1

From Rule of Law to Legal Singularity

SIMON DEAKIN AND CHRISTOPHER MARKOU*

> Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.
>
> **Bertrand Russell**
> *Recent Work on the Principles of Mathematics* (1901)

> In and of itself nothing really matters. What matters is that nothing is ever 'in and of itself'.
>
> **Chuck Klosterman**
> *Sex, Drugs, and Cocoa Puffs* (2003)

I. The Dawn of the All New Everything

Before most had a clue what the Fourth Industrial Revolution entailed,[1] the 2019 World Economic Forum meeting in Davos heralded the dawn of 'Society 5.0' in Japan.[2] Its goal: creating a 'human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space'.[3] Using Artificial Intelligence (AI) and various digital technologies, 'Society 5.0' proposes to liberate people from:

> … everyday cumbersome work and tasks that they are not particularly good at, and through the creation of new value, enable the provision of only those products and services that are needed to the people that need them at the time they are needed, thereby optimizing the entire social and organizational system.

* Faculty of Law and Centre for Business Research, University of Cambridge. We gratefully acknowledge the support of the Leverhulme Trust and of the ESRC through its funding for the Digital Futures at Work Research Centre (grant ES/S012532/1) and the UKRI-JST Joint Call on Artificial Intelligence and Society (grant ES/T006/315/1).

[1] K Schwab, *The Fourth Industrial Revolution* (Crown Publishing, 2017).

[2] H Nakanishi, 'Modern Society Has Reached its Limits. Society 5.0 Will Liberate Us' (World Economic Forum, 9 January 2019), www.weforum.org/agenda/2019/01/modern-society-has-reached-its-limits-society-5-0-will-liberate-us.

[3] Cabinet Office, Government of Japan, 'Society 5.0', www8.cao.go.jp/cstp/english/society5_0/index.html.

The Japanese government accepts that realising this vision 'will not be without its difficulties' but the plan makes clear its intention 'to face them head-on with the aim of being the first in the world as a country facing challenging issues to present a model future society'.

Although Society 5.0 enjoys support from beyond Japan,[4] it bears a familiarly Japanese optimism about the possibilities of technological progress.[5] Yet Japan is not alone in seeing how the technologies of the Fourth Industrial Revolution could enable new systems of governance and 'algorithmic regulation'.[6] And this is particularly the case with regard to a specific type of computation, a family of statistical techniques known as Machine Learning (ML),[7] that is central to engineering futures of vast technological possibility that Society 5.0 exemplifies.

Generally, ML 'involves developing algorithms through statistical analysis of large datasets of historical examples'.[8] The iterative adjustment of mathematical parameters and retention of data enable an algorithm to automatically update (or 'learn') through repeated exposure to data and to optimise performance at a task. Initially, the techniques were applied to the identification of material objects, as in the case of facial recognition. Successive breakthroughs and performance leaps in ML, and the related techniques of Deep Learning (DL),[9] have encouraged belief in AI as a universal solvent for complex socio-technical problems. Tantalising increases of speed and efficiency in decision-making, and reductions in cost and bureaucratic bloat, makes the public-sector fertile ground for a number of AI-leveraging 'Techs'. These include LegalTech,[10] GovTech,[11] and RegTech[12] (short for Legal, Government and Regulatory technology respectively) which involve the development of 'smart' software applications for deployment in legal, political, and human decision-making contexts.

The Society 5.0 plan is not, however, an ex nihilo creation of the Japanese government. Rather, it articulates an emerging orthodoxy – one the 'Techs' are now capitalising on – that the core social systems of law, politics and the economy must adapt or die in the face of new modes of 'essentially digital governance'. This is often idealised as leading to a 'hypothetical new state' with 'a small intelligent core, informed by big data … leading

[4] C Dube and M Minevich, 'Mapping AI Solutions in Japan's Society 5.0' (2018) IPSoft/AI Pioneers White Paper, sifted.eu/articles/europe-can-learn-from-japans-society-5-0/.

[5] J West, 'Utopianism and National Competitiveness in Technology Rhetoric: The Case of Japan's Information Infrastructure' (1995) 12 *The Information Society* 3.

[6] K Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2017) King's College London Dickson Poon School of Law Legal Studies Research Paper Series: Paper No. 2017-27.

[7] *cf* E Alpaydin, *Machine Learning* (MIT Press, 2016); G Marcus and E Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (Pantheon, 2019) 42–48; D Spiegelhalter, *The Art of Statistics: Learning from Data* (Pelican Books, 2019) 143–87.

[8] Spiegelhalter, *The Art of Statistics*, 144.

[9] I Goodfellow, Y Bengio and A Courville, *Deep Learning* (MIT Press, 2016) 1–8.

[10] R Dale, 'Industry Watch: Law and Word Order: NLP in Legal Tech' (2019) 25 *Natural Language Engineering* 1.

[11] IBM, 'Digital Transformation: Reinventing the Business of Government' (2018), www.techwire.net/sponsored/digital-transformation-reinventing-the-business-of-government.html.

[12] DW Arner, JN Barberis and RP Buckley, 'FinTech, RegTech and the Reconceptualization of Financial Regulation' (2016) University of Hong Kong Faculty of Law Research Paper No. 2016/035, ssrn.com/abstract=2847806; G Roberts, 'Fintech Spawns Regtech to Automate Compliance with Regulations' (Bloomberg, 28 June 2016), www.bloomberg.com/professional/blog/fintech-spawns-regtech-automate-compliance-regulations.

government (at last) to a truly post-bureaucratic "Information State"'.[13] Some, such as Tim O'Reilly, argue that the 'old' state model is essentially a 'vending machine' where money goes in (tax) and public goods and services come out (roads, police, hospitals, schools).[14] It is time, he and others suggest, to 'rethink government with AI'[15] now that technological change has 'flattened'[16] the world, eroded state power, and provided models for uncoupling citizenship from territory.[17] Only by seeing government as a 'platform' will it be possible to harness critical network externalities and ensure what Jonathan Zittrain calls 'generativity' – the uncanny ability of open-ended platforms like Facebook or YouTube to create possibilities beyond those envisioned by their creators.[18] For O'Reilly, Big Tech 'succeeded by changing all the rules, not by playing within the existing system'.[19] Governments around the world, he suggests, must now follow their lead. Evidence from Japan, Singapore, Estonia and elsewhere, indicates that many are.[20]

A shift towards increasingly 'smart' and data-driven government is now underway and shows no sign of abating.[21] But the intoxicating 'new government smell' and techno-utopian visions of programmes such as Society 5.0 should not distract from critical questions about what exactly 'techno-regulation' means for human rights, dignity and the role of human decision-makers in elaborate socio-technical systems that promise to more or less run themselves.[22] Frank Pasquale observes that societal 'authority is increasingly expressed algorithmically',[23] while John Danaher warns against the 'threat of algocracy' – arguing it is 'difficult to accommodate the threat of algocracy, i.e. to find some way for humans to "stay on the loop" and meaningfully participate in the decision-making process, whilst retaining the benefits of the algocratic systems'.[24] Both are key observations for Society 5.0 where 'people, things, and systems … [are] all connected in cyberspace and optimal results obtained by AI exceeding the capabilities of humans [are] fed back to physical space'. However, the idea of AI 'exceeding' human capabilities is where Society 5.0's vision comes into sharper focus. Looking past the aspirational rhetoric of a 'human-centred society' it is ultimately a future where Artificial General

[13] P Dunleavy and H Margetts, 'Design Principles for Essentially Digital Governance', 111th Annual Meeting of the American Political Science Association (2015) 26, eprints.lse.ac.uk/64125/.

[14] T O'Reilly, 'Government as a Platform' (2010) 6 *Innovations* 13.

[15] H Margetts and C Dorobantu, 'Rethink Government with AI' (*Nature*, 9 April 2019), www.nature.com/articles/d41586-019-01099-5.

[16] G Hadfield, *Rules for a Flat World: Why Humans Invented Law and How to Reinvent It For a Complex Global Economy* (Oxford University Press, 2016).

[17] M Loughlin, 'The Erosion of Sovereignty' (2016) 45 *Netherlands Journal of Legal Philosophy* 2.

[18] J Zittrain, 'The Generative Internet' (2006) 119 *Harvard Law Review* 1974.

[19] O'Reilly, 'Government as a Platform', 38.

[20] M Goede, 'E-Estonia: The e-government Cases of Estonia, Singapore, and Curaçao' (2019) 8 *Archives of Business Research* 2.

[21] Obama White House, 'Digital Government: Building a 21st Century Platform to Better Serve the American People' (2012), obamawhitehouse.archives.gov/sites/default/files/omb/egov/digital-government/digital-government.html; UK Cabinet Office, 'Government Transformation Strategy 2017 to 2020' (9 February 2017), www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020; Treasury Board of Canada Secretariat, 'Digital Operations Strategic Plan: 2018–2022' (27 December 2018), www.canada.ca/en/government/system/digital-government/digital-operations-strategic-plan-2018-2022.html.

[22] E Medina, 'Rethinking Algorithmic Regulation' (2015) 44 *Kybernetes* 6/7.

[23] F Pasquale, *The Black Box Society* (Harvard University Press, 2015) 8.

[24] J Danaher, 'The Threat of Algocracy: Reality, Resistance and Accommodation' (2016) 29(3) *Philosophy & Technology* 266.

Intelligence (AGI) is no longer hypothetical. The AGI hypothesis is that a machine can be designed to perform any 'general intelligent action'[25] that a human is capable of – an idea with longstanding institutional support in Japan.[26] But the invocation of AGI is what makes 'Society 5.0' hard to reconcile with what exactly it portends for the centrality and need of human decision-makers.[27]

II. From Rule of Law to Legal Singularity

While Society 5.0 perhaps exemplifies what Evgeny Morozov terms the folly of 'solutionism', it is not a uniquely Japanese phenomenon.[28] Indeed, such techno-solutionism has long been part of the 'dotcom neoliberalism' Richard Barbrook and Andy Cameron call 'The Californian Ideology'.[29] This ideology has, however, now crept into the rhetoric of LegalTech developers who have the data-intensive – and thus target-rich – environment of law in their sights. Buoyed by investment, promises of more efficient and cheaper everything, and claims of superior decision-making capabilities over human lawyers and judges, LegalTech is now being deputised to usher in a new era of 'smart' law built on AI and Big Data.[30] For some, such as physicist Max Tegmark, the use-case is clear:

> Since the legal process can be abstractly viewed as computation, inputting information about evidence and laws and outputting a decision, some scholars dream of fully automating it with robojudges: AI systems that tirelessly apply the same high legal standards to every judgment without succumbing to human errors such as bias, fatigue or lack of the latest knowledge.[31]

Others, such as Judge Richard Posner, are cautious but no less sympathetic to the idea:

> The judicial mentality would be of little interest if judges did nothing more than apply clear rules of law created by legislators, administrative agencies, the framers of constitutions, and other extrajudicial sources (including commercial custom) to facts that judges and juries determined without bias or preconceptions. Judges would be well on the road to being superseded by digitized artificial intelligence programs … I do not know why originalists and other legalists are not AI enthusiasts.[32]

Legal scholar Eugene Volokh even proposes a legal Turing test to determine whether an 'AI judge' outputs valid legal decisions. For Volokh, the persuasiveness of the output is what matters:

> If an entity performs medical diagnoses reliably enough, it's intelligent enough to be a good diagnostician, whether it is a human being or a computer. We might call it 'intelligent,' or we

[25] A Newell and HA Simon, 'Computer Science as Empirical Inquiry: Symbols and Search' (1976) 19(3) *Communications of the ACM* 116.

