DEUX MACHINA: A CROSS-DISCIPLINARY APPROACH TO ARTIFICIAL INTELLIGENCE FOR REGULATORY UNDERSTANDING

AYBAN ELLIOTT-RENHARD

ST ANTONY'S COLLEGE

MASTER OF PHILOSOPHY IN LAW

TRINITY TERM 2021

ABSTRACT

DEUX MACHINA: A CROSS-DISCIPLINARY APPROACH TO ARTIFICIAL INTELLIGENCE FOR REGULATORY UNDERSTANDING

AYBAN ELLIOTT-RENHARD | ST ANTONY'S COLLEGE
MASTER OF PHILOSOPHY IN LAW | TRINITY TERM 2021

Legal scholars, policymakers and artificial intelligence researchers disagree about how to typify and describe artificial intelligence; the goalposts are constantly shifting with the advancement of technology, but this is an ancillary matter. Definitions aside, sophisticated decision-making technology is widespread and influential, and a more important normative question needs answering: when should machines be used to make decisions? Since the law is premised on clear avenues of responsibility, rationality, and objective justifications for decisions, the prospect of complex, non-human decision-makers tendering inscrutable decisions with considerable utility but ineffable reasoning is a novelty without comparison in the realm of technology regulation. I contend that this novelty necessitates a deeper, contextual understanding of artificially-intelligent decision-making technology to provide a foundation for regulatory answers to the normative question and resolve ambiguity that pervades more general discussion. Meaningful answers cannot be obtained in the abstract and are tied to the domain in which machine decisions are made; broad doctrinal approaches that begin from the law, and take a technology-neutral approach to regulation, are thus inadequate. Medicine, a science and an art with a long history of integration of artificially-intelligent technology, provides a useful domain for close inquiry. Reliance on machine decision-making by physicians and patients in clinical medical contexts reveals important considerations that high-level discussions do not uncover; the most important of these is that data-driven machines reason in a manner that is alien to human reasoning, and this incongruity is the source of the unintelligibility described vaguely as the 'black box' of machine reasoning. Nonetheless, carefully designed machine reasoning is consistent, insightful, and useful, and should be constrained by predefined structural roles for machines, rather than human notions of explanation. Comprehensive, unambiguous understanding is a fundamental prerequisite to the regulatory prescription of suitable roles for machines, and understanding is what I provide here.

