Technology and Regulation

Debate

All Rise for the Honourable Robot Judge?

Using Artificial Intelligence to Regulate AI

Simon Chesterman*
with responses by Lyria Bennett Moses** and Ugo Pagallo***

There is a rich literature on the challenges that AI poses to the legal order. But to what extent might such systems also offer part of the solution? China, which has among the least developed rules to regulate conduct by AI systems, is at the forefront of using that same technology in the courtroom. This is a double-edged sword, however, as its use implies a view of law that is instrumental, with parties to proceedings treated as means rather than ends. That, in turn, raises fundamental questions about the nature of law and authority: at base, whether law is reducible to code that can optimize the human condition, or if it must remain a site of contestation, of politics, and inextricably linked to institutions that are themselves accountable to a public. For many of the questions raised, the rational answer will be sufficient; but for others, what the answer is may be less important than how and why it was reached, and whom an affected population can hold to account for its consequences.

1. Introduction

The judge's robes are a deep black, though subtle touches of colour complement the national emblem dominating the courtroom wall. Red symbolizes revolution; golden stars rising over the Tiananmen Gate signify the unity of the people under the Party's leadership. Until the turn of the century, judicial officers wore military uniforms — the Supreme People's Court sits at the apex of the legal system but below the Communist Party. By appearance, this judge would not have even been in law school back then. Appearances can be deceiving, of course, since her generic face and simple hairstyle were designed by computer scientists. The avatar's lips move as the synthesized voice asks in Mandarin: 'Does the defendant have any objection to the nature of the judicial blockchain evidence submitted by the plaintiff?'

'No objection,' the human defendant responds.

The video of the pre-trial meeting at Hangzhou's Internet Court, released in late 2019, is part propaganda, part evangelism. Courts were identified as one of the areas ripe for improvement in China's New Generation Artificial Intelligence Development Plan. In a section on social governance [社会治理], it called for the creation of 'smart courts' [智慧法庭].[1]

This builds on moves to digitize and standardize litigation across the country, with experiments like those in Hangzhou paving the way for further advances. The avatar can handle online trade disputes, copyright cases, and e-commerce product liability claims.[2] Hangzhou was chosen because it is the home of Alibaba, enabling integration with trading platforms like Taobao for the purpose of evidence gathering as well as 'technical support'.[3]

Online dispute resolution is not new; eBay has long used it to help parties settle tens of millions of disputes annually.[4] What is interesting in the Chinese context is the extent to which this embrace of technology is permeating the court hierarchy not just in mediating small claims but all the way up to the Supreme People's Court itself.

The Judicial Accountability System [司法责任制] began as a campaign to promote consistency in judgments.[5] Past efforts had relied on

* David Marshall Professor of Law, Vice Provost (Educational Innovation), Dean of NUS College, National University of Singapore; Senior Director (AI Governance), AI Singapore. This article draws heavily on work first published as Simon Chesterman, We, the Robots? Regulating Artificial Intelligence and the Limits of the Law (Cambridge University Press, 2021).
** Lyria Bennett Moses is Director of the UNSW Allens Hub for Technology, Law and Innovation and a Professor and Associate Dean (Research) in the Faculty of Law and Justice at UNSW Sydney.
*** Ugo Pagallo is professor of Jurisprudence at the Department of Law, University of Turin (Italy)

Received 21 Aug 2023, Accepted 24 Sep 2023, Published 3 Oct 2023

1 国务院关于印发新一代人工智能发展规划的通知 [State Council Issued Notice of the New Generation Artificial Intelligence Development Plan] (State Council, Guofa [2017] No 35, 20 July 2017).
2 最高人民法院关于互联网法院审理案件若干问题的规定 [The Provisions of the Supreme People's Court on Several Issues Concerning the Trial of Cases by Internet Courts] (Supreme People's Court, Fasi [2018] No 16, 6 September 2018); Chuanman You, 'Law and Policy of Platform Economy in China' (2020) 39 *Computer Law & Security Review* 1.
3 DU Guodong and YU Meng, 'A Close Look at Hangzhou Internet Court', *China Justice Observer* (3 November 2019).
4 Pablo Cortés, *The Law of Consumer Redress in an Evolving Digital Market: Upgrading from Alternative to Online Dispute Resolution* (Cambridge University Press 2017) 8; Ethan Katsh and Orna Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (Oxford University Press 2017).
5 最高人民法院关于统一法律适用加强类案检索的指导意见（试行）

reviews by superiors, but this was deemed impractical and undermined the authority of the judge who heard the case.[6] AI systems now push similar cases up to a judge prior to a decision, flagging an 'abnormal judgment warning' if a proposed outcome departs significantly from past data.[7] This is part of a suite of technologies that have been adopted, influenced both by the supply of technology companies in China and the demands of a complex and developing legal system. The Wujiang District of Suzhou has trialled a 'one-click' summary judgment process, automatically generating proposed grounds of decision complete with sentence.[8] Other courts are following suit.[9]

Singapore's Chief Justice, Sundaresh Menon, has said that developments in China are making 'machine-assisted court adjudication a reality'. At the same time, he noted, the use of AI within the justice system gives rise to a 'unique set of ethical concerns, including those relating to credibility, transparency and accountability'.[10] To this one might add considerations of equity, since the drive towards greater automation is being dominated by deep-pocketed clients and ever-closer ties to technology companies, with uncertain consequences for the future administration of justice.[11]

The impact of AI on the practice of law goes well beyond the scope of this article.[12] It considers the narrower question of whether and how AI systems themselves could support regulation of AI. Insofar as gaps are revealed by the rise of fast, autonomous, and opaque systems, do new rules and new institutions need to be supplemented by new actors in the form of AI regulators and judges?

Section one briefly sketches out past efforts to automate the law. Though AI judges are the most provocative example,[13] many areas of legal practice and regulation have long been seen as ripe for automation. Despite successes in simple and repetitive tasks, these efforts tended to founder because they were premised on a misconception of law as the mere application of clear rules to agreed facts. In practice, the rules are rarely so clear and disagreement over facts explains a significant portion of legal disputes.

A more promising approach has been to abandon the goal of thinking 'like a lawyer' and approach legal analysis not as the application of rules to facts but as data. Section two discusses this bottom-up approach to legal analytics, which reveals distinct limitations that are not technical so much as social and political. Even though AI systems are getting ever better at forecasting regulatory outcomes, embracing this across the legal system would represent a fundamental shift from making decisions to predicting them.

Even if regulation by AI generally were possible, then, it is not desirable. Can a special case be made, however, for the regulation of AI systems themselves? If the objection to AI regulators and judges is their inability to appreciate the social context within which legal determinations take place, or legitimacy questions about humans having their fate determined by statistics, one response is that this need not apply to regulation of AI. Section three discusses how systems could be made to be self-policing. As we have seen in other areas, for example, one of the virtues of AI is relative transparency in that simulations can be run with slight variations to look for bias. And, unlike humans, a machine is far more likely to admit to its errors.[14]

To the extent that they increase the transparency and human control of AI systems, these developments may be useful. But self-regulation by AI ultimately confronts similar limitations to self-regulation by industry. Though helpful in establishing standards and best practices, red lines will need to be drawn and ultimate oversight conducted by politically legitimate and accountable actors. And, if it is impermissible to outsource inherently governmental functions to fast, autonomous, and opaque machines, enforcement of that prohibition cannot itself be left to those same machines.

2. Automating the Law

In the literature on AI and the law, an early theme was that legal practice — viewed essentially as the logical application of rules to established facts — was a strong candidate for automation. Though initially confined to theory,[15] in the 1980s researchers developed prototype systems based on manually created representations of rules in machine-readable form.[16] The enthusiasm was characteristic of the time, preceding as it did one of the 'AI winters' that has periodically seen inflated expectations crash against reality.[17]

Subsequent decades did see transformations in legal research and document management. These increased lawyers' access to information and their efficiency in using and sharing it, but did not fundamentally alter their role. Even those encouraging the adoption of technology believed that the inability of AI to emulate human qualities

[Guiding Opinions of the Supreme People's Court on Unifying the Application of Laws to Strengthen the Retrieval of Similar Cases (for Trial Implementation)] (Supreme People's Court, Fafa [2020] No 24, 27 July 2020).

6 Cf Margaret Y.K. Woo, 'Court Reform with Chinese Characteristics' (2017) 27 *Washington International Law Journal* 241; Junfeng Li et al, 'Artificial Intelligence Governed by Laws and Regulations' in Donghan Jin (ed), *Reconstructing Our Orders: Artificial Intelligence and Human Society* (Springer 2018) 61 at 67-71.

7 YU Meng and DU Guodong, 'Why Are Chinese Courts Turning to AI?', *The Diplomat* (19 January 2019).

8 '苏州法院刑案简易判决一键生成 [One-click Generation of the Summary Judgment of the Criminal Case in Suzhou Court]', 法制日报 *[Legal Daily]* (19 June 2017).