[26] EY Shapiro, 'The Fifth Generation Project – a Trip Report' (1983) 26(3) *Communications of the ACM* 637.

[27] DJ Chalmers, 'The Singularity: A Philosophical Analysis' (2010) 17 *Journal of Consciousness Studies* 7; N Bostrom, *Superintelligence: Paths, Dangers, and Strategies* (Oxford University Press, 2014).

[28] E Morozov, *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist* (Penguin, 2014).

[29] R Barbrook and A Cameron, 'The Californian Ideology' (1996) 6 *Science as Culture* 1.

[30] MB Chalmers, 'SmartLaw 2.0: The New Future of Law' (*Lexology*, 22 August 2018), www.lexology.com/library/detail.aspx?g=c7605940-ef62-4c49-9451-3799083beb60.

[31] M Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Allen Lane, 2017) 105.

[32] RA Posner, *How Judges Think* (Harvard University Press, 2008) 5.

> might not. But, one way or the other, we should use it. Likewise, if an entity writes judicial opinions well enough … it's intelligent enough to be a good AI judge … If a system reliably yields opinions that we view as sound, we should accept it, without insisting on some predetermined structure for how the opinions are produced.[33]

The import of these views is that human judges are not just replaceable with AI, but that 'AI Judges' should be preferred on the assumption that they will not inherit the biases and limitations of human decision-making.[34] Nonetheless, other scholars, such Giovanni Sartor and Karl Branting, remain sceptical:

> No simple rule-chaining or pattern matching algorithm can accurately model judicial decision-making because the judiciary has the task of producing reasonable and acceptable solutions in exactly those cases in which the facts, the rules, or how they fit together are controversial.[35]

The boldest vision, however, comes from legal scholar and LegalTech entrepreneur Ben Alarie:

> Despite general uncertainty about the specifics of the path ahead for the law and legal institutions and what might be required of our machines to make important contributions to the law, over the course of this century we can be confident that technological development will lead to (1) a significantly greater quantification of observable phenomena in the world ('more data'); and (2) more accurate pattern recognition using new technologies and methods ('better inference'). In this contribution, I argue that the naysayers will continue to be correct until they are, inevitably, demonstrated empirically to be incorrect. The culmination of these trends will be what I shall term the 'legal singularity'.[36]

Although AGI is not seen as a necessary condition, Alarie's legal singularity is described as a point where AI has ushered in a legal system 'beyond the complete understanding of any person.'[37] Seemingly a response to the incompleteness and contingency of the law, the legal singularity is implicitly a proposal for eliminating juridical reasoning as a basis for dispute resolution and normative decision-making. While nothing is said about the role of law in Society 5.0, much less human lawyers and judges, Alarie's legal singularity can be considered a credible interpolation.

Even if the mathematical or symbolic logic used in AI research could, at least in theory, replicate the structure of juridical reasoning, this would not necessarily account for the political, economic, and socio-cultural factors that influence legal discourse and the evolution of the legal system.[38] Jürgen Habermas argues that these factors are important because:

> The positivist thesis of unified science, which assimilates all the sciences to a natural-scientific model, fails because of the intimate relationship between the social sciences and

[33] E Volokh, 'Chief Justice Robots' (2019) 68 *Duke Law Journal* 1135, 1138.

[34] EM Harley, 'Hindsight Bias in Legal Decision Making' (2007) 25 *Social Cognition: Special Issue: The Hindsight Bias* 1; KL Mosier and LJ Skitka, 'Human Decision Makers and Automated Decision Aids: Made for Each Other?' in R Parasuraman and M Mouloua (eds), *Automation and Human Performance: Theory and Applications* (CRC Press, 2009).

[35] G Sartor and LK Branting, 'Introduction: Judicial Applications of Artificial Intelligence' in G Sartor and LK Branting (eds), *Judicial Applications of Artificial Intelligence* (Springer, 1998) 1.

[36] Alarie, 'The Path of the Law: Towards Legal Singularity' (2016) 66 University of Toronto LJ 443, 446.

[37] Alarie, 'The Path of the Law', 455.

[38] P Stein, *Legal Evolution: The Story of an Idea* (Cambridge University Press, 2009); S Deakin, 'Legal Evolution: Integrating Economic and Systemic Approaches' (2011) 7 *Review of Law and Economics* 659.

> history, and the fact that they are based on a situation-specific understanding of meaning that can be explicated only hermeneutically … access to a symbolically prestructured reality cannot be gained by observation alone.[39]

But if mathematical logic cannot capture the 'situation-specific understanding' of legal reasoning and the complexity of the social world it exists in – at least to any extent congruent with how natural language categories cognise social referents and character of meaning – the hypothetical totalisation of 'AI judges' implied by the legal singularity would instantiate a particular view of law: one in which legal judgments are essentially deterministic or probabilistic outputs, produced on the basis of static or unambiguous legal rules, in a societal vacuum. This would deny, or see as irrelevant, competing conceptions of law, in particular the idea that law is a social institution, involving socially constructed activities, relationships, and norms not easily translated into numerical functions.[40] It would also turn a blind eye to the reality that legal decision making involves an exercise of power which is both material and, in Pierre Bourdieu's sense, 'symbolic'.[41] The recursive exercise of this power is only legitimate if the process through which each exercise of it adheres to prevailing procedural expectations that are highly contingent and socially constructed.

As HLA Hart observed, it is critical to 'preserve the sense that the certification of something as legally valid is not conclusive of the question of obedience, and that, however great the aura of majesty or authority which the official system may have, its demands must in the end be submitted to a moral scrutiny'.[42] In short, the legal system as a totality, and its outputs, including legal judgments, should be subject to checks and balances, or 'feedback' to use an algorithmic term, to determine the 'legal validity' of its internal operations and the process of outputting and enforcing them with a society. Thus one particular danger of 'AI judges' is that moral scrutiny would not be a meaningful prophylactic again challenging abuses of power.

This is particularly the case with so-called 'black box' algorithms which are inscrutable at a technical level, or where trade secret-protections prevent diagnostic scrutiny.[43] However, at another level of abstraction, objections to ADM and 'AI Judges' before they have been adopted and normalised are also a way of subjecting the legal system to the moral scrutiny Hart though to be critical. While this perspective does not deny that legal decision-making is some sense structurally algorithmic in nature – insofar as it involves following defined steps to reach a particular output, it is wrong, as Mireille Hildebrandt observes:

> … to mistake the mathematical simulation of legal judgment for legal judgment itself. Whereas machines may become very good in such simulation, judgment itself is predicated

[39] J Habermas, *On the Logic of the Social Sciences* (1967) cited in W Outhwaite, *Habermas: Key Contemporary Thinkers*, 2nd edn (Polity Press, 2009) 22.

[40] H Ross, *Law as a Social Institution* (Hart, 2001).

[41] P Bourdieu, *Language and Symbolic Power* (Harvard University Press, 1991).

[42] HLA Hart, *The Concept of Law*, 3rd edn (Oxford University Press, 2012) 210.

[43] *State of Wisconsin v Loomis* 881 N.W.2d 749 (Wis. 2016); *cf* 'State v. Loomis: Wisconsin Supreme Court Requires Warning Before Using Risk Assessments in Sentencing' (2017) 130 *Harvard Law Review* 1530, harvardlawreview.org/2017/03/state-v-loomis; H-W Liu, C Fu and Y-J Chen, 'Beyond *State v Loomis*: Artificial Intelligence, Government Algorithmization and Accountability' (2019) 27 *International Journal of Law and Information Technology* 2; R Yu and RS Alo, 'What's Inside the Black Box? AI Challenges for Lawyers and Researchers' (2019) 19 *Legal Information Management* 1.

> on the contestability of any specific interpretation of legal certainty in the light of the integrity of the legal system – which goes way beyond a quasi-mathematical consistency.[44]

Neglecting the often subtle ways in which power and legitimacy are transferred and social order maintained within and across generations risks undermining one the principal institutions of a liberal-democratic order through the 'black boxing' of the legal system. As such the idea of 'AI judges' must be treated with utmost caution and scepticism, particularly if their outputs continue to defy both moral and technical scrutiny.[45]

III. The Origins of Digital Computation

Mathematical and statistical calculation have long been recognised as the best methods for inferring from the present what might happen in the future.[46] While AI and various statistical techniques are making prediction, to some degree at least, increasingly feasible, the notion of making law more predictable is intertwined with the development of computers more generally. But what are the basic properties of this computational universe created by digital computers?

At its most fundamental the computational universe is composed of bits, the basic unit of information theory and digital communications. The word 'bit' – a contraction of 'binary digit' – was coined by the American mathematician John Tukey.[47] Whether it is a universe of five kilobytes or zettabytes, there are only two types of bits that make a difference in it: differences in space and differences in time. A digital computer selects between these two forms of information by parsing their structure and sequence according to hard-coded rules. Bits that are embodied as structure vary in space, but do not vary over time, are the basis of computational memory. Whereas bits embodied in sequence, which vary over time and remain invariant over space, are the basis of what we call computer code.

The identification of a fundamental unit of communication, represented by a single distinction between binary alternatives (0/1, off/on) was the central idea of information theorist Claude Shannon's then-secret 'Mathematical Theory of Cryptography' in 1945,[48] which he expanded into the seminal 'Mathematical Theory of Communication'

[44] M Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2017), dx.doi.org/10.2139/ssrn.2983045.

[45] M Annany and K Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability' (2018) 20(3) *New Media & Society* 973; J Pearl and D MacKenzie, *The Book of Why: The New Science of Cause and Effect* (Penguin, 2018) 358.

[46] Spiegelhalter, *The Art of Statistics*, 143–87.

[47] CE Shannon, 'A Mathematical Theory of Communication' (1948) 47 *The Bell System Technical Journal* 3 ('The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly bits, a word suggested by J. W. Tukey').