THESIS LENGTH: 29, 989 words

TABLE OF CONTENTS

INTRODUCTION **4**

PART I: THE LAW **12**

TECHNOLOGY REGULATION AND ARTIFICIAL INTELLIGENCE **12**

Artificial intelligence has a regulatory effect 14

Artificial intelligence is novel and has no antecedent 18

A FLAWED LEGAL APPROACH TO ARTIFICIAL INTELLIGENCE **23**

Future-making 25

Anthropomorphisation 27

Terminological ambiguity 30

PART II: ARTIFICIALLY-INTELLIGENT TECHNOLOGY IN CLINICAL MEDICINE **35**

Introduction 35

ARTIFICIAL INTELLIGENCE **39**

The evolution of definitions over time 40

Definitions proposed by artificial intelligence researchers 40

Definitions used by lawmakers 41

Terminology in medical literature 44

ARTIFICIAL INTELLIGENCE AND MEDICINE **46**

A brief history 46

Applications of artificially-intelligent technology in medicine 49

Clinical medicine and clinical judgement 52

CLINICAL DECISION SUPPORT SYSTEMS **59**

Introduction 59

Existing taxonomies of clinical decision support systems 62

KNOWLEDGE-BASED AND DATA-DRIVEN CDSS **65**

Introduction 65

Distinction in the literature 67

Knowledge-based CDSS 70

Data-driven CDSS 77

PART III: CONSEQUENCES & QUESTIONS **93**

LEGAL & ETHICAL ISSUES IN THE CONTEXT OF MEDICINE AND CDSS **95**

Liability 95

Fiduciary and professional relationships 98

Autonomy and independence: the status of machines 100

CONCLUSION **104**

BIBLIOGRAPHY **107**

INTRODUCTION

What happens when we give decision-making authority to artificially-intelligent technology? Some say the inevitable result will be a gradual loss of human skill,[1] or perhaps a decline in the critical faculties and moral reasoning skills necessary for meaningful political participation.[2] Where decisions are not delegated but rather made off-stage without our conscious awareness, the more nefarious consequence will be an undetectable loss of control over our environment, which shifts and changes to coax us into behaviour that we may not otherwise have chosen for ourselves.[3] On the other hand, though, what can we say about ourselves? Human beings are morally fallible, poor historians, susceptible to myriad cognitive biases and limitations that render even expert judgement unreliable,[4] inclined to post hoc rationalisations of our affective (emotional) preconceptions,[5] often outperformed in predictive tasks by simple mathematical models,[6] and subservient to 'mental processes that are put into motion by features of the environment and that operate outside of conscious awareness and guidance'.[7] In short, we

[1] Robert Lehr Goodman, 'Commentary: Health Care Technology and Medical Education: Putting Physical Diagnosis in Its Proper Place' (2010) 85 Academic Medicine 945; Federico Cabitza, Raffaele Rasoini and Gian Franco Gensini, 'Unintended Consequences of Machine Learning in Medicine' (2017) 318 JAMA 517.

[2] Jamie Bartlett, *The People vs Tech: How the Internet Is Killing Democracy (and How We Save It)* (Ebury Press 2018) ch 1.

[3] See generally Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019); Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar 2016).

[4] See, for example, Eliezer Yudkowsky, 'Cognitive Biases Potentially Affecting Judgement of Global Risks' in Nick Bostrom and Milan M Cirkovic (eds), *Global Catastrophic Risks* (Oxford University Press 2008).

[5] Jonathan Haidt, 'The New Synthesis in Moral Psychology' (2007) 316 Science 998, 998.

[6] See generally the seminal work of Paul E Meehl, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press 1954).

[7] John A Bargh and Tanya L Chartrand, 'The Unbearable Automaticity of Being' (1999) 54 American Psychologist 462, 462.

might not like the idea of machines making decisions for us, but the alternative is not very good, either.

Fortunately, and contrary to the general gist of scholarship on the topic of artificial intelligence, we have increasingly reliable machines to make better decisions for us,[8] and we are already using them, often to great effect. Indeed, for many decades, we have had machines or other statistical methods to helpfully correct our judgement. This is not to dismiss the prevalence of algorithmic bias or systematic discrimination that forms an undeniable part of machine decision-making, and which we are right to guard against by insisting on transparency and human oversight.[9] And of course, there are other reasons to retain our discretion: as Jamie Bartlett points out, a body politic in which machines make political decisions might be 'wiser and more humane', but it would not be democratic;[10] and as Tim Wu reflects, the convenience of automation 'threatens to erase the sort of struggles and challenges that help give meaning to life'.[11] Nonetheless, we should recognise that it is no longer possible to make aversion to the cold, calculating computer a blanket rule; we're not going to stop relying on machine judgement, and nor should we.

In some cases, machine decision-making is more reliable or less biased than human beings because it relies not on human knowledge that has been extracted and formalised,

[8] See, for example, Alex P Miller, 'Want Less-Biased Decisions? Use Algorithms.' [2018] *Harvard Business Review* <https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms> accessed 3 August 2021; Dimitra Kamarinou, Christopher Millard and Jatinder Singh, 'Machine Learning with Personal Data' (Queen Mary University of London, School of Law 2016) Legal Studies Research Paper 247/2016 22 <https://papers.ssrn.com/abstract=2865811> accessed 29 July 2021.

[9] Evidence of bias and discrimination is widespread and of genuine concern. Note, however, that biases inherent in historical data are not always amplified: see, for example, Ashesh Rambachan and Jonathan Roth, 'Bias In, Bias Out? Evaluating the Folk Wisdom' [2020] arXiv:1909.08518 [cs.LG] <http://arxiv.org/abs/1909.08518> accessed 4 August 2021.