9 中国法院的互联网司法 [Chinese Courts and Internet Justice] (Supreme People's Court of the People's Republic of China, 2019) 63-65; Yadong Cui, *Artificial Intelligence and Judicial Modernization* (Springer 2020).

10 Sundaresh Menon, 'Opening of the Legal Year' (Supreme Court, Singapore, 7 January 2019).

11 Seth Katsuya Endo, 'Technological Opacity & Procedural Injustice' (2018) 59 *Boston College Law Review* 821.

12 See, eg, Richard Susskind, *The Future of Law: Facing the Challenges of Information Technology* (Oxford University Press 1996); Richard Susskind, *The End of Lawyers? Rethinking the Nature of Legal Services* (Oxford University Press 2008); Dory Reiling, *Technology for Justice: How Information Technology Can Support Judicial Reform* (Leiden University Press 2010); Richard Susskind, *Tomorrow's Lawyers: An Introduction to Your Future* (Oxford University Press 2013); Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge University Press 2017); Richard Susskind, *Online Courts and the Future of Justice* (Oxford University Press 2019); Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Hart 2020).

13 See, eg, Tania Sourdin, 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making' (2018) 41 *UNSW Law Journal* 1114; Eugene Volokh, 'Chief Justice Robots' (2019) 68 *Duke Law Journal* 1135.

14 See Simon Chesterman, 'Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity' (2021) 69 *American Journal of Comparative Law* 271, 284-85.

15 See, eg, L. Thorne McCarty, 'Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning' (1977) 90 *Harvard Law Review* 837.

16 See, eg, M.J. Sergot et al, 'The British Nationality Act as a Logic Program' (1986) 29 *Communications of the ACM* 370.

17 Anja Oskamp and Marc Lauritsen, 'AI in Law Practice? So Far, Not Much' (2002) 10 *Artificial Intelligence and Law* 227.

limited its scope for taking on the higher functions of lawyers — the role of judges in particular.[18] As we have seen in other areas, however, emulating human methods may not be the right or the best approach for reaping the benefits of AI. Autonomous vehicles, to pick an obvious case, are not driven by humanoid robots controlling speed and direction with mechanical hands and feet in substitution of their absent 'drivers'.

The DoNotPay chatbot, launched in 2015, offered an indication of what might be possible. Written by a seventeen-year-old Stanford student, it followed a series of rules to appeal against parking fines. Similar technology now facilitates other simple tasks from the making of wills to reporting suspected discrimination, yielding efficiencies as well as offering greater access to basic legal services for the wider public.[19] It is also leading to a re-evaluation of what the practice of law means, in the sense of a regulated profession. If a practising certificate or membership of a bar is required to offer legal advice, at what point does an automated system cross that line? Rules-based chatbots do not seem problematic, analogous to a textbook with a flowchart indicating how the law may handle various hypothetical situations. But if an AI system takes in new information, analyses it, and recommends a course of action in a manner that goes beyond the expertise of the programmer, does that become legal advice? Should it be regulated in the same manner as a lawyer?[20]

These are some of the questions raised by legal tech, a growing area of legal practice.[21] Having a lawyer sign off on advice is the current solution, much as a partner in a firm might approve a memo drafted in significant part by an intern.[22] That was the approach accompanying another high profile example of technology making inroads into the legal profession, when white-shoe law firm Baker & Hostetler announced that IBM's Ross was joining its bankruptcy practice.[23] Though routinely referred to as a 'robot lawyer', Ross was neither: a subscription service lacking any physical form (certainly not a humanoid one), it did not provide legal advice as such. It was, however, adept at sifting through vast numbers of documents for relevant information in support of the firm's cases.[24] Ross Intelligence announced in December 2020 that it was shutting down operations — defeated not by the limitations of its programming or the open-textured nature of law, but by a lawsuit from competitors.[25]

Many lawyers long assumed that litigation would be the last part of legal practice to be automated, though the example of China from the introduction to this article points to inroads being made there, also. Online dispute settlement has a long history and, for smaller claims in particular, has been embraced not only by online traders like eBay and PayPal, but also in the legal systems of Canada and Britain.

And yet the tsunami of change long forecast by Richard Susskind and others has not yet occurred.[26]

In part this is due to institutional resistance. Lawyers have defended their domain against encroachment by accounting firms and other actors; some view computers as just the next horde to be repelled.[27] As a profession, lawyers are also notoriously conservative. Though transactional lawyering must accommodate the needs of business, courtroom procedures retain elements both byzantine and archaic. The Covid-19 pandemic forced a reassessment of information technology in law firms and the courtroom.[28] Much as classes at schools and universities utilized video-conferencing services like Zoom, however, this was a change of medium rather than a transformation of the way in which law is practised.

A second reason the legal profession resisted radical change, and may continue to do so, is less self-serving. For it turns out that neither of the assumptions underpinning the hopes for widespread automation — that law is a contained logical system and that facts can be unambiguously established — withstands scrutiny.

2.1 The Inner Illogic of the Law

A preliminary problem is that legal rules are typically expressed in natural language that may be difficult for a computer to parse. This is a familiar issue in linguistics: humans often interpret language consistently, but not logically. Imagine an instruction to go shopping, for example, with the following request: 'Please buy me a newspaper; and if the store has bananas, buy six.' A naïve and literal interpretation could lead an autonomous agent to return with six copies of the newspaper. Similarly, the difference between saying that 'I hunted the bear with my wife' and 'I hunted the bear with my knife' is immedi-

18 Richard Susskind, 'Detmold's Refutation of Positivism and the Computer Judge' (1986) 49 *Modern Law Review* 125.
19 Paul Gowder, 'Transformative Legal Technology and the Rule of Law' (2018) 68(Supplement 1) *University of Toronto Law Journal* 82; Frank Pasquale, 'A Rule of Persons, Not Machines: The Limits of Legal Automation' (2019) 87 *George Washington Law Review* 1, 7-17. It is a stretch, however, to call this automation of certain legal processes 'AI' in any meaningful sense. See also Felicity Bell et al, AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators (Australasian Institute of Judicial Administration, 2022).
20 In October 2019, for example, the Hanseatic Bar Association Hamburg successfully challenged Smartlaw, a bot operated by Wolters Kluwer, in the district court of Cologne for operating inconsistently with Germany's Legal Services Act [*Rechtsdienstleistungsgesetz]. See further* Michael Stockdale and Rebecca Mitchell, 'Legal Advice Privilege and Artificial Legal Intelligence: Can Robots Give Privileged Legal Advice?' (2019) 23 *International Journal of Evidence & Proof* 422; Polly Botsford, Future of Law: Courts Debate Legality of Legal 'Bots' (International Bar Association, 11 March 2020).
21 Sanda Erdelez and Sheila O'Hare, 'Legal Informatics: Application of Information Technology in Law' (1997) 32 *Annual Review of Information Science and Technology* 367; Jens Frankenreiter and Michael A. Livermore, 'Computational Methods in Legal Analysis' (2020) 16 *Annual Review of Law and Social Science* 39.
22 See, eg, Model Rules of Professional Conduct (American Bar Association, 2020), rule 5.3 (responsibilities regarding nonlawyer assistance — though the language of the rule clearly assumes that such assistance comes from a 'person'). Cf Ed Walters, 'The Model Rules of Autonomous Conduct: Ethical Responsibilities of Lawyers and Artificial Intelligence' (2019) 35 *Georgia State University Law Review* 1073; Anthony E. Davis, 'The Future of Law Firms (and Lawyers) in the Age of Artificial Intelligence' (2020) 27(1) *The Professional Lawyer* 3.
23 Michal Addady, 'Meet Ross, the World's First Robot Lawyer', *Forbes* (12 May 2016).
24 See, eg, Dena Dervanović, 'I, Inhuman Lawyer: Developing Artificial Intelligence in the Legal Profession' in Marcelo Corrales, Mark Fenwick, and Nikolaus Forgó (eds), *Robotics, AI and the Future of Law* (Springer 2018) 209 at 226-27; Sergio Alberto Gramitto Ricci, 'Artificial Agents in Corporate Boardrooms' (2020) 105 *Cornell Law Review* 869, 876.
25 Rhys Dipshan, 'ROSS Shuts Down Operations, Citing Financial Burden From Thomson Reuters Lawsuit', *Law.com* (11 December 2020).
26 See, eg, Susskind (n 12).
27 Chay Brooks, Cristian Gherhes, and Tim Vorley, 'Artificial Intelligence in the Legal Sector: Pressures and Challenges of Transformation' (2020) 13 *Cambridge Journal of Regions, Economy, and Society* 135, 148.
28 Julie Marie Baldwin, John M. Eassey, and Erika J. Brooke, 'Court Operations During the COVID-19 Pandemic' (2020) 45 *American Journal of Criminal Justice* 743. Cf Daphne Yong, 'The Courtroom Performance' (1985) 10(3) *The Cambridge Journal of Anthropology* 74.

ately clear to a human but requires additional information outside the text to make sense.[29] Sometimes language may be inherently ambiguous. The statement that 'I saw the girl with the telescope' might mean either that the speaker looked through a telescope or that the girl was carrying one.