[48] CE Shannon, 'A Mathematical Theory of Cryptography – Case 20878' (1945) MM-45-110-92, www.iacr.org/museum/shannon/shannon45.pdf.

in 1948.[49] The cybernetician Gregory Bateson would subsequently reformulate Shannon's theory into the open-ended phrase 'any difference that makes a difference'.[50] Within the digital universe of computation, however, the only difference that matters is that between zero and one.

The idea that two symbols were enough to encode all communications was not a new one. Rather, it was established by Francis Bacon (1561–1626). In his 1623 tract *De Augmentis Scientarium*, Bacon reasoned:

> The transposition of two Letters by five placeings will be sufficient for 32 Differences [and] by this Art a way is opened, whereby a man may expresse and signifie the intentions of his minde, at any distance of place, by objects … capable of a twofold difference only. [sic]

He went on to provide examples of how this binary coding could be conveyed at the speed of paper, the speed of sound, or the speed of light.[51] Bacon's insight, that zero and one were sufficient for arithmetic, would be expanded upon by Thomas Hobbes (1588–1679), most notably in his *De Corpore* (1655) and *Computation (or 'Logique')* in 1656. It is in this earlier work that Hobbes first explains his idea that human reasoning is a form of computation. 'By reasoning' he states, 'I understand computation':

> And to compute is to collect the sum of many things added together at the same time, or to know the remainder when one thing has been taken from another. To reason therefore is the same as to add or to subtract.[52]

In the next section, Hobbes offers some examples of how addition works in human reasoning, concluding that adding ideas together allows for the formulation of more complex ones. Accordingly, in Hobbes' view, it is 'from the conceptions of a quadrilateral figure, an equilateral figure, and a rectangular figure the conception of a square is composed'.[53] Though he concedes this accounts for only a fraction of human reasoning, he goes on to describe propositions and syllogisms in terms of addition:

> A syllogism is nothing other than a collection of a sum which is made from two propositions (through a common term which is called a middle term) conjoined to one another; and thus a syllogism is an addition of three names, just as a proposition is of two.[54]

In this way, Hobbes can be seen as one of the progenitors of the Computational Theory of Mind (CTM). The central idea of CTM is that the human mind is a form of computer. As Fodor puts it: 'the immediately implementing mechanisms for intentional laws are computational … [Computations] viewed in intension, are mappings from symbols under syntactic description to symbols under syntactic description'.[55] While there is a degree of concordance between Hobbes' work and subsequent theories of mind, the claim that Hobbes was 'prophetically launching Artificial Intelligence' seems a stretch.[56]

[49] Shannon, 'A Mathematical Theory of Communication'.

[50] GA Bateson, 'Re-examination of 'Bateson's Rule'' (1971) 60 *Journal of Genetics* 230; *cf* MH Schroeder, 'The Difference that Makes a Difference for the Conceptualization of Information' (2017) 1 *Proceedings* 221.

[51] F Bacon, *De Augmentis Scientiarum* (1623) translated by G Wats as 'Of the Advancement and Proficiency of Learning, or the Partitions of Sciences' (1640) 265–66.

[52] T Hobbes, *De Corpore in Part I of De Corpore* AP Martinich (trans) (Abaris Books, 1981) 1.2.

[53] Hobbes, *De Corpore*, 1.3.

[54] Hobbes, *De Corpore*, 4.6.

[55] JA Fodor, *The Elm and the Expert* (MIT Press, 1994) 8.

[56] B Haugeland, *Artificial Intelligence; The Very Idea* (MIT Press, 1985) 23.

IV. The Leibniz Dream and Mathematisation of Law

It is much less a stretch, however, to attribute the conceptual organs of Artificial Intelligence to German polymath Gottfried Wilhelm Leibniz (1646–1716). Nonetheless, Hobbes' influence of Leibniz's thinking is made clear by Leibniz himself:

> Thomas Hobbes, everywhere a profound examiner of principles, rightly stated that everything done by our mind is a computation, by which is to be understood either the addition of a sum or the subtraction of a difference … So just as there are two primary signs of algebra and analytics, + and −, in the same way there are as it were two copulas, 'is' and 'is not'.[57]

Following Hobbes, and in advance of German mathematician David Hilbert,[58] Leibniz believed that it was possible to develop a consistent system of logic, language, and mathematics using an alphabet of unambiguous symbols that could be manipulated according to mechanical rules. In a 1675 letter to the secretary of the Royal Society and his middle-man for correspondence with Isaac Newton, Leibniz wrote that 'the time will come, and come soon, in which we shall have a knowledge of God and mind that is not less certain than that of figures and numbers, and in which the invention of machines will be no more difficult than the construction of problems in geometry'.[59] Foreshadowing what we now refer to as software, Leibniz saw a bi-directional connection between logic and mechanism. In a letter sent to Dutch mathematician Christiaan Huygens in 1679, Leibniz appended the observation that 'one could carry out the description of a machine, no matter how complicated, in characters which would be merely the letters of the alphabet, and so provide the mind with a method of knowing the machine and all its parts'.[60]

Dissatisfied with the laborious arithmetic enabled by the common decimal system, Leibniz declared '[i]t is unworthy of excellent men to lose hours like slaves in the labour of calculation which could safely be relegated to anyone else if machines were used'.[61] To alleviate the burden, Leibniz proposed to 'develop a generalized symbolic language, and an algebra to go with it, so that the truth of any proposition in any field of human inquiry could be determined by simple calculation'.[62] In his *Art of Discovery* (1685) he thus asserted:

> The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate, without further ado, to see who is right.[63]

[57] GW Leibniz, 'Of the Art of Combination' in GHR Parkinson (ed), *Leibniz: Logical Papers* (Clarendon Press, 1966) 3.

[58] D Hilbert, 'Die Grundlegung der elementaren Zahlenlehre' 104 *Mathematische Annalen* 485; translated by W Ewald as 'The Grounding of Elementary Number Theory' in P Manscou (ed), *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s* (Oxford University Press, 1998) 266–73.

[59] 'GW Leibniz to Henry Oldenburg, December 18, 1675' in HW Turnbull (ed), *The Correspondence of Isaac Newton, Vol. 1* (Cambridge University Press, 1959) 401.

[60] 'GW Leibniz – Supplement to a letter to Christiaan Huygens, September 8, 1679' in LE Loemeker (ed), *Philosophical Papers and Letters, Vol. 1* (University of Chicago Press, 1956) 384–85.

[61] G Leibniz quoted in M Nadin, 'Predictive and Anticipatory Computing' in PA Laplante (ed), *Encyclopaedia of Computer Science and Technology: Volume 2* (CRC Press, 2017).

[62] FC Kreiling, 'Leibniz' (1968) 18(5) *Scientific American* 95.

[63] GW Leibniz quoted in M Gelford and Y Kahl, *Knowledge Representation, Reasoning, and the Design of Intelligent Agents* (Cambridge University Press, 2014) 7.

Using his logical calculus, Leibniz embarked on a major project to develop his vision of a 'universal symbolistic in which all truths of reason would be reduced to a kind of calculus'. Central to his project was the idea that: 'a kind of alphabet of human thoughts can be worked out and that everything can be discovered and judged by a comparison of the letters of this alphabet and an analysis of the words made from them'.[64] Leibniz worked out a system of universal coding in which primary concepts could be represented by prime numbers, thus providing a comprehensive framework for mapping numbers to ideas. Having done the groundwork himself, Leibniz thought a complete networking of numbers to ideas was not only feasible, but that 'a few selected men could finish the matter in five years … It would take them only two, however, to work out, by an infallible calculus, the doctrines most useful for life, that is, those of morality and metaphysics'. This was so that: 'the human race will have a new kind of instrument which will increase the power of the mind much more than optical lenses strengthen the eyes and which will be as far superior to microscope or telescopes as reason is superior to sight'.[65]

Although Leibniz believed binary coding to be the basis for a universal language,[66] he credited its creation to the Chinese. Specifically, it was in the hexagrams of the Yi-Jing – an ancient Chinese book of philosophy compiled in the latter part of the ninth century BCE – that Leibniz saw elements of 'a Binary Arithmetic … which I have rediscovered some thousands of years later'. Although Leibniz's ontology differed between writings,[67] some suggest his Dream was premised on a misunderstanding of the Yi-Jing's ontological reality.[68] In short, the Yi-jing treats reality as not entirely real, but more akin to a dream or illusion.[69] The 'dream' of reality humans experience is said to emerge from the binary oppositions of Yin and Yang as they play out in their infinite combinations. Smith notes that '[t]he binary structure of the *Yijing* entranced and inspired Leibniz [although] the number symbolism of the *Yijing* remained numerological and … never truly mathematical'.[70] As a consequence, Leibniz's system integrated a dualist view of reality in which everything could be represented and understood with 1s and 0s, or: Yin and Yang. While binary code was later refined by English mathematician George Boole, from whom 'Boolean logic' derives,[71] modern computation inherited Leibniz's dualist ontology.[72]

[64] GW Leibniz, 'On the General Characteristic' in LE Loemker (ed), *GW Leibniz: Philosophical Papers and Letters* (Kluwer, 1989) 224.

[65] GW Leibniz, 'ca. 1679' in LE Loemker (ed), *GW Leibniz: Philosophical Papers and Letters* (Kluwer, 1989) 344.

[66] G Leibniz, 'Explanation of Binary Arithmetic', www.leibniz-translations.com/binary.htm.

[67] T Tho, 'What is Leibniz's Ontology? Rethinking the Role of Hylomorphism in Leibniz's Metaphysical Development' (2015) 4(1) *Journal of Early Modern Studies* 79; J Whipple, 'Leibniz on Fundamental Ontology: Idealism and Pedagogical Exoteric Writing' (2017) 4(11) *Ergo* 311.

[68] ES Nelson, 'The *Yijing* and Philosophy: From Leibniz to Derrida' (2011) 38(3) *Journal of Chinese Philosophy* 377.

[69] C-Y Cheng, 'The Yi-Jing and Yin-Yang Way of Thinking' in B Mou (ed), *The Routledge History of Chinese Philosophy* (Routledge, 2008).

[70] RJ Smith, *Fortune-Tellers and Philosophers: Divination in Traditional Chinese Society* (Westview Press, 1991) 205.

[71] G Boole, *The Mathematical Analysis of Logic* (MacMillan, Barclay, & MacMillan, 1847).

[72] J Teixeira, 'Computational Complexity and Philosophical Dualism' (1998), www.bu.edu/wcp/Papers/Cogn/CognTeix.htm; D King, 'Cartesian Dualism, and the Universe as Turing Machine' (2003) 47(2) *Philosophy Today* 379.