[10] Bartlett (n 2) 21.

[11] Tim Wu, 'The Tyranny of Convenience' *The New York Times* (16 February 2018) <https://www.nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html> accessed 4 August 2021.

but on the 'magic' of data, which reveals sparse, complex relationships that human beings can neither identify nor understand in complete detail. Deep learning (a subset of machine learning), in particular, is a computational technique that is sensitive to such relationships and underpins the rapid advancement of many forms of machine learning over the last decade, but many simpler and older computational methods also have this capacity. In these situations, we not only delegate our decision-making role, but also give up our ability to ground the decisions that are produced on human-comprehensible reasoning. Depending on the decision-making task, this may or may not be acceptable; in the case of some exercises of state authority, or the use of autonomous weapons,[12] it clearly isn't.

Inevitably, though, where we can ensure reliability and fairness, we will have access to insightful decision-making that we can depend on but not fully explain. That is to say, although our understanding of technologies such as deep learning will mature, human beings are subject to cognitive limits, since 'we cannot physically expand the human brain',[13] but the volume of data that may inform and sharpen machine decisions has no such limitation—'[v]ariable-size learners can in principle learn *any* function given sufficient data'[14]—and there is little reason to limit relevant insight to that which can be explained in human terms. This is especially the case in medicine, where we are not only used to the idea of inexplicable decision-making (clinical intuition), but where the complex interrelationships between biological processes in the human body far exceed the human ability to accurately model them. In this domain, inexplicability is a poor basis on which

[12] Although automating warfare might mean fewer combatant deaths overall, there is still something unsettling about delegating authority to take a human life to a machine. For an interesting narrative review of the state of the art of autonomous weapons, see generally Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (WW Norton & Company 2019).

[13] Tom Stonier, 'Machine Intelligence and the Long-Term Future of the Human Species' (1988) 2 AI & Society 133, 138.

[14] Pedro Domingos, 'A Few Useful Things to Know about Machine Learning' (2012) 55 Communications of the ACM 78, 85 (emphasis added).

to deny the application of more precise medical practices (assuming that precision translates to quality of care and outcome), given that the practical enterprise is concerned with more health, not more understanding.

Given the assumption that some domains are amenable to machine decision-making without loss, we might ask when machines *should* be employed to do something for us, and the regulatory task becomes more apparent. If we are concerned with outcomes—as opposed to the benefits of deliberation,[15] or the retention of human autonomy (as in liberal freedom)—and if we can show that machines make decisions in a less biased and more efficient fashion than humans would, the decision to delegate is easy. Where the primary concern is efficiency or the commercial imperative, we seem to attend to the latter requirement of reduced bias only after our machines are sent out into the wild and their inadequacies become apparent. This method of trial and error is employed by Silicon Valley entrepreneurs (the 'High Priests of Silicon Valley'),[16] and naturally contributes to the perception of automated decision-making as unreliable and unfair. However, in the case of machine decision-making in medicine, we have the use of long-standing institutional and ethical frameworks directed towards the attainment of better patient outcomes as the primary objective, and this provides at least two benefits, in theory: an expert to oversee and review a machine decision; and a system that subjects novel medical practices, devices, therapies, and medications to scrutinous validation to ensure that no harm comes about as a result of changes in practice.

---

[15] For example, Frank Pasquale contends that the expression of law in language is necessary for a 'just and accountable social order', so the process of 'doing' law is as important as the outcome: see Frank Pasquale, 'A Rule of Persons, Not Machines: The Limits of Legal Automation' (2019) 87 George Washington Law Review 1.

[16] Karen Yeung and Martin Lodge, 'Algorithmic Regulation: An Introduction' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019) 2. See also Jonathan Taplin, *Move Fast and Break Things: How Facebook, Google and Amazon Have Cornered Culture and What It Means for All of Us* (Macmillan 2017).