Advances in natural language processing have overcome many of these difficulties, though statutes and case law may be more challenging than the average text.[30] Indeed, the profession of law depends on the ability to charge clients for advice as to how to structure their activities to comply with the law, and advocating on their behalf to enforce it in support of their interests. There may be multiple plausible constructions of a given text — even a carefully drafted one. And until statutes and judgments are written in a manner that can be represented using formal logic, the authoritative text is the original one.[31]

This points to a more fundamental problem, which is that many laws are not reducible to logical representation.[32] To be sure, some may be. Road traffic laws, for example, state that exceeding a given speed limit constitutes an offence. Many jurisdictions use speed cameras that automatically record infringements and issue fines. Yet it is telling that these laws — among the most commonly experienced, for much of the population — rarely feature in law school curricula, precisely because they are so clear.[33]

Others are not. The tort of negligence, for example, is not representable as duty of care plus breach plus causation minus defences equals liability. It explicitly incorporates judgments based on human experience — the famous 'man on the Clapham omnibus'[34] — and notions of reasonableness. In other areas of law, terms such as 'good faith' or 'unconscionability' are notoriously difficult to define in terms that would be useful to a machine.[35] Pretending otherwise is to delegate the interpretive task from the judge not to the machine but to the programmer who establishes its parameters.[36] More formally, it is sometimes argued that efforts to treat the law as a logical system susceptible to automation will fail due to the necessary incompleteness of that system — and all such systems.[37]

In any case, few legal theorists today would adhere to a strictly formalist position that law can or should be interpreted mechanically. Ronald Dworkin, for example, did hold that there is one correct answer to legal questions — even the difficult ones — but he explicitly rejected the notion that this implied that the answer was reachable by a computer designed by an 'electronic magician'.[38] On the contrary, the difficulty in applying the law is that it is always an exercise in political morality, interpreting the law in its best light on behalf of a community in search of a justification for state coercion.[39] Joseph Raz rejected Dworkin's view of uniquely correct solutions, arguing that judges in such cases are analogous to subordinate legislators, with legal duties to enact particular rules.[40] The positivist tradition is often seen as the most sympathetic to automation of legal processes, but even HLA Hart held that judges must make choices where existing law fails to dictate that any decision is the 'correct' one.[41] Legal realists and critical legal studies scholars, who emphasize the role of judges and the influence of power on the social order, would regard the question of automating the law as so ridiculous to not be worth taking seriously.[42]

2.2 In Fact

In his confirmation hearings before the US Senate, Chief Justice John Roberts deflected criticisms of partisanship by quipping that his job was merely 'to call balls and strikes'. The answer was disingenuous regarding the politicized nature of the court, but Roberts also underestimated the moves to automation in major league sport. In baseball in particular, there have been many calls for umpires to be assisted by a computerized strike zone or replaced entirely. If the role of judges was as simple as determining whether a leather encased ball passed within a three-square-foot zone or not, then they probably should be replaced by machines — it would be both more efficient and consistent.[43]

Even if a law appears on its face to be expressed clearly, however — 'no vehicles in the park', to pick a well-known example first offered by Hart — how it is to be applied in practice may be less so. We might agree that it covers automobiles, but what about bicycles, roller skates, toy cars?[44] How about a stroller? Or the statue of a Second World War tank?[45]

29 Ian McEwan, *Machines Like Me* (Vintage 2019) 178.

30 See, eg, Livio Robaldo et al, 'Introduction for Artificial Intelligence and Law: Special Issue "Natural Language Processing for Legal Texts"' (2019) 27 *Artificial Intelligence and Law* 113; Loïc Vial, Benjamin Lecouteux, and Didier Schwab, 'Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation' (2019) arXiv 1905.05677v3; Boon Peng Yap, Andrew Koh, and Eng Siong Chng, 'Adapting BERT for Word Sense Disambiguation with Gloss Selection Objective and Example Sentences' (2020) arXiv 2009.11795v2; Zakaria Kaddari et al, 'Natural Language Processing: Challenges and Future Directions' in Tawfik Masrour, Ibtissam El Hassani, and Anass Cherrafi (eds), *Artificial Intelligence and Industrial Applications* (Springer 2021) 236.

31 L. Karl Branting, 'Artificial Intelligence and the Law from a Research Perspective' (2018) 14(3) *Scitech Lawyer* 32.

32 Cf H. Patrick Glenn and Lionel D. Smith (eds), *Law and the New Logics* (Cambridge University Press 2017).

33 Note that many jurisdictions allow 'reasonable excuse' as a defence, so perhaps even this example is not so simple.

34 *McQuire v. Western Morning News* (1903) [1903] 2 K.B. 100, 109 (Collins MR).

35 See, eg, Mindy Chen-Wishart and Victoria Dixon, 'Humble Good Faith: 3 x 4' in Paul Miller and John Oberdiek (eds), *Oxford Studies in Private Law Theory* (Oxford University Press 2020) forthcoming.

36 Francesco Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings' (2020) 2(1) *Law, Technology, and Humans* 4, 7.

37 C.F. Huws and J.C. Finnis, 'On Computable Numbers with an Application to the AlanTuringproblem' (2017) 25 *Artificial Intelligence and Law* 181, 183.

38 Ronald Dworkin, *Law's Empire* (Harvard University Press 1986) 412.

39 Brian Sheppard, 'Warming Up to Inscrutability: How Technology Could Challenge Our Concept of Law' (2018) 68(Supplement 1) *University of Toronto Law Journal* 36, 60.

40 Joseph Raz, *Ethics in the Public Domain: Essays in the Morality of Law and Politics* (Clarendon Press 1995) 249-50.

41 H.L.A. Hart, *The Concept of Law* (3rd edn, Clarendon Press 2012) 273. Cf Abdul Paliwala, 'Rediscovering Artificial Intelligence and Law: An Inadequate Jurisprudence?' (2016) 30 *International Review of Law, Computers & Technology* 107.

42 Cf Sangchul Park and Haksoo Ko, 'Machine Learning and Law and Economics: A Preliminary Overview' (202) 11(2) Asian Journal of Law and Economics, 15 (adopting a law and economics analysis and concluding that such systems might be treated as expert witnesses but not as substituting for the human judge).

43 Jennifer Walker Elrod, 'Trial by Siri: AI Comes to the Courtroom' (2020) 57 Houston Law Review 1085; Mariano-Florentino Cuéllar and Aziz Z. Huq, 'Artificially Intelligent Regulation' (2022) 151 *Daedalus* 335.

44 H.L.A. Hart, 'Positivism and the Separation of Law and Morals' (1958) 71 *Harvard Law Review* 593, 607.

45 Pierre Schlag, 'No Vehicles in the Park' (1999) 23 *Seattle University Law Review* 381; Frederick Schauer, 'A Critical Guide to Vehicles in the Park' (2008) 83 *New York University Law Review* 1109.

The underlying problem is that the strength and the weakness of language is that it is open textured, an idea traceable back to Wittgenstein.[46] Even when there may be near-universal agreement on many applications of the law, marginal cases will arise. The open-textured nature of language and law has an important connection to time, since future cases may arise that were unknowable by the drafter of a rule. Twentieth century legislators, for example, could be forgiven for failing to contemplate whether the vehicles prohibited from entering the park include drones.[47]

The need for flexibility in applying the law to particular facts is not merely hypothetical. In the late nineteenth century, the New York State Court of Appeals heard a case in which the plain language of a will and the relevant legislation made clear that the grandson of Francis B Palmer should inherit his estate. Yet the fact that the younger Mr Palmer had poisoned his late grandfather gave them pause. Dworkin uses this example to argue that nearly universal principles of justice may require a departure even from clear textual rules. (The murderer did not get his inheritance.)[48]

Perhaps the strongest illustration of the difficulty of applying law to facts is the market for legal services, in particular litigation. If laws were clearly drafted and easily applied, few disputes would go to court because rational, well-informed actors would reach the correct conclusion on their own. There would be no need for appellate courts. The reason cases end up in court is only rarely because one side is objectively and obviously 'wrong'. This is borne out in practice. Assuming that potential litigants in civil suits make rational estimates of the likely outcome at trial, for example, the individual maximizing decisions of parties should mean that their success rate approaches 50 percent, regardless of the substantive area of law.[49] That figure is a limit case only — approached as the standard of decision is clearer, parties' estimate of the quality of their own cases is more accurate, and the stakes on either side are of similar value. But it finds empirical support.[50]

3. Law as Data

Inherent in many of the debates over AI and legal regulation are fundamental differences in the understanding not of AI, but of law. If law is understood in a narrowly formalistic way — the blind application of rules to uncontested facts — then processing it through algorithms makes sense, in the same way that it would be inefficient to have regulators or judges doing long division by hand instead of using a calculator.[51] But, to state the obvious, law is not long division. The simplest of cases aside, regulation of behaviour and the resolution of disputes is an inherently agonistic enterprise that involves values and meaning that are necessarily contested.[52] As Oliver Wendell Holmes famously said, 'The life of the law has not been logic: it has been experience.'[53]

Ah yes, the computer scientist might respond. But experience is precisely what machine learning can replicate now.