Leibniz's surviving notes show his iterative development of simple algorithms for selecting between decimal and binary notation, as well as those for performing basic arithmetic function using strings of 1s and 0s. 'In binary arithmetic' he observed, 'there are only two signs, 0 and 1, with which we can write all numbers'.[73] With this system and a mechanical calculating machine he termed the 'Stepped Reckoner', Leibniz became convinced that formalising human thought with logico-mathematical calculations was not only possible, but would introduce mathematical rigour and precision into all the human sciences. This idea, commonly referred to as the Leibniz Dream (*characteristica universalis*), was an important precursor to the development of computer science and foreshadowed subsequent research into cognitive enhancement and extension.[74]

But his misunderstanding also influenced Leibniz's thinking about something of particular interest to him: law.[75] For Leibniz, law exemplified how society should resolve 'the most serious deliberations on life and health, on the state, on war and peace, on the moderation of conscience, [and] on care of eternity'.[76] He praised law as the most advanced instrument of human rationality, particularly in the 'balance' of reasons or 'logometric scales' judges used to evaluate the relative weight of 'arguments of discussants, opinions of authors, [and] voices of advisors'.[77] For law to be as precise as mathematics, a just legal rule was one that merged abstract legal axioms, such as the principle that harm should not be done to others, with empirical insights from the natural sciences. In his *Doctrina Conditionum* (1669) he tested this hypothesis with Roman law.[78] While his ambitions were certainly grand, Leibniz did not look to impose a monolithic 'scientific model' on law. Rather, he believed that the axiomatic method would help make it more precise, and ideally, predictable.

It was in 1679, however, that these ideas began to coalesce in Leibniz's imagination into what we might now call a digital computer in which binary numbers could be represented by marbles and governed by mechanical gates. 'This [binary] calculus' Leibniz wrote:

> … could be implemented by a machine (without wheels) in the following manner, easily to be sure and without effort. A container shall be provided with holes in such a way that they can be opened and closed. They are to be open at those places that correspond to a 1 and remain closed at those that correspond to a 0. Through the opened gates small cubes or marbles are to fall into tracks, through the others nothing. It [the gate array] is to be shifted from column to column as required.[79]

[73] GW Leibniz, 'Discourse on the Natural Theology of the Chinese' translated from 'Lettre sur la philosophie chinoise à Nicolas de Remond' H Rosemont and DJ Crook (trans and eds), *Monographs of the Society for Asian and Comparative Philosophy, No. 4* (University of Hawaii Press, 1977) 158.

[74] S Toulmin, 'The Dream of an Exact Language' in B Göranzon and M Florin (eds), *Dialogue and Technology: Art and Language* (Springer, 1991); *cf* N Bostrom and A Sandberg, 'Cognitive Enhancement: Methods, Ethics, Regulatory Challenges' (2009) 5(3) *Science and Engineering* 311; M Dresler, A Sandberg, C Bublitz, et al., 'Hacking the Brain: Dimensions of Cognitive Enhancement' (2019) 10(3) *ACS Chemical Neuroscience* 1137.

[75] *cf* M Armgardt, 'Leibniz as Legal Scholar' (2014) 20 *Fundamina* 1.

[76] GW Leibniz quoted in 'Introductory Essay' in M Dascal (ed), *GW Leibniz: The Art of Controversies* (Springer, 2006) 38.

[77] GW Leibniz quoted in *GW Leibniz: The Art of Controversies* 36.

[78] P Boucher, 'Leibniz: What Kind of Legal Rationalism?' in M Dascal (ed), *Leibniz: What Kind of Rationalist?* (Springer, 2008).

[79] GW Leibniz, 'De Progressione Dyadica – Pars I (15 March 1679)' in E Hochstetter and H-J Greve (eds), *Herrn von Leibniz' Rechnung mit Null und Einz* (Siemens Aktiengesellschaft, 1966).

While he did not term it as such, Leibniz had basically invented the shift register some 270 years before it was implemented in the Colossus Mark 2 – a British code breaking computer developed in 1944 – and used by the Allies in the Normandy invasions.[80] In the shift registers at the heart of all modern computers, voltage gradients and electron pulses have replaced the marbles and gravity of his original plans, but for all practical purposes modern computers still function largely identically to how Leibniz envisioned them in 1679.

V. *Calculemus!* Leibniz's Influence on Law

Posthumous English translations of Leibniz's work brought it to the attention of the English-speaking world. It particularly appealed to legal scholars, such as the formalist Christopher Columbus Langdell, who presented an axiomatic conception of law in his *Cases on Contracts* (1871).[81] Echoing many of Leibniz's claims, Langdell argued that 'Law, considered as a science, consisted of certain principles or doctrines. To have such mastery of these as to be able to apply them with constant facility and certainty to the ever-tangled skein of human affairs, is what constitutes a true lawyer ….'[82] They differed, however, on the source of legal axioms. Whereas Leibniz considered a combination of existing domestic and natural law as their source, Langdell believed that legal axioms – or to use his term: 'principles' – were empirically derived as they were in the natural sciences. According to Langdell the data for an empirical 'legal science' was contained in the cases decided and reported by courts.[83] This case 'data', he believed, would allow for predicting the outcomes of contract disputes – his primary focus.[84]

Despite vehement disagreement with both Leibniz and Langdell, American jurist Oliver Wendell Holmes was an avid admirer of the German polymath. As a lecturer at Harvard Law School, where Langdell was contemporaneously dean, Holmes spent his early career rejecting Leibnizian-Langdellian axiomatics, and instigating a turn in jurisprudence wary of any formalism or 'logic' in law. In *The Common Law* (1881) Holmes famously proclaimed that 'The life of the law has not been logic: it has been experience.'[85] In his subsequent article 'The Path of the Law' (1897), shrewdly published just after Langdell resigned his deanship, Holmes denounced legal axiomatics as 'a fallacy which I think it important to expose'.[86]

It was in this later article that Holmes articulated his 'Prediction Theory of Law'[87] central to which was the rhetorical 'bad man', a person devoid of ethics or lofty notions

[80] J Copeland and others (eds), *Colossus: The Secrets of Bletchley Park's Codebreaking Computers* (Oxford University Press, 2006) 74–77, 100.

[81] CP Wells, 'Langdell and the Invention of Legal Doctrine' (2010) 28(3) *Buffalo Law Review* 551.

[82] CC Langdell, *Cases on Contracts* (Little Brown & Co., 1871) vi.

[83] MH Hoeflich, 'Law & Geometry: Legal Science from Leibniz to Langdell' (1986) 30(2) *The American Journal of Legal History*.

[84] CP Wells, 'Langdell and the Invention of Legal Doctrine' 599–605.

[85] OW Holmes, *The Common Law* (Little Brown & Co., 1881) 1.

[86] OW Holmes, 'The Path of the Law' (1897) 110(5) *Harvard Law Review* 997.

[87] MH Fisch, 'Justice Holmes, the Prediction Theory of Law, and Pragmatism' (1942) 39(4) *The Journal of Philosophy* 85; RA Posner, 'The Path Away from the Law' (1996) 110 *Harvard Law Review* 1039.

about the jurisprudential role of courts and concerned only with avoiding payment of damages and staying out of jail. Holmes' enduring contribution to jurisprudence was rejecting the deterministic and nonpolitical conceptions of law advanced by legal formalists like Langdell who regarded law a consistent set of rules and norms from which a 'right' answer could always be derived. As Scott Brewer observes, '… the theorist who aspires to high-density axiomatization, as did Leibniz and Langdell, would seem to have to try and steer between the Scylla of generating axioms that are so open-textured and often so vague that one could never settle on the final set of axioms, but would forever have to be revising them'.[88]

Holmes' rejection of formalism provided the ideological foundation for the American Legal Realism school's critique of legal axiomatics.[89] Roscoe Pound, for instance, derided 'mechanical jurisprudence'[90] as the lazy practice of judges to formulaically apply precedents to cases with reckless disregard for consequences. For Pound, the logic of precedents could not solve jurisprudential problems; he warned that axiomatics risked ossifying socially constructed and politically influenced legal concepts into self-evident truths. It was precisely the desire to avoid such ossification – and not let the 'bad man' win – that led Holmes and the realists to conceive law as inherently indeterminate and defying understanding as a coherent or complete system of rules and principles from which there was always a 'right' answer.[91] In 'Logical Method and Law' (1924), John Dewey expanded Holmes' argument that 'general propositions do not decide concrete cases' – a position known as 'rule skepticism'. For Dewey, no legal argument could be validly represented as a deductive inference, and law never accurately represented with deductive axiomatic systems, arguing:

> There is of course every reason why rules of law should be as regular and as definite as possible. But the amount and kind of antecedent assurance which is actually attainable is a matter of fact, not of form. It is large wherever social conditions are pretty uniform, and when industry, commerce, transportation, etc., move in the channels of old customs. It is much less wherever invention is active and when new devices in business and communication bring about new forms of human relationship.[92]

Generally, the realists regarded judicial opinions and judges with scepticism and condescension. Because written rules (statutes, cases) did not determine what the law was, the purpose of juridical reasoning was not explaining how the court arrived at a decision and providing guidance to judges and litigants encountering similar situations in the future. The real purpose was to 'rationalise' and 'legitimise' decisions and conceal from the public and other judges the real, and often unsavoury, justifications. As such,

[88] S Brewer, 'Law, Logic, and Leibniz: A Contemporary Perspective' in A Artosi, B Pieri and G Sartor (eds), *Leibniz: Logico-Philosophical Puzzles in the Law* (Springer, 2013).

[89] SR Ratner, 'Legal Realism School' in R Wolfrum (ed), *Max Planck Encyclopedia of Public International Law Online* (Oxford University Press, 2007).

[90] R Pound, 'Mechanical Jurisprudence' (1908) 8 *Columbia Law Review* 605.

[91] C Gray, *The Nature and Sources of Law* (Columbia University Press, 1909); R Pound, 'Justice According to Law' (1914) 1(3) *The Midwest Quarterly* 223; KN Llewellyn, *The Bramble Bush: The Classic Lectures on the Law and Law School* (Oxford University Press, [1930] 2008).

[92] J Dewey, 'Logical Method and Law' (1924) 10(1) *Cornell Law Review* 25.

no deductive axiomatic method could account for these influences, and the gradual incorporation of social-scientific insights into legal scholarship helped substantiate the realists' scepticism.