This scenario presents an ethical and moral quandary: under what circumstances is it permissible to give up human understanding of (and hence authority over) decisions—and effectively cede the cognitive primacy we have enjoyed from time immemorial—in exchange for the attainment of a more desirable outcome? Evidently, this question cannot be answered in the abstract, and I propose that we can only answer it with a proper understanding of the technology we purport to install in such a role; indeed, merely adverting to unspecified 'complexity' as a reason to step aside leaves the premise rather vague. The question regulators need to ask—how can we identify domains in which inexplicable machine decision-making has a justifiable benefit and should be permitted?—cannot come before we enjoy a widespread, consistent understanding of artificial intelligence and its related concepts, since non-human decision-making is a technological novelty with no immediate precedent. We need to know exactly what we are talking about, but unfortunately, we often don't. **This is the interdisciplinary gap I propose to contribute to.**

THESIS STRUCTURE

In this thesis, I assume the inevitability of increasing reliance on machine decision-making, and establish the utility of close scrutiny of artificially-intelligent technology for the purpose of contributing to regulatory understanding of those technologies and machine decision-making. Crucially, machine decision-making presents novel regulatory and ethical issues that warrant departure from the standard regulatory approach of technology-neutral, problem-oriented analysis and necessitate a specific technological focus. I will discuss the kinds of legal problems that we must provide normative answers for and in relation to which doctrinal approaches are inadequate; and give evidence of different legal approaches to discussing the problems that artificial intelligence purportedly presents. As I will show, defining artificial intelligence is difficult and not particularly useful, although

I use that term generally to describe the subdiscipline of computer science and related concepts of technological complexity, intelligence, and machine decision-making; I refer to 'artificially-intelligent technology'—because it connotes a more appropriate level of generality than artificial intelligence (which instead connotes a specific class of technological artifact)—to describe machines that perform intelligent functions in the sense that their functions would otherwise require human expertise.[17] This is an admittedly inelegant solution, but as discussed below, the breadth of intelligent technology and disagreement about the meaning of underlying concepts (such as intelligence) renders high-level discussion unavoidably imprecise.

I proceed to clarify properties, make relevant distinctions, resolve terminological ambiguity, and incorporate environmental factors in a discussion of artificial intelligence as it manifests in **clinical medicine**. I intend that my observations be generalisable beyond the immediate medical context, but focus on medicine specifically because **(a)** specificity is necessary to avoid ambiguity; and **(b)** medicine has a long history of engagement with machine decision-making and allows me to expound several important principles that are of general significance. In this way, I show that although much technology regulation scholarship legitimately begins with the law, in the case of artificial intelligence, such an approach is ill-advised. The long and evolving history of the field; the various manifestations of artificially-intelligent technology; the ambiguity that pervades technical, legal, policy, and popular discourse; domain-specific environmental factors; and the novelty of non-human decision-making mean that regulators need to pin down as many of the known unknowns as possible or risk speculation leading them astray.

[17] Edward Feigenbaum, 'Knowledge Engineering For The 1980s' (1982) 1 <https://exhibits.stanford.edu/feigenbaum/catalog/bf210yv9971> accessed 19 April 2021.

This thesis is thus divided into three parts: in the **first**, I justify my focus on technology as useful for regulatory decision-making on account of the novelty of artificially-intelligent technology and the regulatory effect it exerts, and describe the shortcomings of existing doctrinal approaches; in the **second**, I review the concept of artificial intelligence and its use in medicine with a view to elucidating as much practical knowledge as possible and as an example of how such an approach might be undertaken in other contexts; and in the **third**, I discuss relevant consequences and questions that should be asked in light of my findings.

At the heart of this thesis is the idea that machines should be understood as machines, not intelligent beings, and not as capable of possessing human faculties whatsoever.[18] At the same time, although the law conceives of knowledge as being exclusively human (and thus explicable), we can make meaningful use of decisions made by machines even though their reasoning is neither causal nor referable to human epistemological frameworks (and thus inexplicable). In facilitating machine decision-making, applying extant law may not be sufficient, and instead, it is necessary to examine technological artifacts in depth and in their context to understand what rules might usefully apply to them in characterising their role in society.

