Cambridge, September 15, 2014). <https://www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20140915>.

Pfefferkorn, Riana. "Deepfakes in the Courtroom" 29 *B.U. Pub. Int. L.J.* 245 (2020)

Rao, Dana. "Creating a 'lie detector' for deepfakes" interview by David Pogue, CBS, 29 January 2023, Audio, 1.28, <https://www.youtube.com/watch?v=Bf7AS1HBeaM>

Reynolds, Matt. "Courts and Lawyers struggle with growing prevalence of deepfakes." *American Bar Association Journal* (June 2020)

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." *arXiv.org*, 2016. <https://arxiv.org/abs/1602.04938>.

Ryder, Sir Ernest. "Securing Open Justice" (lecture, Max Plank Institute Luxembourg for Procedural Law & Saarland University, February 1, 2018).

Santus, Enrico, Nicolas Christin and Harshimi Jayram "Technology Factsheet Series, Artificial Intelligence" *Harvard Kennedy School*. (2020)

Schmelzer, Ron. "The Achilles' Heel of AI". *Forbes*. March 17, 2019. <https://www.forbes.com/sites/cognitiveworld/2019/03/07/the-achilles-heel-of-ai/>.

Schoenholtz, Andrew I., Jaya Ramji-Nogales, Phillip G. Schrag. "Refugee Roulette: Disparities in Asylum Adjudication". *Stanford Law Review* 60, no. 2. (2007): 295-411.

Shay, Lisa A., Woodrow Hartzog, John Nelson, and Gregory Conti. "Do robots dream of electric laws? An experiment in the law as algorithm". In *Robot Law*, edited by Ryan Calo, A. Michael Froomkin, and Ian Kerr. 274-305. Cheltenham: Edward Elgar Publishing, 2016.

Škop, Michal, Vincze Orsolya, Art Alishani, Goce Arsovski, and Krzysztof Izdebski. "AlGovrithms 2.0: The State of Play". *Open Data Kosovo* (2021).

Snow, C.P. *The Two Cultures and the Scientific Revolution: The Rede Lecture, 1959*. New York, NY: Cambridge University Press, 1961.

State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

Sunstein, Cass R. "The Use of Algorithms in Society". Preliminary draft (2013).

Taylor, Telford. *The Anatomy of the Nuremberg Trials: A Personal Memoir*. New York: 1992:146-9

Tyler, Tom R. *Why people obey the law*. Princeton, NJ: Princeton University Press, 2006.

United Nations Institute for Disarmament Research (UNIDR). "Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence". *UNIDR* (2018).

United States v. Doolin, No. 21-cr-00447 (D.D.C. 2. September 2022), ECF No. 151

United States v. Doolin, No. 21-cr-00447 (D.D.C. 5, August 2022), ECF No. 135

United States v. Doolin, No. 21-cr-00447 (D.D.C. Aug. 12, 2022), ECF No. 140

United States Department of Justice, 2023. “Florida Man Sentenced on Felony and Misdemeanor Charges for Actions During Jan 6. Capitol Breach” *Departmental Press Release*. 16 August 2023. <https://www.justice.gov/usao-dc/pr/florida-man-sentenced-felony-and-misdemeanor-charges-actions-during-jan-6-capitol-breach>

Venema, Agnes E and Zeno J. Geradts. “Digital Forensics, Deepfakes and the Legal Process” *SciTechlawyer*, Vol.16, no.4. (American Bar Association 2020)

West, Jevin and Carl Bergstorm (2019)

https://www.whichfaceisreal.com/results.php?r=1&p=1&i1=image-2019-02-17_123752.jpeg&i2=35757.jpeg

Whittaker, Lucas Andrew Park and Jan Keitzmann. “Synthetic Media: What Managers Need to Know about this Emergent Phenomenon.” *California Management Review*, Vol.65, no.1 (Berkeley 2022)

Xu, Zichun. “Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities”. *Applied Artificial Intelligence* 36, no. 1 (2022).

Yang, Crystal S., and Will Dobbie. “Equal Protection Under Algorithms: A New Statistical and Legal Framework”. *Michigan Law Review* 119, no. 2 (2020): 291-395.

Zalnieriute, Monika, Lyria Bennett Moses, and George Williams. “The Rule of Law and Automation of Government Decision-Making”. *Modern Law Review* 82, no. 3 (2019): 425-455.

Zou, Caixia. “Achievements and Prospects of Artificial Intelligence Judicature in China”. *Chinese Studies* 11, no. 4 (2022).

Zuckerman, Adrian. “Artificial Intelligence – Implications for the Legal Profession, Adversarial Process and the Rule of Law”. *UK Constitutional Law Association*. March 10, 2020. <https://ukconstitutionallaw.org/2020/03/10/artificial-intelligence-implications-for-the-legal-profession-adversarial-process-and-the-rule-of-law/>.