VI. *Characteristica Universalis Lex*

But this was far from the last word on the idea of law as axiom. In the twentieth century the Leibniz Dream would inspire Alfred Tarski[93] and others, such as AI pioneer John McCarthy, to investigate whether the axiomatic method could be applied beyond mathematics. Throughout the 1950s McCarthy developed a methodology within the then emerging field of Artificial Intelligence (AI) known as the logic-based approach.[94] McCarthy's idea was to have a computer program capture and store large amounts of 'human knowledge' about a specific domain (such as law or medicine) and use mathematical logic to represent it and logical inference to determine the 'best' actions to achieve a desired 'output' from a knowledge-base.[95] Since McCarthy formalised the logic-based approach, AI research has remained oriented by the long term goal of AGI or 'Strong AI'.[96] In contrast, 'Weak AI' involves machines performing specific problem-solving or 'reasoning' tasks in 'narrow' domains not requiring wider facets of human intelligence.

An initially promising application of logical-AI was the so-called legal expert systems (LES) movement which was at its height in the 1970s and 1980s. Because similar expert systems worked in more 'complicated' domains like medicine, it was widely assumed that they would work in the comparatively 'easy' domain of law.[97] A simple idea underpinned LES: '… that one can take rules of law, mould them into a computer-based formal system, and advice will come out the other end'.[98] For instance, the computer scientist L Thorne McCarty confidently proclaimed:

> [Law] seems to be an ideal candidate for an artificial intelligence approach: the 'facts' would be represented in a lower level semantic network, perhaps; the 'law' would be represented in a higher level semantic description; and the process of legal analysis would be represented in a pattern-matching routine.[99]

According to enthusiastic developers at the time, all that was needed to build a useful LES was a way to: (1) translate legislation or 'rules' into a machine readable format; (2) write software that could interpret them, and; (3) gather some legal experts to

[93] AB Feferman and S Feferman, *Alfred Tarski: Life and Logic* (Cambridge University Press, 2004).

[94] J McCarthy, 'Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I' (1960), www-formal.stanford.edu/jmc/recursive.pdf; *cf* SL Andersen, 'John McCarthy: Father of AI' (2002) 17(5) *IEEE Intelligent Systems*.

[95] J McCarthy, 'Philosophical and Scientific Presuppositions of Logical AI' in HJ Levesque and F Pirri (eds), *Logical Foundations for Cognitive Agents* (Springer, 1999).

[96] P Wang and B Goertzel (eds), *Theoretical Foundations of Artificial General Intelligence* (Atlantis Press, 2012); L Muehlhauser, 'What is AGI' (*Machine Intelligence Research Institute*, 2013), intelligence.org/2013/08/11/what-is-agi/.

[97] R Susskind, *Transforming the Law* (Oxford University Press, 2000) 162–206.

[98] P Leith, 'The Rise and Fall of the Legal Expert System' (2010) 1(1) *European Journal of Law and Technology* 1.

[99] LT McCarty, 'Some Requirements for a Computer-based Legal Consultant' quoted in RN Moles, *Definition and Rule in Legal Theory: A Reassessment of HLA Hart and the Positivist Tradition* (Basil Blackwell, 1987) 269.

explicate the legal rules so they could then be imputed by a computer programmer. LES developers thus understood their task as formalising what Hart had termed the 'open texture'[100] of legal rules into mathematical conditionals upon which, as Leibniz aspired to centuries earlier, there could be no disagreement.[101] By attempting this, they sought to reanimate the Leibniz Dream using computation to achieve the 'mechanical jurisprudence' Pound and the legal realists bemoaned as radically simplifying the complex and indeterminate nature of law and legal reasoning.

Despite some middling successes, LES proved both overly ambitious and myopic in the attempt to take Occam's Razor to jurisprudence and the sociological context of law to extract a purified mathematical 'essence'. While the reasons for its failure were several and complex, the collapse of the LES project was primarily due to technical limitations and a gross underestimation of the complexity in applying axiomatic methods to law.[102] However, the seeds of failure were the a priori denial of law's social context, purpose, and anthropological role being relevant to its interpretation. Although LES operated in the 'narrow' domain of law, the nature of legal reasoning, deduction and inference made it clear that they would require wider facets of human intelligence. Because humans learn to perform tasks in ways that allow acquired knowledge to become both tacit and implicit,[103] LES engineers could not use axiomatics to extract and represent legal knowledge and intuition in a workable way.[104]

By the end of the twentieth century, a shift away from the logic-AI approach to the use of connectionist models changed things again.[105] Their success in real-world tasks, such as the defeat of chess grandmaster Gary Kasparov by IBM's Deep Blue, renewed interest in AGI research.[106] In the early twenty-first century the sensationalised victories of IBM's Watson at Jeopardy and DeepMind's AlphaGo at the board game Go seemed to make the goal of 'solving intelligence' both technically feasible and commercially viable.[107] Unlike the logical-AI approach, these newer systems, inspired by neuroscientific models of the brain, did not try to exhaustively formalise human expertise and knowledge to generate axioms. Instead, using both ML and DL techniques, they used vast numbers of historical examples (training data) that allowed them to iteratively update their mathematical parameters and optimise performance at directed tasks. While their results have been undoubtedly impressive, David Spiegelhalter reminds us:

> … that these are technological systems that use past data to answer immediate questions, rather than scientific systems that seek to understand how the world works: they are to be

[100] Hart, *The Concept of Law* 124–54.

[101] WG Popp and B Schlink, 'Judith, A Computer Program to Advise Lawyers in Reasoning a Case' (1974) 15(303) *Jurimetrics Journal* 303.

[102] MZ Bell, 'Why Expert Systems Fail' (1985) 36(7) *Journal of the Operational Research Society* 613; P Leith, 'The Rise and Fall of the Legal Expert System'.

[103] M Polanyi, *The Tacit Dimension*, rev edn (University of Chicago Press, 2009) 3–25.

[104] P Leith, 'Fundamental Errors in Legal Logic Programming' (1986) 29(6) *The Computer Journal* 545.

[105] SI Gallant, 'Connectionist Expert Systems' (1988) 32(2) *Communications of the ACM* 137.

[106] Y Seirawan, HA Simon and T Munakata, 'The Implications of Kasparov vs. Deep Blue' (1997) 40(8) *Communications of the ACM* 21; M Newborn, 'Deep Blue's contribution to AI' (2000) 28(1–4) *Annals of Mathematics and Artificial Intelligence* 27.

[107] D Hassabis D Kumaran, C Summerfield and M Botvinick, 'Neuroscience-inspired Artificial Intelligence' (2017) 95(2) *Cell* 245; JE Laird, C Lebiere and PS Rosenbloom, 'A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics' (2017) 38(4) *AI Magazine* 13.

> judged solely on how well they carry out the limited task at hand, and, although, the form of the learned algorithms may provide some insights, they are not expected to have imagination or have super-human skills in everyday life.[108]

Even with this caveat, we can safely predict that interest in the application of ML to law will continue to grow. This is not just because it is a new, if still largely untried, technology. As our brief historical review has shown, the idea of a functionally complete legal system, one capable of being described using the formal logic and axiomatic reasoning associated with mathematical models, is not new. Despite serial refutations and reversals, it has a deep hold on the legal imagination. Thus today's law-AI debate, in common with earlier iterations between law and technology, poses in a new form a fundamental question about the nature of law as a mode of governance: is law computable? This is the issue which, in varying ways, the contributions to this collection set out to address.

VII. Computationalism and the Mathematisation of Reality

With growing acceptance that biological and artificial intelligence differ only in their substrates,[109] some suggest that AI and data science will someday enable system-level modelling of social reality and the complete prediction of human behaviours.[110] While this is the long term goal, the more immediate goal of using AI to replicate the cognitive domain of lawyers and judges has spurred confidence that a new generation of AI-leveraging LegalTech and ADM systems will succeed where their logic-based forebears could not: replicating legal reasoning, deduction, and inference. With them comes a familiar refrain: LegalTech will improve access to justice, lower costs, and improve the efficiency of legal administration among many other practical benefits.[111] While the functional capabilities of this new generation have undoubtedly improved, the LegalTech enterprise still rests on the Leibnizian-Langdellian assumption that there is a purified essence to law and legal reasoning there to be mathematised. This leads physicist Max Tegmark to wonder:

> … why has our physical world revealed such extreme mathematical regularity that astronomy superhero Galileo Galilei proclaimed nature to be 'a book written in the language of mathematics,' and Nobel laureate Eugene Wigner stressed the 'unreasonable effectiveness of mathematics in the physical sciences' as a mystery demanding an explanation?[112]

[108] Spiegelhalter, *The Art of Statistics* 145.

[109] PS Churchland, 'Can Neurobiology Teach Us Anything About Self Consciousness?' (1994) 42(3) *Proceedings and Addresses of the American Philosophical Association* 23; M Allen and KJ Friston, 'From Cognitivism to Autopoiesis: Towards a Computational Framework for the Embodied Mind' (2018) 195(6) *Synthese* 2459; *cf* M Velmans, 'Is Human Information Processing Conscious?' (1991) 14(4) *Behavioral and Brain Sciences* 651.

[110] M Morrison, Reconstructing Reality: Models, Mathematics, and Simulations (Oxford University Press, 2015); Tegmark, *Our Mathematical Universe*.

[111] Susskind, *Transforming the Law* 237–43; R Susskind, *Tomorrow's Lawyers*, 2nd edn (Oxford University Press, 2017) 93–121.

[112] Tegmark, *Our Mathematical Universe* 6.

Ultimately it is belief in the 'unreasonable effectiveness of mathematics' that drives much of the current AI enterprise, including its use in the legal realm. For instance, contract analytics is a popular focus for LegalTech applications, with a number of start-ups competing to have their proprietary interpretations of contractual terms established as definitive referents. While the business use case is clear, among the many things it neglects is that widely framed general clauses such as 'reasonableness' or 'good faith' operate symbiotically with lower-level concepts and, ultimately, with fact-specific instances of individual disputes. The application of a legal rule to a set of social facts is, in this sense, algorithmic, involving the addition of concepts to form more complex ones, but contingent upon interactions between concepts and rules expressed at different levels of generality. A deterministic or 'mechanical jurisprudence' approach ignoring the interaction between concepts at different levels of generality accomplishes little more than ossifying legal concepts into self-evident computational 'truths'. As a consequence, it advances a radically simplified legal ontology which assumes that these concepts are stable referents part of tractable computational problems decidable by a Turing Machine.[113] Thus the implicit assumption underpinning the LegalTech enterprise is that 'law' is ultimately Turing Complete, and that 'legal problems' are ultimately decidable as 'right' or 'wrong' – or '1' or '0'.[114]

With seeming indifference to why LES failed in the first place, LegalTech developers remain beholden to an axiomatic conception of law, but also an implicit understanding of reality known as pancomputationalism – the view that physical reality is describable by information and that the universe is the deterministic or probabilistic output of a computer program or network of computational processes.[115] According to its most familiar formation in Tegmark's mathematical universe hypothesis (MUH): 'our physical world not only is described by mathematics, but … is mathematics, making us self-aware parts of a giant mathematical object … [and] forcing us to relinquish many of most deeply ingrained notions of reality'.[116] Observers, including humans, existing within this mathematical object are 'self-aware structures' that 'subjectively perceive themselves as existing in a physically "real" world'.[117] Because the MUH suggests that mathematical existence equals physical existence – and all structures that exist mathematically exist physically – reality is ultimately defined by computable functions. By defining law, legal concepts, norms and the range of human and non-human behaviours they pertain to as ultimately 'computable' functions, LegalTech applications and calls for 'AI Judges' endorse a specific understanding about the nature of social reality and information itself. On one hand, they endorse the dualism Leibniz derived from the Yi-jing expressed in binary code but also, on the other, the purely digital ontology posed by the MUH, where everything is computable, because everything is computation.