Indeed, more recent innovations reflect a shift in the approach to the law analogous to the move in AI research towards machine learning. Rather than trying to encode legal rules in fixed systems that can then be applied to sanitized facts — top down, as it were — key achievements have been made in analysing large amounts of data from the bottom up. This approach does not seek to answer an individual case, but offer a prediction as to the outcome based on past experience.[54] It represents, as Mireille Hildebrandt observes, a shift 'from reason to statistics and from argumentation to simulation'.[55]

The turn to AI in this context has proven useful in identifying relevance for the purposes of legal research, contract review, and discovery.[56] But if extended to regulation and adjudication it would fundamentally change the task from making a decision to predicting it.[57] Rather than being part of an ongoing social process in the development of the law, such determinations are more akin to forecasting the weather.[58] Analytics may provide more information to disputing parties and encourage efficient resolution of disputes while reducing bias and error,[59] but they could not be a replacement of the judicial function itself.[60]

Indeed, in some jurisdictions the approach has been met with outright hostility. In 2019, for example, France adopted a law prohibiting the publication of data analytics that reveal or predict how particular judges decide on cases, with a maximum punishment of five years in prison.[61] Though France will likely remain an outlier, AI systems will not replace lawyers or judges in the near term. A more probable scenario is increasing use of AI systems

46 Hart (n 41) 124; Ralf Poscher, 'Ambiguity and Vagueness in Legal Interpretation' in Lawrence M. Solan and Peter M. Tiersma (eds), *The Oxford Handbook of Language and Law* (Oxford University Press 2012) 128.
47 Michael A. Livermore, 'Rules by Rules' in Ryan Whalen (ed), *Computational Legal Studies: The Promise and Challenge of Data-Driven Research* (Edward Elgar 2020) 238 at 246-47.
48 Ronald Dworkin, *Taking Rights Seriously* (Harvard University Press 1977) 23, citing *Riggs v. Palmer*, 115 N.Y. 506 (1889).
49 George L. Priest and Benjamin Klein, 'The Selection of Disputes for Litigation' (1984) 13 *Journal of Legal Studies* 1.
50 Simon Chesterman, 'Do Better Lawyers Win More Often? Measures of Advocate Quality and Their Impact in Singapore's Supreme Court' (2021) *Asian Journal of Comparative Law* forthcoming.
51 Mireille Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (2016) 79 *Modern Law Review* 1. For an example of using AI to rethink the notion of legal logic, see Douglas Walton, *Argumentation Methods for Artificial Intelligence in Law* (Springer 2005).
52 Jeremy Waldron, 'The Rule of Law and the Importance of Procedure' in James E. Fleming (ed), *Nomos L: Getting to the Rule of Law* (New York University Press 2011) 3 at 22.
53 Oliver Wendell Holmes, Jr., *The Common Law* (Little, Brown 1881) 1.
54 Maxi Scherer, 'Artificial Intelligence and Legal Decision-Making: The Wide Open?' (2019) 36 *Journal of International Arbitration* 539, 569-71. See, eg, Nikolaos Aletras et al, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2:e93 *PeerJ Computer Science*.
55 Mireille Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2018) 68(Supplement 1) *University of Toronto Law Journal* 12, 29.
56 See Robert Dale, 'Law and Word Order: NLP in Legal Tech' (2019) 25 *Natural Language Engineering* 211.
57 Cf Oliver Wendell Holmes, Jr., 'The Path of the Law' (1897) 10 *Harvard Law Review* 457, 461 ('The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by law').
58 Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) 68(Supplement 1) *University of Toronto Law Journal* 63, 64-65.
59 Daniel L. Chen, 'Judicial Analytics and the Great Transformation of American Law' (2019) 27 *Artificial Intelligence and Law* 15.
60 Compare the requirements of transparency in medicine, where opaque treatments are routinely prescribed: Chesterman (n 14) 275.
61 Loi no 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice 2019 (France), art 33; 'France Bans Judge Analytics, 5 Years in Prison for Rule Breakers', *Artificial Lawyer* (4 June 2019).

as part of legal services, a partnership sometimes compared to the pairing of humans and machines to play advanced chess, also known as centaur or cyborg chess.[62]

In this context, it is common to draw a distinction between technology assisting in the retrieval of information and in the exercise of judgment.[63] The former is analogous to use of a calculator and deemed unproblematic; the latter raises troubling questions about who is exercising discretion. But when the 'information' being retrieved goes to the heart of a decision, that distinction may be artificial. As we have seen in other areas, reliance on opaque systems to make recommendations on matters like sentencing are an abdication of the judicial function not because they may be incorrect but because they are illegitimate.[64] More generally, automation bias raises concerns that human agency may diminish in favour of reliance on the machine.[65] Even for sophisticated decision-makers, it can be difficult to tell where an algorithm's 'nudge' ends and the accountable individual's choice begins.[66]

For present purposes, it is sufficient to conclude that AI will continue to transform the legal profession and the role of lawyers — but not to replace them completely. The limits are not so much technical as inherent in the nature of law and the legitimacy accorded to it through political structures in most well-ordered societies.

4. Law as Code

Is there, however, a special case to be made for AI playing a larger role in regulating AI itself?

The speed, autonomy, and opacity of AI systems do occasionally give rise to practical and conceptual difficulties for human regulators. In some cases, the response has been to slow them down, as in the case of high-frequency trading.[67] In others, it has been to ensure the possibility of accountability through requiring that actions be attributable to traditional legal persons — typically the owner, operator, or manufacturer.[68] In still others, it has been to call for prohibiting certain activities entirely — most prominently the use of lethal force.[69]

AI does offer means of supporting regulation of AI, though the traditional justifications for regulation do not translate easily onto AI systems themselves. In particular, in the absence of AI with legal personality,[70] the targets of regulation are not the AI systems themselves but those who own, operate, and make those systems. That said, the unique features of AI suggest two avenues for a form of self-regulation. First, regulatory objectives can be built into the software itself. Analogous to requirements that privacy values be incorporated into software harvesting personal data, this may be termed regulation by design. Secondly, AI systems allow for interrogation of mistakes and adverse outcomes in a manner not possible with traditional legal actors. This should enable greater transparency concerning errors, but the consequences should also be different than for traditional legal persons. It will be described here as regulation by debugging.

4.1 Regulation by Design

The idea of incorporating law-compliant behaviour into an AI system may seem self-evident. Autonomous vehicles should comply with traffic laws; algorithms allocating social benefits or recommending loans should not discriminate on the basis of gender or race. But it is possible to go far beyond this.

The notion that regulation can be achieved through design is not new. Though legal scholars often focus on 'command and control' approaches, design standards can gather information, set standards, and shape behaviour for regulatory ends.[71] The usual tools of regulation — commands, incentives, influence — presume the need to compel or persuade human actors (or their corporate proxies) to do or refrain from doing certain actions.[72] Programmable devices and systems, which include most applications of AI considered here, offer the possibility of incorporating regulatory standards directly into their code.

There are limits. As I've argued elsewhere, proposals analogous to Asimov's laws of robotics misconceive the nature of law and will never be a complete solution to the regulatory challenges posed by AI systems.[73] But as a restriction on what such systems can do, they point to a promising path forward. Effective standard-setting will, in some cases, require global rules.[74] Yet implementing those rules should not rely upon state enforcement alone — to the extent possible, they should be encoded into AI systems themselves. As for the content of those rules, most will be the same that would apply to any product or service. Rather than requiring robots not to murder humans, for example, the prohibition would be against producers making devices that could do so.

Of more interest is how regulation by design might support the two areas that do suggest potential gaps: human control and transparency.

62 See, eg, Rebecca Crootof, '"Cyborg Justice" and the Risk of Technological–Legal Lock-In' (2019) 119 *Columbia Law Review Forum* 233, 243; John Morison and Adam Harkens, 'Re-engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making' (2019) 39 *Legal Studies* 618, 634-35.

63 Zihuan Xu et al, 'Case Facts Analysis Method Based on Deep Learning' in Weiwei Ni et al (eds), *Web Information Systems and Applications* (Springer 2020) 92.

64 Chesterman (n 14).

65 Raja Parasuraman and Dietrich Manzey, 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52 *Human Factors* 381, 392.

66 Mariano-Florentino Cuéllar, 'Cyberdelegation and the Administrative State' in Nicholas R. Parrillo (ed), *Administrative Law from the Inside Out: Essays on Themes in the Work of Jerry L. Mashaw* (Cambridge University Press 2017) 134 at 159.

67 See Simon Chesterman, '"Move Fast and Break Things": Law, Technology, and the Problem of Speed' (2021) 33 *Singapore Academy of Law Journal* 5.

68 See Simon Chesterman, *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law* (Cambridge University Press 2021) 88-91.

69 See Simon Chesterman, 'Artificial Intelligence and the Problem of Autonomy' (2020) 1 *Notre Dame Journal on Emerging Technologies* 210, 234-37.