Thus, ultimately, my contention is this: if we are to make appropriate decisions about how artificially-intelligent technology interacts with society, what responsibilities we should be comfortable delegating, and how regulation might achieve those and other goals, we must closely examine the outcome of machine decision-making in the specific contexts of its use, with the understanding that we cannot proceed on the basis of broad, untested

[18] See generally Ronald J Brachman, '(AA)AI More than the Sum of Its Parts' (2006) 27 AI Magazine 19.

assumptions, which we should limit as much as possible. In this case, the law should not inform our approach to such novel technology; the technology should inform the law.

PART I: THE LAW

TECHNOLOGY REGULATION AND ARTIFICIAL INTELLIGENCE

The modern paradigmatic example of a technology that heralded a new body of law is the Internet, and numerous legal scholars railed against the suggestion that new technology necessarily meant a new legal subdiscipline.[19] Here, in the case of artificial intelligence, I propose no such thing, and in fact, I do not even contend that extant law is poorly adapted to dealing with artificially-intelligent technologies, although that seems quite likely. My contention is more subtle: we cannot assess the capacity of the law to deal with the widely observed societal implications of artificial intelligence unless we cut through the noise and look directly and specifically at technology *first*, rather than conjecture about how the law might apply to one particular conceptualisation of it. This thesis is a prescription for a different approach to novel technological complexity. In that sense, it may be conceived of as an attempt to both **(a)** alleviate some of the obscurity described by the first hurdle of the Collingridge dilemma (lack of information about a new technology's impact);[20] and **(b)** cast light along the technology axis of Bert-Jaap Koops' ten dimensions of technology regulation.[21]

With that understanding, there are two primary reasons (from the perspective of regulation) for lawyers to subject the concept of artificial intelligence to in-depth scrutiny. First, artificially-intelligent technology, by virtue of its complexity and consequent

[19] See, for example, Joseph H Sommer, 'Against Cyberlaw' (2000) 15 Berkeley Technology Law Journal 1145.

[20] David Collingridge, *The Social Control of Technology* (Pinter 1980).

[21] Bert-Jaap Koops, 'Ten Dimensions of Technology Regulation: Finding Your Bearings in the Research Space of an Emerging Discipline' in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Technology Regulation* (Wolf Legal Publishers 2010) 314–315.

unpredictability, is widely regarded to exert a regulatory effect on those with whom it shares an environment. Specificity is necessary to understand more about its effect and adapt regulatory practices to mitigating risk; and regulation of behaviour by technology—in the various senses of Roger Brownsword's 'West Coast' techno-regulation,[22] Lawrence Lessig's 'architecture' or 'code',[23] Karen Yeung's 'hypernudge',[24] Mireille Hildebrandt's 'data-driven architecture' and 'digital unconscious',[25] or Shoshana Zuboff's 'instrumentarianism'[26]—cannot be explained by looking first or only to the law. Second, artificially-intelligent technology is novel in the sense that the problems arising from its use do not have useful, complete, and relevant analogies offered by extant technology that might allow regulators to theorise in the abstract; actually, 'the power and sophistication of these machines has no historical antecedents'.[27] By comparison to the latter, consider the illustrative case of nanotechnology, in relation to which Moses observes that

> the risks presented are closely related to the risks associated with other chemical substances and methods of manufacture. … Whatever risk-avoidance mechanisms are prescribed for chemicals generally can be used for nano-materials …[28]

---

[22] Roger Brownsword, 'The Technologies of the 21st Century: Regulatory Challenge and Regulatory Opportunity', *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008); Roger Brownsword, 'Code, Control, and Choice: Why East Is East and West Is West' (2005) 25 Legal Studies 1.

[23] Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books 1999).

[24] Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118.

[25] See generally Hildebrandt (n 3).

[26] Zuboff (n 3) 8ff.