[113] MacCormick, *What Can Be Computed* 317–30.
[114] MacCormick, *What Can Be Computed* 103–13.
[115] G Piccinini, *Physical Computation: A Mechanistic Account* (Oxford University Press, 2015) 51–73; M Miłkowski, 'From Computer Metaphor to Computational Modelling: The Evolution of Computationalism' (2018) 28(3) *Minds and Machines* 515.
[116] Tegmark, *Our Mathematical Universe* 17.
[117] M Tegmark, 'Is "The Theory of Everything" Merely the Ultimate Ensemble Theory?' (1998) 270(1) *Annals of Physics* 1.

This is problematic, not only because of the unresolved theoretical and practical limits of computation[118] – which raises the wider questions: what is computable? and is law computable? – but because the binary nature of computation means that all legal problems must ultimately be decidable using binary logic.

While Tegmark's hypothesis has attracted considerable attention within the AI community, the MUH remains fiercely debated.[119] Gualtiero Piccini, for instance, notes the deeply contested nature of pancomputationalism: 'some philosophers find it obviously false, too silly to be worth refuting; [while] others find it obviously true, too trivial to require a defense'.[120] Dan McQuillan, meanwhile, connects it to the wider claims of data science, which he describes as:

> … an echo of the neo-platonism that informed early modern science in the work of Copernicus and Galileo. That is, it resonates with a belief in a hidden mathematical order that is ontologically superior to the one available to our everyday senses.[121]

Despite robust critiques of the MUH's 'digital ontology'[122] and broad support for an informational approach to structural realism whereby 'knowledge of the world is knowledge of its structures',[123] current AI research is largely guided by pancomputationalist and materialist views of intelligence.[124]

Seeing the potential of AI across a number of contexts, legal scholars have become somewhat enamoured with AGI. Most often, however, their enthusiasm about the boundless potential of AI runs the risk of presenting it as a 'magic problem solver' for everyday legal issues (as in the case of corporate governance[125]). This usually means neglecting the wider scope of the societal transformation posed by AGI, and assuming that the legal system will remain more or less stable in a post-AGI paradigm. A different account, can however be found in Ben Alarie's prediction of a forthcoming 'legal singularity'.[126] Derived from the technological singularity hypothesised by Ray Kurzweil,[127] Alarie's 'legal singularity' envisions a scenario where machines designing increasingly capable and powerful machines trigger an 'intelligence explosion' and the creation of a 'superintelligence' vastly exceeding human understanding and control.[128] Once this point is reached, 'disputes over the legal significance of agreed facts will be rare. [There] may be disputes over facts, but the [once] found, the facts will map on to clear legal consequences. The law will be functionally complete'.[129]

[118] MacCormick, *What Can Be Computed* 3–11, 294–311.
[119] J Schmidhüber, 'Algorithmic Theories of Everything' (2000), arxiv.org/abs/quant-ph/0011122; P Hut, M Alford and M Tegmark, 'On Math, Matter and Mind' (2006) 36(6) *Foundations of Physics* 765.
[120] Piccinini, *Physical Computation* 51.
[121] D McQuillan, 'Data Science as Machinic Neoplatonism' (2018) 31(2) *Philosophy & Technology* 254.
[122] L Floridi, 'Against Digital Ontology' (2009) 168(1) *Synthese* 151.
[123] J Smithies, *The Digital Humanities and the Digital Modern* (Palgrave, 2017) 55; L Floridi, 'A Defence of Informational Structural Realism' (2008) 161(2) *Synthese* 219.
[124] Hassabis, Kumaran, Summerfield and Botvinck, 'Neuroscience-Inspired Artificial Intelligence'.
[125] M Petrins, 'Corporate Management in the Age of AI' (UCL Working Paper Series, Corporate Management in the Age of AI No. 3, 2019), https://dx.doi.org/10.2139/ssrn.3346722.
[126] Alarie, 'The Path of the Law' 443.
[127] R Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (Gerald Duckworth & Co. 2006).
[128] Bostrom, *Superintelligence*.
[129] Alarie, 'The Path of the Law' 3.

While results should not be overstated, some systems have already 'outperformed' human legal experts at the task of predicting case outcomes.[130] There is, however, a gulf between using statistical techniques to predict case outcomes using superficial criteria, such as the jurisdiction and political affiliation of a judge, and replacing what judges do with AI. Nevertheless, the ambitions of what we might call the 'legal singularitarians' do not stop at predicting case outcomes. They have in their sights the eventual replacement of juridical reasoning as the basis for dispute resolution and the substitution of some protean triumvirate of powers, rights, and responsibilities for legitimate legal authority.

VIII. Chapter Overview

The papers collected in this volume were first presented as part of 'Lex Ex Machina: A Conference on Law's Computability' held at Jesus College, in the University of Cambridge, on 13 December 2019. This conference was borne out of the desire to gather together some of the most influential scholars working at the intersection of law/technology and explore what 'computational law', 'robot judges' and the 'legal singularity' mean for the future of law as a social institution, and to push back against some of the more sensationalist claims emanating out of legal scholarship about the use of AI in law.

A. Christopher Markou and Simon Deakin – *Ex Machina Lex*: Exploring the Limits of Legal Computability

In their chapter *'Ex Machina Lex*: Exploring the Limits of Legal Computability' Christopher Markou and Simon Deakin pose the question: to what extent do the statistical techniques of ML/DL lend themselves to formalisation of legal reasoning? While new use cases are being identified for AI in law, their article assesses the feasibility of their success and what 'success' would indeed mean in legal contexts. Although there are material and practical limits to computation and data storage, and theoretical limits to the computability of all problems, this does not mean that legal problems are necessarily non-computable. Some might be, but not necessarily all. The authors thus explore the extent to which the legal system – a system in the sense implied by the theory of social systems – is amenable to computation and automation, and how far the replacement of juridical reasoning with strategic and computational reasoning might impact the autonomy of the legal system, erode the rule of law, and diminish state authority in structuring and mediating legal relations. Their approach adopts a systemic-evolutionary understanding of law to identify unifying principles that help explain the legal system's mode of operation with respect to other social sub-systems, including the economy, politics and technology itself, and which help to clarify the role of juridical reasoning in facilitating legal evolution.

[130] J Goodman-Delahunty, PA Granhag, M Hartwig and EF Lotus, 'Insightful or Wishful: Lawyer's Ability to Predict Case Outcomes' (2010) 16(2) *Psychology, Public Policy and Law* 133; B Alarie, A Niblett and A Yoon, 'Using Machine Learning to Predict Outcomes in Tax Law' (2017), dx.doi.org/10.2139/ssrn.2855977; DM Katz, MJ Bommarito I and J Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' (2017) 12(4) *PLoS One* e0174698.

The chapter suggests that the hypothetical 'legal singularity' – which presumes the elimination of all legal uncertainty – conflates simulation and the probabilistic capabilities of ML and Big Data for the process of legal judgment. The authors build on Mireille Hildebrandt's observation that '[w]hereas machines may become very good in such simulation, judgement itself is predicated on the contestability of any specific interpretation of legal certainty in the light of the integrity of the legal system – which goes beyond a quasi-mathematical consistency'[131] and, following Alain Supiot's identification of the 'anthropological function' of law as a 'technique [for the] humanization of technology'[132] contend that the replacement of juridical reasoning with computation would ultimately result in the subordination of the 'rule of law' to a new 'rule of technology'.

The chapter develops this critique through an examination of one of the issues high on the agenda of those arguing for a computational approach to law, namely the determination, for tax and employment purposes, of the distinction between employees and independent contractors. This distinction is shown to be historically contingent and to have been shaped by numerous economic and political factors. To reduce the juridical task of work classification to an automated process would conceal the political choices which are unavoidably present in these areas of law.

B. Mireille Hildebrandt – Code Driven Law: Freezing the Past and Scaling the Future

In 'Code Driven Law: Freezing the Past and Scaling the Future', Mireille Hildebrandt examines 'cryptographic law' or 'smart regulation' (that is, self-executing code as a new type of legal regulation). She begins by drawing an important distinction between data-driven law and code-driven law. The former pertains to predictions of legal decisions or mining of legal arguments, which is often discussed in Legal AI or LegalTech literature, and in Hildebrandt's wider works on 'law as computation'.[133] The latter, on the other hand, concerns the 'legal norms or policies that have been articulated in computer code, either by a contracting party, law enforcement authorities, public administration or by a legislator'. Her focus is on how these code-driven laws establish pre-determined thresholds that trigger 'smart'/automated regulations (as in the case of social security fraud detection).

The first part of her paper describes what code-driven law 'is' by explaining what it 'does' with respect to text-driven law. She provides examples of code-driven laws in smart contracts (which are 'not only articulated in computer code but also self-executing'), public administration ('a decision-support or a decision-making system that is articulated in computer code'), and legislatures. Hildebrandt argues that these examples of code-driven laws raise questions about principles in private law, public law, criminal

[131] M Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2017), www.ssrn.com/abstract=2983045.

[132] A Supiot, *Homo Juridicus* 117.

[133] M Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence'.

law and constitutional law. She concludes that what the code-driven law does is to 'fold enactment, interpretation and application into one stroke, collapsing the distance between legislator, executive and court'. The implicit assumption of code-driven law is that it can foresee all potential scenarios in order to cover all future interactions, which out of necessity must be 'highly dynamic and adaptive to address and confront what cannot easily be foreseen by way of unambiguous rules'.

Hildebrandt then turns to the capacity assumption of code-driven law – that is, our ability to sufficiently foresee the future at the moment legal norms are encoded into law. She argues that compared to text-driven laws – which are structured on natural language and speech acts – code-driven law has certain constraints in its computational architecture and design. These include the requirement of formal deduction ('if this then that', IFTTT), disambiguation of terms and rules, and the incompleteness and inconsistency of computations.[134] These constraints are all 'related to the uncertainty that inheres in the future'. ML embraces a false assumption that the distribution of training data can be a close approximation of the distribution of future. In reality, the future distribution of data can only be predicted rather than learnt by ML.