70 See Simon Chesterman, 'Artificial Intelligence and the Limits of Legal Personality' (2020) 69 *International and Comparative Law Quarterly* 819.

71 Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 *Information, Communication & Society* 118, 120. See generally Lawrence Lessig, *Code: Version 2.0* (first published 1999, Basic Books 2006); Mireille Hildebrandt, 'Saved by Design? The Case of Legal Protection by Design' (2017) 11 *Nanoethics* 307; Nynke Tromp and Paul Hekkert, *Designing for Society: Products and Services for a Better World* (Bloomsbury Visual Arts 2019).

72 See Chesterman (n 68) 185-92.

73 See ibid 192-94.

74 See Simon Chesterman, 'Weapons of Mass Disruption: Artificial Intelligence and International Law' (2021) 10 *Cambridge International Law Journal* 181.

On human control, building in capability restrictions and a 'kill switch' may sound like obvious design solutions. For the time being, that is true — though I've argued for a global agency to support a ban on the creation of uncontrollable or uncontainable AI.[75] Projecting into the future, however, if the emergence of a superintelligence ever moves from science fiction to plausible reality, such constraints could bring about the evil that they are intended to prevent; it may be more prudent to seek to instil alignment with human values instead.[76]

In terms of transparency, different degrees are appropriate depending on the type of decision or activity in question. Generally, however, AI systems should be designed to identify themselves as such and in a manner that enables identification of a legal person who is the owner, operator, or manufacturer.[77] In addition, systems should at a minimum maintain a basic audit trail of how decisions are made.[78] This points to the second way in which AI could assist in its own regulation, which is through enabling interrogation of its failures.

4.2 Regulation by Debugging

When one human kills another, it may give rise to criminal prosecution and lawsuits — these raise legal questions to be resolved. When a machine kills someone, there may be an investigation of its owner, operator, or manufacturer. But with regard to the machine itself, the problem is more likely seen to be an engineering one. Much as airplane crashes are studied using information from flight data recorders, audit trails in AI systems offer the chance to review how and why errors occurred. If these disclose culpability on the part of the owner, operator, or manufacturer, legal remedies may follow. As for the AI system itself, however, punishment for an error would make no more sense than punishing a plane for its engine failure.

If a system is deemed unsafe it may be removed from the market; a more likely scenario is that it would be improved. Much as software is now continuously updated with patches as bugs and vulnerabilities are discovered, AI systems operating in the world should be expected to evolve in response to their environment. Market pressure will encourage such updates, but they could also be the subject of regulations or a court order.[79]

Debugging in this way satisfies the aims of regulation at far less cost. Assuming the improvements do not introduce other errors, it may also be more reliable than traditional regulatory tools if an AI system cannot be tempted once more into deviance. It presumes, of course, a degree of transparency that is unavailable in traditional regulatory settings. If one asks a human driver whether she ran a red light, or a human manager if he discriminated on the basis of race, the answer may be unreliable. Proper audit logs should avoid this problem with respect to AI systems.

This ability to get straight answers also points to another potential strength of such systems, which is that they could be tasked with monitoring themselves. Two broad theories of oversight are known as 'police patrol' and 'fire alarm', depending on whether it is conducted through periodic surveys or waiting for problems to be escalated.[80]

75 See ibid 192-98.
76 See Chesterman (n 68) 138-41.
77 See ibid 213-16.
78 See ibid 156-57.
79 Mark A. Lemley and Bryan Casey, 'Remedies for Robots' (2019) 86 *University of Chicago Law Review* 1311, 1386-89.

AI systems offer a third possibility of self-investigation. This would be more than a regime of self-regulation, as it would not rely on the good faith of actors with incentives to defect. Provided the instructions were clear, a system could report on its compliance with rules and policies, among other things examining its conduct for bias with a degree of candour not possible with humans.[81] Problems disclosed in this way would also point to a need to rethink the remedies available — not as sins to be punished, but errors to be corrected.

5. The Prospects for Regulation

After the avatar's brief interaction with the parties concludes, the video celebrating Hangzhou's Internet Court shows an interview with its very human Vice President, Ni Defeng. 'What we are doing now,' he enthuses, 'you can't understand it as merely improving efficiency. It also speaks to the issue of legal justice. The faster speed — is kind of justice on its own, because justice delayed is justice denied.'

The desire for efficiency and consistency is driving China's push to digitize its court system, with strong endorsement by government as well as the judiciary, and strong support from industry. Though judges themselves remain, for the most part, human, Shanghai's courts are replacing law clerks with AI systems to perform basic legal research — another step in the push to modernize the judicial system through the use of technology.[82] These developments have been matched by the embrace of computational legal studies in Chinese legal academia. The past decade has seen a turn to empirical legal studies more comprehensive than in the United States; computational methods are now routinely used in articles published in the top generalist Chinese law journals.[83]

A partial explanation of the greater traction of computational approaches in theory as well as practice is that China's embrace of the rule of law is more instrumental than its Western counterparts.[84] Chinese judges refer to interpretation and the exercise of discretion in the context of 'judicial measurement' [裁判尺度], a term without a precise equivalent in the Western tradition but routinely invoked in China with a view to unifying judicial standards.[85] Judgments at the district and intermediate level tend to be short — a couple of

80 Mathew D. McCubbins and Thomas Schwartz, 'Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms' (1984) 28 *American Journal of Political Science* 165.
81 But see Casey Chu, Andrey Zhmoginov, and Mark Sandler, 'CycleGAN, a Master of Steganography' (2017) 1712.02950v2 arXiv; Joel Lehman et al, 'The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities' (2018) arXiv 1803.03453v1; Tom Simonite, 'When Bots Teach Themselves to Cheat', *Wired* (8 August 2018) (describing AI systems that learned to 'cheat').
82 Sarah Dai, 'Shanghai Judicial Courts Start to Replace Clerks with AI Assistants', *South China Morning Post* (1 April 2020).
83 Yingmao Tang and John Zhuang Liu, 'Computational Legal Studies in China: Progress, Challenges, and Future' in Ryan Whalen (ed), *Computational Legal Studies: The Promise and Challenge of Data-Driven Research* (Edward Elgar 2020) 124.
84 Randall Peerenboom, *China's Long March Toward Rule of Law* (Cambridge University Press 2002) 280-330; Cong-rui Qiao, 'Jurisprudent Shift in China: A Functional Interpretation' (2017) 8(1) *Asian Journal of Law and Economics*; Simon Chesterman, 'Can International Law Survive a Rising China?' (2020) 31 *European Journal of International Law* 1507.
85 统一裁判尺度 规范法律适用 [Uniform Judgment Standards and Standardize the Application of Law] (Supreme People's Court of the People's Republic of China, 12 January 2018). Cf JIANG Na, 'Old Wine in New Bottles? New Strategies for Judicial Accountability in China' (2018) 52 *International Journal of Law, Crime and Justice* 74.

paragraphs stating the facts, an outline of the applicable law and responses to the parties' arguments, and a decision.

Nevertheless, Chinese judges also express wariness about 'black box' decision-making.[86] In part this is due to concerns about the accuracy of the outcomes. Initial efforts to train computers on murder cases had to be shelved, for example, because there was an insufficient number of cases and the facts in each varied so greatly.[87] But it also goes to the trust that underpins the legal system and the rule of law itself.

It remains to be seen whether China represents the future of regulation by AI or its limit case. This article has argued that some of the qualities of AI systems that make them hard to regulate through traditional processes may also offer tools to regulate them through new ones. Regulation by design and regulation by debugging suggest ways in which AI systems can be built to comply with the law and tasked with investigating their own biases and failings in a way that most humans would find uncomfortable or impossible.

Yet there are limits to this role. Even if AI systems are more efficient and more consistent than human regulators and judges, that would not justify the handover of their powers more generally.

For the authority of law depends not only on its processes in a formal sense but in a substantive sense also. Regulation, legal decisions, are not mere Turing Tests in which we speculate whether the public can guess if the regulator or judge is a person or a robot. Legitimacy lies in the process itself, the ability to tie the exercise of discretion to a being capable of weighing uncertain values and standing behind that exercise of discretion.[88] Accepting otherwise would be to accept that legal reasoning is not a mix of doctrinal, normative, and interdisciplinary scholarship. Rather, it would come to be seen as a kind of history — the emphasis on appropriate categorization of past practice rather than participation in a forward-looking social project.[89]

As Robert H Jackson, another US Supreme Court judge, once observed: 'We are not final because we are infallible, but we are infallible only because we are final.'[90] Many decisions might therefore properly be handed over to the machines. But the final exercise of discretion, public control over the legal processes that regulate our interactions with the world around us, should only be transferred when we are also prepared to transfer political control also — when we give up the ballot box for the X-Box.