[27] Yeung and Lodge (n 16) 12.

[28] Lyria Bennett Moses, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5 Law, Innovation and Technology 1, 16. See also Lyria Bennett Moses, 'Regulating beyond Nanotechnology: Do Nano-Specific Problems Require Nano-Specific Solutions?', *2010 IEEE International Symposium on Technology and Society* (2010).

We have never before had machines that can make decisions in ways we cannot understand—machines that operate by using knowledge we did not have when we prescribed their task. For these reasons, it is necessary to look closely at manifestations of artificially-intelligent technology in a specific context to make meaningful observations about its properties, lest we apply analogies that are unsuitable or misleading.

ARTIFICIAL INTELLIGENCE HAS A REGULATORY EFFECT

Regulation was traditionally conceived as emanating exclusively from the state in the form of explicit rules ('command and control') or the softer pursuit of policy objectives,[29] and '[l]awyers have tended to focus on the narrower definitions' of regulation,[30] but development of the scholarly field of regulation has led to recognition of the fact that other sources of influence are clearly active in society. At least insofar as regulation is directed towards achieving a particular goal, the source or form of regulation is less relevant than the overall behavioural outcome, which is dependent not only on the state, but on the 'complexity of interactions between actors in society';[31] that is, regulation is 'the product of interactions, not of the exercise of the formal, constitutionally recognized authority of the government'.[32]

[29] Robert Baldwin, Martin Cave and Martin Lodge, *Understanding Regulation: Theory, Strategy, and Practice* (Oxford University Press 2011) 3.

[30] Bronwen Morgan and Karen Yeung, *An Introduction to Law and Regulation: Text and Materials* (Cambridge University Press 2007) 4.

[31] Julia Black, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103, 106.

[32] ibid 110.

Regulation scholarship attesting to the emergence of the post-regulatory state (especially in the 21st century) and decentralisation of regulation,[33] the cybernetics conceptualisations of regulatory environments (in particular, information-gathering to identify system changes and misalignment),[34] and of course, the notion of good governance and small government,[35] affirm the importance of understanding the network of actors that participate in and influence the effect of regulation beyond the state. Indeed, Michel Foucault's discussion of governmentality—'conduct of conduct'—reinforces the idea that relevant influence emerges from all actors, not just the state, but such an abstract discussion is antithetical to the purposes I have laid out here. In any case, while it may be useful to confine analyses of regulation (of any modality) to only regulation that emanates from the state if we want to understand or critique the legitimacy of its operations,[36] it is clear that for many purposes, including achieving an in-depth understanding of a regulatory environment, such a limited definition of regulation is arbitrary; Julia Black's highly-cited definition of regulation,[37] and Karen Yeung's more recent definition of

[33] See especially Black, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (n 31); Colin Scott, 'Regulation in the Age of Governance: The Rise of the Post-Regulatory State' in Jacint Jordana and David Levi-Faur (eds), *The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance* (Edward Elgar 2004).

[34] See Brownsword, 'The Technologies of the 21st Century: Regulatory Challenge and Regulatory Opportunity' (n 22) 14–16; Morgan and Yeung (n 30) 3. See generally Andrew Murray and Colin Scott, 'Controlling the New Media: Hybrid Responses to New Forms of Power' (2002) 65 The Modern Law Review 491.

[35] See generally David Osborne and Ted Gaebler, *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* (Addison-Wesley 1992); Jon Pierre and B Guy Peters, *Governing Complex Societies: Trajectories and Scenarios* (Palgrave Macmillan 2005) ch 6; Nikolas Rose, *Powers of Freedom: Reframing Political Thought* (Cambridge University Press 1999) ch 1.

[36] Roger Brownsword intentionally takes this approach: Brownsword, 'The Technologies of the 21st Century: Regulatory Challenge and Regulatory Opportunity' (n 22).