Hildebrandt claims that predictions influence the very behaviours they supposedly predict, and rephrasing the Goodhart/Campbell/Lucas effect argues that 'when a measure becomes a target it ceases to be a good measure'. While we could design many present futures, there is only one future present (reality), and 'the best way to predict the future is to create it'. She concludes by claiming that Arendt's 'human condition' can be best explained in terms of the Parsonian/Luhmannian 'double contingency', and that natural language makes the process of anticipating a process of anticipating how others anticipate us. The co-evolutionary nature of anticipation based on interpretable natural language thus generates radical uncertainty, which in turns demands the institutionalisation of specific patterns of behaviours to consolidate, stabilise, and reduce both complexity and uncertainty in social systems. Legal certainty is thus crucial to this process of consolidation and stabilisation, and occurs without freezing the future based on a scaling of the past due to its text-driven nature.

Hildebrandt's paper concludes with an inquiry into the nature of code-driven law, pointing out the problems in the concept of 'legal by design' and the formulation of her own approach – 'legal protection by design'. The latter does not focus on compliance or enforcement of legal norms, but strategies for embedding legal protections into the technological infrastructure of code-driven and data-driven environments. At the very least, she contends, code-drive decision-making systems must come with an effective right to appeal against automated decisions and to give legal justification. Her ultimate conclusion, however, is that law is not computable due to its text-driven and multi-interpretable nature, which mean it could be computed in various ways. The choice for how to design 'legal computation' must therefore belong to the 'people' and the courts rather than software developers in big tech or big law.

[134] *cf* K Gödel, 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls' (1930) 37 *Monatshefte für Mathematik und Physik* 349; JM Rogers and RE Molzon, 'Lessons about the Law from Self-Referential Problems in Mathematics' (1992) 90 *Michigan Law Review* 5.

C. John Morison – Toward a Democratic Singularity?

In the chapter 'Toward a Democratic Singularity?' John Morison considers foundational issues surrounding the implementation of automated systems and their attendant impact on democracy and the legitimacy of legal regimes. The paper first outlines shortcomings in the promise of online consultations, and how they have yet to realise their democratic potential. Morison then speculates on the development of new surveillance technologies, which purport to transform virtually all aspects of human existence into quantifiable data that is used to monitor behaviours and allow for greater predictive capabilities. He then examines how these technologies have been used to enable 'biosurveillance' in the context of the Covid-19 pandemic.

Morison's argument is that the convergence of digital technologies amounts to a new, pervasive, and all-encompassing form of surveillance and social control he terms 'algorithmic governmentality'. Building upon insights from the shortcomings of democratic consultations, he argues that these developments run the risk of rendering ideas about consultation and deliberative democracy all but redundant 'as actual preferences can be measured directly without the need for an intermediary political process to represent preferences'. The totalisation of these technologies, he suggests, allows for the elicitation of individual 'preferences' through various means of statistical inference, profiling, and the wider process of 'radical datafication' which 'offers a false emancipation by appearing to be, by its very nature, all-inclusive and accurate'.

Morison's conclusion is that these developments amount to a novel form of governance, one that is post-political, and has the potential to first 'undermine, and then transcend, many of fundamental attributes of citizenship which presently appear as part of the bargain within the government-governed relationship'. His conclusion is that we should resist efforts to de-politicise politics by removing it from the process of detailed decision-making and replacing it with algorithmic governmentality, which in turn must be resisted by debating and questioning the processes themselves.

D. Jennifer Cobbe – Legal Singularity and the Reflexivity of Law

In 'Legal Singularity and the Reflexivity of Law' Jennifer Cobbe argues that those pursuing legal AI – particularly those interested in the so-called 'legal singularity' – misunderstand both the nature of law and of technology. She contends that not only would they fail to solve the very real problems of the law, but they could potentially make them worse and cause new and greater problems.

Drawing on insights and concepts from a variety of disciplines, Cobbe's argument is premised on the idea that law functions as a reflexive societal institution. In her view, law not only reflects society but significantly influences society and its institutions. As such, she argues, the law cannot be neutral, as it is inherently contextual and contingent on the circumstances of the time, and imbued with normative assumptions and priorities of that time. Ultimately, the law reifies the interest and goals of its creators (legislatures), practitioners (lawyers), and adjudicators (judges). This reflexive functioning should be understood as distinct from the role that law plays in society as a result of its reflexivity. In Cobbe's view, law's role in society, both historically and in the present, has been to entrench the power of capital, strengthen the position of the wealthy, reinforce

inequalities, and protect established interests from outside challenges. Functioning as a reflexive societal institution, Cobbe argues, the law's role has not only been to reflect the inequalities and injustices in society, but establish societal conditions that repeat, reinforce, and re-encode them back into society. Cobbe further argues that algorithms, too, are reflexive: 'Just as law in its reflexivity moulds society according to the subjective assumptions, understandings, and goals of those who write and practice it', she says, 'so too with algorithms'. As such, just as it matters what goals and priorities those working within the law are pursuing, and whose interests they serve, it matters what goals and priorities are being pursued in the design, deployment, and use of algorithmic systems.

Having laid the groundwork by describing the reflexivity of both law and AI, Cobbe proceeds to develop her argument in two parts. In the first, she argues that algorithmic systems are not – and may never be – capable of replacing human lawyers and judges. She highlights some faulty assumptions relied upon by legal AI proponents, such as the idea that AI systems can engage in legal reasoning or be neutral and objective arbiters. She emphasises, though, that critiques of legal AI that focus on the technical limitations of algorithmic systems, while important, do not get to the heart of the structural questions in which she is primarily interested.

In the second part of her argument, therefore, Cobbe formulates a critique of the power relations and structural effects of legal AI as it is commonly envisaged. Adopting Foucauldian theories of governmentality, she begins by locating the ideological underpinnings and motivations of legal AI as being part of a process of neoliberal rationalisation; replacing the qualitative, normative values of law with supposedly rational, objective, quantitative metrics and logics based on statistical and economic thinking. As part of this process of rationalisation, the law is often problematised by legal AI proponents as slow, costly, inefficient, complex, unpredictable, in need of optimisation, and thus amenable to techno-solutionist interventions. By framing the case for legal AI in terms that are fundamentally concerned with the quality of the law's functioning, advocates for legal AI – and the 'legal singularity' more generally – not only fail to consider the nature of the law's role in society but also prioritise the kind of market-oriented and commercially-driven ways of thinking that contribute to the development of problems with that role in the first place. Without a critical examination of the law's role in society, Cobbe argues, legal AI proponents therefore risk developing systems that will primarily make law 'better' at extending and reinforcing hierarchies, maintaining the law's exclusionary effects, and reifying the dominance and power of capital.

E. Roger Brownsword – Artificial Intelligence and the Functional Completeness of Law

Roger Brownsword's chapter 'Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law' starts with the observation that 'a number of technological wedges are being driven under the idea of law as a rule-based enterprise; that the wedges that are being driven in relation to the channelling and re-channelling function of law are much more significant than those being driven into adjudication' and that at least some of these technological wedges are 'going in thick end first'. Brownsword contends that these challenges may well necessitate

a 'radical rebooting of our legal thinking' and potentially reshape 'the Rule of Law and our conception of coherence in the law'.

The argument of his paper can be broken down into three axes. The first is that the use of algorithmic or 'AI' tools by both public and private regulators must be guided by a comprehensive technological management strategy, and grounded in a revised conception of what regulatory responsibility entails. Secondly, he suggests that this revised understanding, which he describes as a 'new benchmark for legality', must be circumscribed into the rules governing the exercise of regulatory power that depend on these technological measures. Thirdly, Brownsword suggests that the revision of these rules requires what he terms a 'new coherentism' which focuses on the compatibility, and inter-contingency, of regulatory measures with newly established benchmarks for legality. By revising 'traditional coherentist thinking' – which is concerned with how general legal principles apply to particular fact patterns – Brownsword formulates 'how a new coherentist mind-set needs to be cultivated so that there is a constant scrutiny of technological measures to check that they are compatible with the benchmark regulatory responsibilities'.

Brownsword concludes his paper with reflections on the need for new institutional configurations that are better suited to nurturing and sustaining coherentist legal reasoning, and supporting 'the stewardship responsibilities that regulators have for the global commons'.

F. Sylvie Delacroix – Automated Systems and the Need for Change

Sylvie Delacroix's chapter, 'Automated Systems and the Need for Change', examines the use of AI systems in moral decision making. One of the problems with this idea, she suggests, is that AI systems would not arrive at a form of moral consciousness in any way that we would recognise as characterising our own experience. In particular, current efforts to develop automated systems to be deployed in morally loaded contexts pay little attention to the difficulties that stem from the unavoidable need for moral change. If we take the view that ethics is a 'work in progress', it would be essential to retain a role for human agents in the process of value formation.

Systems like law which are designed to provide simplified guidance on how to live together can induce conformity and dissipate pressure for change. Automated systems are similarly 'morally risky'. Dangers include the backward-looking nature of machine learning. Delacroix explores the use of inverse reinforcement learning (IRL) as a possible solution. IRL involves the design of 'systems that are meant to infer from the behaviour of surrounding (human) agents a morally-loaded "utility function"'. There is interest in the use of AI to allow data to train the system, avoiding the need to select data 'by hand'. The utility function is not fixed but evolves in the light of new data on agents' moral judgments.

The possibility of automating moral choices may seem attractive in the light of the difficulty human beings have in assessing practices to which they become habituated. Moral capacity cannot be preserved through cognitive vigilance alone, however, as the latter is conditioned by the habits of thought acquired through immersion in a particular

social environment. For some who take a 'realist' stance on issues of morality, the possibility of a superintelligence that would set us on the path to righteousness is attractive. However, Delacroix warns that 'without regular exercise, our moral muscles will just wither away, leaving us unable to consider alternative, better ways of living together'.

A more pragmatic aim, she argues, would be to 'build human-computer interactions that are apt at dispelling moral torpor'. One way to do this would be to include human beings as 'end-users' in the decision loop, as in the case of 'interactive machine learning'. But 'in the absence of a radical shift in the design choices that preside over the way those systems call for interaction with us, lazy normative animals, that effect will be dramatic, to the point of possibly undermining the very possibility of human-triggered change'.

G. Ryan Abbott and Alex Sarch – Punishing Artificial Intelligence: Legal Fiction or Science Fiction?

Ryan Abbott and Alex Sarch's chapter, 'Punishing Artificial Intelligence: Legal Fiction or Science Fiction', explores the question of whether it is feasible to consider applying criminal sanctions to AI entities. They point out that the idea of applying the criminal law to non-human entities is nothing new, as the criminal responsibility of corporations is, in principle, well established. Corporate persons are subject to criminal penalties, they suggest, where organisations cause systemic harms which are not reducible to the actions of individual human beings. If an AI system has the potential to cause similarly 'irreducible' harms – 'Hard AI crimes' – the possibility of punishing AI needs to be at least carefully considered.