86 郭富民 [GUO Fumin], '人工智能无法取代法官的审慎艺术 [Artificial Intelligence Cannot Replace the Prudential Art of Judges]', 中国法院网 *[China Court Network]* (5 July 2017); Jie-jing Yao and Peng Hui, 'Research on the Application of Artificial Intelligence in Judicial Trial: Experience from China' (2020) 1487 *Journal of Physics: Conference Series* 012013, 4.
87 Jinting Deng, 'Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System' (2019) 3 *Georgetown Law Technology Review* 223, 275.
88 See also John Tasioulas's discussion of the importance of reciprocity in the rule of law: John Tasioulas, 'The Rule of Algorithm and the Rule of Law' (forthcoming).
89 Cf Michael A. Livermore (ed), *Law as Data: Computation, Text, and the Future of Legal Analysis* (Santa Fe Institute Press 2019).
90 *Brown v. Allen*, 344 U.S. 443, 540 (1953) (Jackson, J concurring).

Copyright (c) 2023, Simon Chesterman.

Creative Commons License
This work is licensed under a Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License.

Comment on ‘All Rise for the Honourable Robot Judge?’

Lyria Bennett Moses

Simon Chesterman’s article is not only a thought-provoking discussion about the use of AI by the judiciary, it also makes an intriguing connection between the idea of AI judges and the use of AI as a mechanism to regulate AI. In particular, Chesterman discusses whether AI regulating AI is a ‘special case’ exception to the undesirability of ‘regulation by AI’, linked to the problems he identifies with the use of AI in legal institutions. In this commentary, I discuss the relationship between these two core ideas in Chesterman’s article – in what sense is AI regulating AI related to use of AI in formal legal decision-making?

The article begins with the idea of AI in the judicial system, using the extreme example of China’s proposal for avatar judges handling disputes alongside the reality of their Judicial Accountability System that guides judges to enhance ‘consistency’ in the legal system. The article outlines critiques of both ‘top-down’ rules-based automation and ‘bottom up’ data-driven prediction as methods for using AI to replace or support judges, before asking whether the use of AI to regulate AI might be a special case.

Chesterman describes two ways in which AI might regulate AI. The first is to build regulatory objectives into the software itself, a method often described as regulation or compliance ‘by design’. This can be achieved through self-regulation (programmers choose to build compliance features into systems) or as a result of legal requirements (for example, a requirement that AI systems not misleadingly impersonate a human). The second is what Chesterman names ‘regulation by debugging’, which refers to how systems can better respond to errors and mistakes. This involves features such as audit logs, not only to capture the circumstances of any error or mistake, but also potentially to facilitate self-reporting of compliance.

To understand the connection between these two ideas in Chesterman’s article, it is useful to pause and consider the relationship between law and regulation. These terms are commonly connected, but occasionally confused in legal and policy discourse. An example of potentially constructive confusion is Lessig’s statement that ‘code is law’[1] – Lessig did not mean that computer code on its own had the status of a legal instrument but rather than code, like law, can be used to regulate. Another example is where the term Regulation (capitalized) is used to describe a kind of legal instrument, as in the General Data Protection Regulation in Europe or the Corporations Regulations 2001 (Cth) in Australia.

The term ‘regulation’ is not defined in Chesterman’s article, although the book from which the article is drawn[2] focuses on public control of a set of activities. This is narrower, but falls within the definition of ‘regulation’ offered by Yeung building on Black, being “organized attempts to manage risk or behavior in order to address a collective problem or concern”[3]. Code, in this sense, can be regulatory. AI regulating AI essentially means either (1) designing an AI system with compliance and/or debugging features, or (2) building a logically separate AI system to control and/or monitor other AI systems. In both cases, what is being managed is the risk of occasional or ongoing non-compliance with a set of rules including, but not necessarily limited to, legal rules. However, code will only be ‘public control’, which how Chesterman uses the term in his book, where it is authored or deployed by government.

Separating the concept of regulation from law can help clarify the relationship between the two issues raised in Chesterman’s article – AI in judicial decision-making and AI as regulator (particularly in regulating AI). AI avatar judges as imagined in China are not necessarily “regulation by AI” – the judge’s role in a particular case is not primarily to manage risk or behavior but to determine a matter in accordance with the law. There are nevertheless two ways of linking the idea of AI regulating AI with judicial decision-making, depending on whether the AI being regulated is inside or outside the courtroom.

In the one scenario, the legislature passes a law requiring that AI regulate AI in a particular industry. For example, they might require manufacturers to design cars that (1) are programmed to comply with road traffic laws, and/or (2) automatically produce audit logs (as in an aircraft black box) in the event of an incident that can be used to check whether the cause of the incident was related to non-compliance with road traffic laws or another bug in the software, and/or (3) are promptly updated in the event of identified bugs or non-compliance. It is irrelevant for current purposes whether this is part of the AI system operating the vehicle, or a logically independent system. This feeds into the court system in the event of an incident where liability by the vehicle manufacturer and/or driver is contested. For example, if the software was not programmed to ‘comply by design’, that could be evidence of a safety defect, resulting in liability of the manufacturer. Conversely, if audit logs demonstrate that the driver was not monitoring road and traffic conditions, that would suggest they might be liable to anyone injured. One could vary this scenario slightly and remove the legislature’s intervention – even without a law, the existence of ‘AI regulating AI’ under a voluntary self-regulatory regime would be relevant evidence in court. One could also vary it in the other direction – the legislature could adopt a ‘rules as code’ approach and write road traffic laws in the form of computer code to be directly incorporated into automated vehicles[4].

Alternatively, the idea of AI regulating AI could be applied directly to the use of AI in courtrooms. For example, a tool used by judges in assessing the risk of re-offending could be linked with an embedded or independent tool to ensure fairness. Fairness would have to be

1 Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York: Basic Books.

2 Chesterman, S. (2021). *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*. Cambridge: Cambridge University Press Available at: https://doi-org.wwwproxy1.library.unsw.edu.au/10.1017/9781009047081.

3 Black, J. (2014). Learning from Regulatory Disasters. *London School of Economics and Political Science Law, Society and Economy Working Papers 24/2014*. Available at: https://core.ac.uk/download/pdf/35434101.pdf; Yeung, K. (2017). “Are human biomedical interventions legitimate regulatory policy instruments?,” in *Oxford Handbook on the Law and Regulation of Technology*, eds. R. Brownsword, E. Scotford, and K. Yeung (Oxford: OUP).

4 See Bennett Moses, L., Boughey, J., and Crawford, L. B. (2021). “Laws for Machines and Machine-made Laws,” in *The Automated State: Implications, Challenges and Opportunites for Public Law*, eds. J. Boughey and K. Miller (Alexandria NSW: Federation Press).

defined[5], but either the first tool could be designed to limit itself to models that met the relevant fairness criteria or a second tool could monitor outputs of the first to measure compliance with fairness requirements over time. In this context, AI regulating AI is a mechanism for ensuring that legal decision-making is itself compliant with a rule, here a rule about fairness, which could be embedded in a legal requirement (say, in legislation) or simply employed by software companies or courts on their own initiative.

Only in the second scenario does AI regulating AI *directly* relate to the issues with which Chesterman's article begins. While the first scenario could be relevant to the question of whether AI judges are feasible, there are contingencies. If the law stated that manufacturers of automated vehicles were liable if and only if they did not comply with the hypothesized requirements, a computer system could be built to adjudicate whether a manufacturer in a particular case was liable. The manufacturer could submit a formal proof of the first requirement, the parts of the code related to data capture and audit, and the series of versions of the software alongside a comprehensive database of audit logs related to incidents. Unlike assessments of whether behavior is 'reasonable', evaluation of the manufacturer's submissions could be accomplished by a computer program. It is analogous to Chesterman's example of judges who only need determine 'whether a leather encased ball passed within a three-square-foot zone or not'. However, absent such a specific test for liability, the fact that AI is regulating AI does not necessarily link to questions around AI replacing judges, although it may produce data that can be introduced into evidence in particular proceedings.

In his article, Chesterman engages primarily with the question of AI in the context of judicial decision-making, commenting on the limitations of AI judges when considering the nature and role of law. His discussion of AI regulating AI as a potential 'special case' exception to the undesirability of 'regulation by AI' leads to questions about the relationship between 'regulation by AI' and 'AI regulating AI' on the one side, and the role of judges in our legal system on the other.

This Commentary sought to explore that relationship. When AI is used to regulate AI systems in the judicial system, it can operate to reduce *some* of the concerns raised about AI in the legal system. When AI is used to regulate AI systems in other contexts, at most it offers the possibility of turning issues raised in legal proceedings into the kind that can be resolved computationally. However, in neither case does AI regulating AI address Chesterman's ultimate concern with preserving values central to our legal and political systems.

Having commented briefly the issue that most intrigued me, it is worth mentioning the broad fields that the article opens up for further exploration. What mix of formal law and technological management (including 'AI regulating AI') is optimal, from the perspective of the rule of law and political control?[6] To what extent are concerns about AI judges linked to the limitations of technology, which may be overcome, and to what extent are they linked to more fundamental concerns about the appropriate use of technology?[7] Given developments in chatbot technology, will we even know when judges are assisted by AI tools? These are questions that will increasingly be asked, long before we are told to "Rise for the Honourable Robot Judge".