[37] Julia Black, 'Critical Reflections on Regulation' (Centre for Analysis of Risk and Regulation, London School of Economics and Political Science 2002) CARR Discussion Paper 4 20 <http://eprints.lse.ac.uk/35985> accessed 9 August 2021; Julia Black, Martin Lodge and Mark Thatcher (eds), *Regulatory Innovation: A Comparative Analysis* (Edward Elgar 2005) 11.

'algorithmic regulation',[38] both recognise this breadth. The modern world is awash with non-state 'architectural' regulation, and many scholars have built on Lawrence Lessig's classic regulatory modality taxonomy to incorporate and characterise the behavioural influence of machine decision-making on regulatees.[39]

Given this broad understanding of regulation, it makes sense to examine the regulatory effect of artificially-intelligent technology under the auspices of regulation scholarship, and there are many scholarly contributions (the most influential of which are mentioned above) that attempt to map and understand its effect. The most important and wide-ranging regulatory effect, in my view, is the capacity for artificially-intelligent technology to exert a subliminal influence on human decision-making, or similarly, to provide information that implies or asserts (by design) that the environment or available options are more limited than they are.[40] Probably the most prominent and comprehensive discussion of this effect is by Shoshana Zuboff,[41] who warns of the 'instrumentarian power' exercised by smart technology that predicts and guides behaviour, engendering

> a population that ebbs and flows in perfect frictionless confluence, shaped by the means of behavioural modification that elude our awareness and thus can neither be mourned nor resisted.[42]

There are also more practical (or at least, ostensibly more focused) lenses through which to view this regulatory effect; one is offered by Lee Bygrave in his reclamation of

[38] Karen Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2018) 12 Regulation & Governance 505.

[39] Lessig (n 23).

[40] One classic example of bias guiding decision-making in a limiting or coercive way is predictive policing: see Kristian Lum and William Isaac, 'To Predict and Serve?' (2016) 13 Significance 14.

[41] But for an interesting earlier take, see also Hildebrandt (n 3) 9.

[42] Zuboff (n 3) 410–411 (and see generally chs 10, 13, and 14).

the concept of 'cognitive sovereignty' originally conceived by Ulrich Beck.[43] Through the lens of cognitive sovereignty, Bygrave analyses the utility of existing data protection rights in attempting to protect

> a human being's ability and entitlement to comprehend with a reasonable degree of accuracy their environs and their place therein, particularly the implications these hold for their exercise of choice.[44]

This entitlement, in Bygrave's estimation, is undermined by society's increasing reliance on machines as 'external knowledge producers' to evaluate risk and make decisions; cognitive sovereignty is compromised when a person does not have equal access to knowledge that forms the basis of decisions affecting their environment. Bygrave says that regulators recognise this interest in various forms—such as 'autonomy' or 'privacy', which terms are comparatively 'nebulous and pregnant with definitional variation'[45]—and consequently data protection norms are

> rooted in a concern to ensure that humans are able to participate in, shape and retain responsibility for decisions that significantly impact upon other humans, and that humans accordingly maintain the primary role in 'constituting' themselves.[46]

Thus, not only is the regulatory effect exerted by artificially-intelligent technology closely related to privacy and data protection (an uncontentious observation), but more generally, it threatens to militate against the legitimacy of public and political decision-making and

[43] Lee A Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' (University of Oslo Faculty of Law 2020) Research Paper 2020–35 <https://papers.ssrn.com/abstract=3721118> accessed 28 July 2021.

[44] ibid 9.

[45] ibid 13.

[46] ibid 16.

pervade the exercise of even small decisions made in relation to day-to-day activities.[47] For these reasons, it is necessary for regulators to closely examine the effect of machine decision-making in the environment, rather than more abstractly as a legal problem to be solved; only from that vantage point can a balance between autonomy and efficiency be struck.

ARTIFICIAL INTELLIGENCE IS NOVEL AND HAS NO ANTECEDENT

Sufficiently complex artificially-intelligent technology, by virtue of that complexity and its capacity for independent, intelligent decision-making, has no useful antecedent that properly captures and accounts for the rights and interests its use affects. Attempts to describe it in terms of human faculties can mislead, and instead, it is necessary to examine specific technologies to determine the extent to which the law might be applied.