While there is no immediate prospect of a 'strong AI' which can exactly replicate human cognition, it is possible to envisage cases in which AI systems could behave not just unpredictably and unexplainably, but also autonomously of any direct human control. Machines may then cause harms which would be categorised as criminal but where no identifiable human person has acted with criminal culpability, or when it is not practicably defensible to attach criminal culpability to any one human actor, as in the case of actions taken by many individuals over a period of time or in a complex organisational setting. In these circumstances, punishment of an AI entity could indirectly deter the individuals who develop and market it, and the law's expressive function might be satisfied in the legal condemnation of certain outcomes associated with AI.

Abbot and Sarch then discuss whether it is possible to attribute mental states to AI systems. This is already done, they point out, in the case of corporate entities, through theories of agency, as in the case of the respondeat superior principle. This involves looking to the mental state of the user or developer responsible for the AI. However, it runs up against the problem that it may not be clear who is responsible for an AI. Use of criminal law techniques for imposing liabilities on those ultimately responsible for serious harms, such as the doctrine of constructive liability, runs the risk of imposing excessive liabilities on developers and stifling valuable innovation.

On the vexed question of whether to attribute legal personality to AI systems, Abbot and Sarch warn against the danger of 'rights creep', as has happened already with the attribution, in a number of jurisdictions, of constitutionally-protected human rights to corporations. One solution, they suggest, would be to designate a 'responsible person', in

practice likely to be a corporation, as the ultimate guarantor of an AI. They also consider the costs and benefits of requiring mandatory registration and reporting for certain AI systems.

They conclude that notwithstanding 'the growing possibility of Hard AI Crime', the 'radical tool of punishing AI' would be an overreaction. Preferable would be 'modest expansions to criminal law, including, most importantly, new negligence crimes centered around the improper design, operation, and testing of AI applications as well as possible criminal penalties for designated parties who fail to discharge statutory duties'.

H. Lyria Bennett Moses – Not a Single Singularity

Lyria Bennett Moses' chapter, 'Not a Single Singularity', envisages the development of AI in law occurring along three dimensions: an 'x axis' describing what is currently possible at any given time, a 'y axis' referring to what the technology is capable of becoming, and a 'z axis' along which decisions about the legitimacy of AI will be made. She argues that rather than there being a 'straight path' towards a legal singularity, there are likely to be periods of alternating progress and stasis along each of the three axes, and that the final outcome is unlikely to be a fully computable legal system.

In the context of legal AI, human and machine capabilities are more likely to be complements than substitutes. Machines can currently simulate certain human tasks, and if the focus is on observed functions they may be understood as simulating certain forms of rational behaviour. However, the current state of AI systems is such that they are very far from achieving anything approaching general intelligence. When particular techniques are closely scrutinised, their limitations with respect to law come into sharp relief. Thus attempts to make legislation machine-readable through the use of expert systems techniques have shown that it is most straightforward to code those statutory texts which embed an element of calculation or which rely on relatively straightforward decision rules. The danger in seeking to translate legislation into code is that it will change the way in which laws themselves are drafted. Expert systems cannot code for open-ended concepts such as 'reasonableness', but general clauses of this kind which, by virtue of their incompleteness, permit legal adaptation.

Similar problems arise when natural language processing and machine learning are used for case prediction. The current state of these techniques is inadequate given their tendency to replicate racial and other biases, as the case of the COMPAS software has highlighted. But even if these capabilities can be enhanced in future, there are problems of legitimacy associated with their use. In the context of legal decision making, judgment is not a purely predictive process; it is the outcome of deliberation and contestation.

She concludes that an uneven development of AI in law is likely, not just because of limits to what can be achieved technically, but by virtue of unavoidable concerns over the legitimacy of automated decision making. Technological advances 'will lead to both progress and error, sometimes expanding what is available, what is possible and what is appropriate and legitimate in ways that are both evolutionary and revolutionary. But there are many thresholds to cross, and it is hard to imagine a system that would render law fully computable without changing the nature of law itself'.

I. Dilan Thampapillai – The Law of Contested Concepts: Reflections on Copyright Law and the Legal and Technological Singularities

Dilan Thampapillai's chapter 'The Law of Contested Concepts? Reflections on Copyright Law and the Legal and Technological Singularities' argues that while the idea of the legal singularity might seem convincing enough at a high level of abstraction, it runs up against serious objections in context of an applied field of law such as intellectual property ('IP'). IP law reflects a complex bargain around notions of property, human rights, free expression, technological development, and education, among other things. Copyright can only function as an imperfect property system in which the courts have discretion to balance competing interests and values on a case by case basis.

As a form of property law, IP gives expression to the right of exclusion. However, exclusion has always been an imperfect vehicle for realising the goals of IP law. Copyright affords the right to claim to redress for violations of exclusive rights, but it cannot guarantee that the right-holders' interests will be completely protected. The legitimacy of the law depends also on recognising alternative interests, and in allowing carve-outs such as the fair use doctrine. It is only by accepting that it serves a range of interests that copyright law retains its legitimacy.

The legal singularity, by contrast, describes a version of a complete legal system, overseen by a superhuman intelligence. Such a system is premised on the possibility of the perfect enforcement of legal rights. This, Thampapillai suggests, is inherently antithetical to the kind of multi-factorial decision making which occurs in copyright cases. The fair use doctrine has evolved in such a way as to allow courts to strike complex compromises in the cases that come before them. The concepts they use to strike these balances are useful precisely because they are open-textured and contestable at the point of interpretation. Concepts also express a value dimension which no amount of data can straightforwardly capture.

For these reasons, Thampapillai suggests that scepticism is in order when contemplating the legal singularity. The example of copyright highlights the degree to which 'human systems and reasoning have proved resilient and adaptable' in responding to the challenges of technological change; the vision of the legal singularity, by contrast, is one in which not just human labour but the human subject itself is pushed to the margins.

J. Christopher Markou and Lily Hands – *Capacitas Ex Machina*: Are Computerised Systems of Mental Capacity a 'Red Line' or Benchmark for AI?

In the volume's concluding contribution, '*Capacitas Ex Machina*: Are Computerised Assessments of Mental Capacity a "Red Line" or Benchmark for AI?', the prospect of using AI and advanced technologies such as fMRI and Automated Mental Stated Detection is taken up by Christopher Markou and Lily Hands. Their paper begins by reviewing the history of AI in medicine, starting with the advent of Expert Systems (ES) in the mid-twentieth century, to the rise of connectionist AI research in its latter

half, to the development of Automated Mental State Detection (AMSD) and related bio-cognitive interfaces. It examines theoretical and practical problems for implementing these systems in the real world, and how psychiatry is likely to be impacted in the near term by technological advances.

The paper situates this as for now hypothetical problem within the context of historical efforts to use computers to assist with medical diagnosis, and potentially replace human doctors and psychiatrists. It then examines whether and how computational reasoning could, and indeed should, operate in the context of capacity decisions in England and Wales. Following the critique of computer scientist Joseph Weizenbaum, Markou and Hands highlight not only the technical challenges faced in capturing, encoding, and applying domain expertise in medicine, but the normative – and deeply political question – of whether a computer should under any circumstances be given the authority to make legally consequential judgements turning upon subjective psychological and psychiatric phenomena. Highlighting many of the problems identified earlier in the volume by Hildebrandt, Morison and others, they argue that allowing a machine to make determinations of mental incapacity is not just, at least for now, technically infeasible, but that 'the essential humanity and consequence of capacity decisions on not just the individual, but their community, demands that capacity be not only defined, but assessed and imputed by members of that community'. They conclude that 'anything else would result in a machine being elevated to the role of arbiter of human behaviour and experience within a social reality it cannot access'.

IX. Conclusion

The chapters that comprise this volume are part of a deliberate effort to push-back against the more hagiographical accounts of AI in law, and to help delineate domain specific challenges that must be considered in the course of their development before they can be reliably deployed in real-world legal contexts. We hope that this volume serves as a socratic entry point for lawyers, legal scholars, and students alike, but that it also helps bridge the gap between the technical dimensions of AI research and its normative implications for effective policy and governance.

Although the contributions to this volume identify a range of legal, ethical, and political challenges related to the automation of justice more generally, they help us better connect present and near-term challenges to those that might be faced in the longer-term. Perhaps the real danger is not that the legal singularity occurs, but that what is willed to truth as 'just' by the modern AI enterprise is, in fact, nothing more than what is 'satisficed' as 'just enough'. In an AGI paradigm the prospect of AI judges is, perhaps, a trivial one. We simply do not, and cannot, know. But we do have reason to think that it might be a scenario where all bets are off for humanity and by some accounts our odds aren't good.[135]

[135] E Yudkowsky, 'Artificial Intelligence as a Positive and Negative Factor in Global Risk' in N Bostrom and MM Cirkovic (eds), *Global Catastrophic Risks* (Oxford University Press, 2008).

Yet pinning a hypothetical argument on long-term and unspecified conditionals is just an elaborate way of counting '1, 2, 3, a million' and calling it good numerical sense and a sound business proposition. While for some the latter seems nonetheless true, all it does is frame the totalisation of AI in law as inevitable once an arbitrarily defined threshold of 'intelligence' is attained or various functional capabilities achieved. Indeed, the perspectives contained in this volume are advanced in an attempt to fulfill Ray Kurzweil's suggestion that concerned and constructively critical observers of technology will be needed to manage what he sees as the inevitable transition to a post-singularity world:

> My own expectation is that the creative and constructive applications of this technology will dominate, as I believe they do today. But there will be a valuable (and increasingly vocal) role for a concerned and constructive Luddite movement (i.e., anti-technologists inspired by early nineteenth century weavers who destroyed labour-saving machinery in protest).[136]

It is hoped that this volume, its contributions, and the ongoing work of its contributors inspires like-minded sceptics, critical thinkers, and rebels to challenge the orthodoxy of Big Tech and its salvific claims of limitless potential and untold prosperity. As Julia Powles and Helen Nissenbaum remind us:

> … the preoccupation with narrow computational puzzles distracts us from the far more important issue of the colossal asymmetry between societal cost and private gain in the roll-out of automated systems. It also denies us the possibility of asking: Should we be building these systems at all?[137]

[136] R Kurzweil, 'The Law of Accelerating Returns' (2001), www.kurzweilai.net/the-law-of-accelerating-returns.
[137] J Powles and H Nissenbaum, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence' (Medium, 7 December 2018), medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.