5 Verma, S., and Rubin, J. (2018). Fairness definitions explained. in *Proceedings of the International Workshop on Software Fairness* FairWare '18. (New York, NY, USA: Association for Computing Machinery), 1–7. doi: 10.1145/3194770.3194776.

6 See also Brownsword, R. (2019). *Law, Technology and Society: Reimagining the Regulatory Environment*. Routledge doi: 10.4324/9781351128186.

7 Deakin, S., and Markou, C. eds. (2020). *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*. Hart Publishing.

Copyright (c) 2023, Lyria Bennett Moses.

Creative Commons License
This work is licensed under a Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License.

Handel's Messiah and the Robotic Courts

Comment on 'All Rise for the Honourable Robot Judge?

Ugo Pagallo

1. Intruduction

There is an Anglo-Saxon tradition, according to which audiences stand up for the Hallelujah chorus towards the end of Handel's Messiah. An apocryphal story tells that the custom may have begun at the Messiah's 1743 premiere in London. Apparently, during the famous chorus, King George II rose from his seat. Some reckon that the British king simply thought that the long oratorio – which lasts circa 2 hours and 20 minutes – was finally luckily close to the end. The audience comprehensively followed suit, standing up to not offend the king. The anecdote fits like hand into glove the analysis of Simon Chesterman in *All Rise for the Honourable Robot Judge?* In both cases, i.e., Handel's Messiah and the robotic courts, people would stand up either due to a misunderstanding, or according to a custom born for the wrong reasons.

Chesterman's work represents a reference point for today's debate on the law and artificial intelligence (AI).[1] In his recent article on using AI to regulate AI,[2] Chesterman provides a brilliant analysis on efforts to automate the law, on how AI systems could be made to be self-policing, and why legal regulation by AI – even if it were possible at all – is not desirable. These arguments can be summed up in accordance with two main points. They regard the description of today's state-of-the-art vis-à-vis a normative stance to take sides in current efforts on the automation of the law.

2. Misunderstanding AI

Concerning the descriptive part of the analysis in *All Rise*, the attention should be drawn to the limits of technology. AI systems still fall short in dealing with both functions and requisites of the courts, generally speaking, the adjudicatory powers of the law. At their best possible light, efforts of engineers and computer scientists to set up AI systems for decision-making in the legal domain, aim to predict what human judges could rule in such cases either through neural networks or machine learning techniques. As Chesterman claims, "rather than being part of an ongoing social process in the development of the law, such determinations are more akin to forecasting the weather."[3]

The predictive functions of AI applications in the legal domain could captivate fans of Oliver Wendell Holmes and his jurisprudential philosophy of law, according to which the law is supposed to be a prediction of what courts will do in fact. As a matter of fact, however, the predictive powers of AI have their limits. Also, but not only in the legal domain, AI systems critically lack robustness. Even the most powerful AI models are sensitive to small changes and perturbations. Small changes in the input data can dramatically affect the output and lead to different results. This can have fatal consequences in, e.g., a criminal trial. It is noteworthy that the use of AI systems in the judicial domain is deemed as 'high-risk' in the 2021 proposal of the European Commission for a new *Artificial Intelligence Act*. The fact that China – as illustrated by Chesterman in his paper – is creating a sort of robotic courts through the use of AI-assisted adjudication systems can thus be interpreted either as intrepid legal experimentation, or as a straight way to suppress dissenting opinions and conflicting values.

Mind, the limits of technology and hence, of legal regulation of AI do not mean that we cannot automate several procedural steps of the law. This is what some colleagues in Bologna with my team in Turin are actually doing with the processing of asylum applications in Italian immigration law. For that matter, also Chesterman admits "that AI will continue to transform the legal profession and the role of lawyers."[4] However, the development of AI systems for information retrieval and support for decision-making poses some problems of its own. Indeed, it can be really tricky to determine to what extent the output of the algorithm leaves room to human autonomy. It is worth mentioning once again the AI Act of the European Commission and how the 'high-risk' use of AI in the judicial system is subordinated to a considerable burden of tests, authorizations, and obligations. Far from approving the administrative approach of the AI Act, what is relevant here to stress is the complex balance that shall be struck between human autonomy and legal automation. Judges have to have a meaningful control over their AI assistants, which means, on the one hand, that AI systems should respond to the reasons and intentions of their human controllers, and on the other hand, that such humans should be "cognitively, physically and morally capable to perform their assigned tasks and fulfill their obligations."[5]

The normative part of the analysis should be complemented with the breathtaking advancements of technology, e.g., Moore's law on computational power. In fact, what should our stance be if AI systems break down current limits of technology and become as good as – if not better than – current judges?

3. Tricky Robots that Shall Abide by the Law

The scenario of smart robots that shall abide by the law is not simply theoretical. The problem has been discussed in international humanitarian law (IHL) over the past 15 years. It is striking that, back in 2010, two special rapporteurs of the United Nations (UN) delivered quite different opinions in their reports to the UN General Assembly. According to Philip Alston, on the one hand, the use of smart robots on the battlefield does not entail any matter of principle: "a missile fired from a drone is no different from any other commonly used weapon, including a gun fired by a soldier or a helicopter or gunship that fires missiles. The critical legal question is the same for each weapon: whether its specific use complies with IHL."[6] On the other hand, according to Christof Heyns, it makes a lot of difference whether a missile is fired by a human soldier, or its robotic counterpart, therefore raising "the fundamental question of whether lethal force should ever be permitted to be fully automated."[7]

1 See S. Chesterman, *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*, Cambridge University Press, 2021.
2 S. Chesterman, *All Rise for the Honourable Robot Judge? Using Artificial Intelligence to Regulate AI* (October 19, 2022). NUS Law Working Paper No. 2022/019, at SSRN: https://ssrn.com/abstract=4252778 or http://dx.doi.org/10.2139/ssrn.4252778.
3 S. Chesterman, All Rise for the Honourable Robot, *supra* note 2, at 13.
4 S. Chesterman, All Rise for the Honourable Robot, *supra* note 2, at 14.
5 See F. S. Santoni de Sio, G. Mecacci, S. Calvert *et al.*, Realising Meaningful Human Control Over Automated Driving Systems: A Multidisciplinary Approach, *Minds & Machines*, 2002, available at https://doi.org/10.1007/s11023-022-09608-8.
6 See U. Pagallo, *The Laws of Robots: Crimes, Contract, and Torts*, Springer, Dordrecht 2013, at 59.
7 *Ibid*.

At the end of the day, this latter opinion prevailed. After a series of informal discussions at Certain Conventional Weapons (CCW) in Geneva between 2014 and 2016, a Group of Governmental Experts (GGE) was set up with a formal mandate at the Fifth Review Conference of the High Contracting Parties to CCW in December 2016. The aim was to establish limits to the use of emerging technologies in the area of lethal autonomous weapons systems. Although the GGE is far away from attaining any reasonable compromise, the basic principle of current negotiations seems pretty clear: humans shall stay in the loop.

This conclusion goes hand-in-hand with Chesterman's main thesis in *All Rise for the Honourable Robot Judge?* "Even if regulation by AI generally were possible, then, it is not desirable."[8] In other words, "red lines will need to be drawn and ultimate oversight conducted by politically legitimate and accountable actors."[9]

Chesterman's main thesis is more than acceptable; however, I have some problems with the conclusion of his analysis, in particular, the final sentence of the article: "Many decisions might therefore properly be handed over to the machines. But the final exercise of discretion" shall be upon us, the humans.[10] The assumption is problematic because it seems to suggest that the distinction between humans and machines is coextensive with the difference between non-automatic and automatic decisions. Since they are not, it is crucial to understand how this final part of the article may affect its main thesis. How should we grasp the interplay between human discretion and legal regulation by AI?

4. Human discretion meets legal automation

Humans do not simply decide as a result of meditation, criticism and a prudent evaluation of the circumstances. Human decisions are often the result of reiterated applications of already existing competences, patterns, heuristics. This sort of human automatism matters not only in the fields of cognitive psychology and behavioral economics, but it is relevant in the legal domain as well. In the words of Herbert Hart, there is indeed a whole set of legal cases where issues are "plain," that is, "where the general terms seem to need no interpretation and where the recognition of instances seems unproblematic or 'automatic'... where there is general agreement in judgements as to the applicability of the classifying terms."[11]

We can adopt in this context Hart's distinction of plain cases and hard cases, leaving aside his opinions on the rule of recognition, the role of natural law, or a more appropriate theory of legal positivism after Hans Kelsen's. The distinction between plain and hard cases is fruitful because it sheds light on the extent to which legal systems can go on delegating cognitive tasks and decisions to smart machines. AI systems can indeed be more efficient and more consistent than human judges and regulators in many fields of the legal domain. Most litigation in several countries that regard issues of tax law, banking law, immigration law, or traffic law, are amenable to automation. Moreover, most of these issues often end up with Hart's 'plain cases.'[12] Correspondingly, the problem does not consist either in delegating decisions to the machine, or in ensuring that the final decision will be taken by a human. The final exercise of discretion – e.g., admitting an asylum request – can fairly be delegated to the machine in all plain cases. The use of technological systems for decision-making that occurs in the plain cases of the law is thus opposed to what occurs in such fields as, e.g., soccer. We do not have to ask for technological support, e.g., the VAR in soccer, to determine if the ball has crossed the goal line. Just the opposite, we have the right to ask for human intervention during the automated process, e.g., the right enshrined in Art. 22 of the EU's general data protection regulation, any time we think something went wrong.

In addition to the plain cases of the law, there are of course the hard cases, namely, cases of disagreement that may depend on semantics, legal reasoning, or the role and logic of the principles governing the legal system. Accordingly, general disagreement may not only concern the interpretation of legal texts, but also, different values of the normative context under investigation. To prevent the suppression of these different values and multiple dissenting opinions, it seems fair to concede that the delegation of cognitive tasks and decision-making to AI systems shall be deemed as inadmissible vis-à-vis the hard cases of the law. Furthermore, it should always be up to humans determining whether we are dealing with a plain case, or a hard one. This assumption converges with Chesterman's claim that public control over legal processes shall remain a human affair. Whether dealing with hard cases or plain cases, the distinction rests on an institutional space of interpretation in which legal rules and principles are evaluated.[13]

This common institutional space does not entail that critical differences between plain and hard cases of the law – e.g., the use of AI systems for decision-making in such cases – simply collapse. To understand why such differences shall be kept firm, another kind of legal disagreement should be under scrutiny, namely, the meta-hard case of jurisprudence on how the law should address its own hard cases.

5. Dworkin's Paradox

Contrary to the ideas of Hart on the hard cases of the law and how to tackle them, I may concede that there could be room for a uniquely right answer, namely, the well-known thesis of Ronald Dworkin for the hard cases of the law. Against every general disagreement, jurists could identify the principles of the system that fit with the established law, to apply such principles in a way that interprets the case in the best possible light. As Dworkin claims, we "must read through what other judges in the past have written not only to discover what these judges have said, or their state of mind when they said it, but to reach an opinion about what these judges have collectively done, in the

8 S. Chesterman, All Rise for the Honourable Robot, *supra* note 2, at 4.
9 S. Chesterman, All Rise for the Honourable Robot, *supra* note 2, at 5.
10 S. Chesterman, All Rise for the Honourable Robot, *supra* note 2, at 20.
11 H.L.A. Hart, *The Concept of Law*, Oxford University Press, 1961, at p. 121.
12 See T. Agnoloni and U. Pagallo, The case law of the Italian constitutional court between network theory and philosophy of information. In: Winkels R, Lettieri N (eds) 2d International Workshop on Network Analysis in Law (with JURIX 2014), Krakow, December 2014, pp 26–38. http://www.leibnizcenter.org/~winkels/NAiL2014-pre-proceedings.pdf.; T. Agnoloni and U. Pagallo, The case law of the Italian constitutional court, its power laws, and the web of scholarly opinions. In: ICAIL'15 proceedings of the 15th international conference on artificial intelligence and law. ACM Digital Library, New York, pp 151–155. https://dl.acm.org/citation.cfm?id1/42746108; T. Agnoloni and U. Pagallo, The power laws of the Italian constitutional court, and their relevance for legal scholars. In: Rotolo A (ed) Legal knowledge and information systems - JURIX 2015: the twenty-eighth annual conference. IOS Press, Amsterdam, pp 1–10.
13 See U. Pagallo and M. Durante, The Pros and Cons of Legal Automation and its Governance, *European Journal of Risk Regulation*, 2016, 7(2): 323-334.

way that each of our novelists formed an opinion about the collective novel so far written."[14] As far as I know, nobody has developed a sort of Dworkinian judge supported by AI techniques, although the task seems not impossible. Yet, Chesterman reminds us of Dworkin's opinion that a Herculean robot would not be feasible. In my view, the difficulty of the task has to do more with the legal assumptions of Dworkin, than the limits of technological know-how.

It is my personal experience, collaborating over the past decades with such institutions as the European Commission, or the World Health Organization, that Hart was right when claiming that before the hard cases of the law, no right answer should be found, but rather, a reasonable compromise between many conflicting interests.[15] The disruption of autonomous lethal weapons in the field of IHL illustrates this point. In the phrasing of Hart, whereas the plain cases of the law refer to a general agreement as the condition for the existence and functioning of the law through standards of conduct – such as norms, values, and principles that need "no further direction" – the different kinds of disagreement making a legal case hard stress the relevant standards of conduct that can be adopted as the basis of legal decisions; and yet, they need a supplement of direction in terms of human intelligence.

Dealing with the hard cases of the law, the final exercise of discretion shall be up to human responsibility and accountability, not because a Herculean robot judge would be untenable, but rather, because that Herculean robot should compromise with us, the humans.

6. The Limits of Legal Automation

The more we insist on the hard cases of the law, the less we should overlook its plain cases. The exercise of human discretion in the legal domain is not necessarily a final safeguard against automation, but rather, a preliminary condition for automation and its expansion in the legal domain. The dream of who must be considered as the grandfather of AI & law, information technology law, and more – that is, the great German polymath Gottfried W. Leibniz – materialized 350 years after his PhD thesis on digital expert systems in the legal domain.[16]

As a courtier, diplomat, and wannabe politician, Leibniz was well aware that technological revolutions do not occur in a legal vacuum, but should be grasped in accordance with both the institutional and social dimensions of the law. The institutional dimension is stressed by Chesterman's concluding remarks on the distinction between a ballot box and the X-box. An institutional forum is indispensable to define the legal and ethical framework for any public deliberation on how law and technology must interplay. Leibniz conceded that legal reasoning often depends on presumptions, rather than proofs; on equity, rather than strict law; on governance and institutions, rather than the deontic logic of the commands of a sovereign.[17] Centuries later, this institutional framework shapes the different ways in which AI systems could be designed to be self-policing in accordance with the social dimension of the law. Whether the regulation of AI entails plain cases, or hard cases, it concerns also – but admittedly not only – the degree of social acceptability regarding the risk inherent in every automation process, and moreover, the degree of social cohesion that hinges on the values and principles that are stake with the delegation of cognitive tasks and decision-making to more or less smart machines.[18]

Such different levels of cohesion and social acceptability go hand-in-hand with different kinds of policies across multiple jurisdictions, traditions, and legal systems. We already noted that, in Europe, the legal regulation by AI is considered as a high-risk use of such AI systems, thus recommending a strict top-down regulatory set of commands. In Japan, the approach to the legal governance of AI rests on the soft law developed by Japanese ministries and governmental agencies, as much as the guidelines provided by the private sector, such as the ethical guidelines of the Japanese Society for AI (JSAI). In China, it is still an open question the degree of public acceptability and social cohesion for the push to automatize the entire court system. The same holds true in the U.S.A. as regards the traditional self-regulatory approach of that country and the first controversial applications of smart technologies in the judiciary.[19]

These discrepancies do not trigger any relativism. We noted that the limits of the state-of-the-art in technology recommend prudence and circumspection in the legal regulation of AI, and still, a well-established tradition in legal philosophy similarly warns against either a simple refusal of every kind of automation, or the abdication of human judges and regulators with their traditional powers.

The hard cases of the law posed by the legal regulation of AI need a public discussion and deliberation that shall address the interplay between law and technology. The question is not about how far the process of legal regulation of AI can go without the final exercise of discretion by humans. This trend on legal regulation of AI is coextensive with the set of plain cases that we find in every legal system. The limits of automation have thus to find elsewhere their regulatory boundaries. The contention of this article has been that the distinction between plain and hard cases cannot be subject to the legal regulation of AI. On the contrary, this distinction shall be reserved to humans that still bear full responsibility for the assessment of what is ethically, socially, and legally plain, or hard.

Every legal system that will do otherwise would incur in the same kind of alternative that audiences raise every time they stand up for the Hallelujah chorus. It is ironic that the tolerant Leibniz shared with a young Handel the troubles of the father of King George II, i.e., Georg I from Hannover. The alternative of people raising for the honourable robot is either misunderstandings on what AI systems can really do, or the ways in which AI systems could throw down roots for the wrong reasons. Leibniz warned against both threats.

Copyright (c) 2023, Ugo Pagallo.

Creative Commons License
This work is licensed under a Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License.

14 R. Dworkin, *A matter of principle*, Oxford University Press, 1985, at 159.
15 H.L.A. Hart, The Concept of Law, *supra* note 11, at 128.
16 See A. Artosi and G. Sartor, Leibniz as Jurist, in M.R. Antognazza (ed.), *The Oxford Handbook of Leibniz*, Oxford University Press, 2016.
17 See U. Pagallo, *Leibniz: Una breve biografia intellettuale, Kluwer, Milan (Italy), 2016.*
18 See U. Pagallo, Algo-Rhythms and the Beat of the Legal Drum, *Philosophy & Technology*, 2018, 31(4): 507-524.
19 See W. Barfield and U. Pagallo, *Advanced Introduction to the Law and Artificial Intelligence*, Elgar, Cheltenham, 2020, at 111.