Lawyers, as members of the general public, have a relatively poor understanding of artificial intelligence that arises in part (1) from the vagueness of the term; (2) from the technical sophistication of artifacts that are described as artificial intelligence; and (3) from our tendency, motivated by a desire for understanding, or as a product of human cognition,[48] to describe technological capabilities in anthropocentric terms—as machines that 'think', 'intuit', 'deliberate', or are 'autonomous'—or by reference to mythology and 'magic'.[49] The consequence of this technologically reductive treatment is that we misapprehend the features of artificial intelligence that commend it to us as a novel

[47] John Danaher, 'The Ethics of Algorithmic Outsourcing in Everyday Life' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019); John Danaher, 'The Threat of Algocracy: Reality, Resistance and Accommodation' (2016) 29 Philosophy & Technology 245.

[48] Ryan Calo, 'Robotics and the Lessons of Cyberlaw' (2015) 103 California Law Review 513, 545–546.

[49] See generally Alexander Campolo and Kate Crawford, 'Enchanted Determinism: Power without Responsibility in Artificial Intelligence' (2020) 6 Engaging Science, Technology, and Society 1.

technological concept deserving of independent regulatory scrutiny,[50] rather than just another iteration of existing technology. In fact, artificial intelligence is different to existing technology and prevailing perceptions in counter-intuitive, contradictory, and important ways; and it is neither human nor mere tool.[51] Consequently, the extent to which existing legal rules do or should apply to artificial intelligence is not actually very clear.[52] Thus, it merits closer attention.

A lot is said about identifying legitimate regulatory targets, and whether a particular technology is a good candidate for investigation by legal scholars; this is often because new technology is exciting but not novel in the sense that it throws up issues the law has never dealt with before.[53] Certainly, we want to 'avoid rehashing doctrinal and regulatory debates for each [new] technology' that emerges;[54] as Judge Frank Easterbrook famously opined in the case of cyberlaw, peering through the lens of technology is a rather roundabout way of trying to understand the law, when in fact, 'the best way to learn the law applicable to specialized endeavors is to study general rules'.[55] In the case of artificial

[50] On the contrary, as I discuss below, artificial intelligence is not really an identifiable technology, but it is a unifying concept, and it makes sense to refer to artificial intelligence in order to focus regulatory analysis of the actors, things, and relationships that surround it. See generally Brownsword, 'Code, Control, and Choice: Why East Is East and West Is West' (n 22).

[51] People experience difficulty assigning robots to an ontological category on account of the simultaneous presentation of their animate and inanimate properties: see Peter H Kahn and others, 'The New Ontological Category Hypothesis in Human-Robot Interaction', *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2011) <https://doi.org/10.1145/1957656.1957710> accessed 5 August 2021.

[52] For example, when might an artificially-intelligent machine warrant the imposition of responsibilities associated with legal personhood? When should decision-makers acting in reliance on machine recommendations be liable for their actions?

[53] See generally Moses, 'How to Think about Law, Regulation and Technology' (n 28).

[54] Ignacio N Cofone, 'Servers and Waiters: What Matters in the Law of A. I.' (2018) 21 Stanford Technology Law Review 167, 197.

[55] Frank H Easterbrook, 'Cyberspace and the Law of the Horse' (1996) 1996 University of Chicago Legal Forum 207, 207.

intelligence, though, I contend that this situation is one that Roger Brownsword would describe as 'normative' and 'productive disconnection',[56] and further inquiry is warranted.

In one sense, we are simply faced with smarter computers that exhibit more nuance in their outputs, and as a consequence, they can perform more tasks to an acceptable standard; they are still machines, and they are still conceivably products, software, or tools that fall within existing regulatory regimes. However, what accompanies competence in a certain domain is increasing (and eventually, exclusive) authority to make decisions within it, and thus to decide what is best in an undeniable exercise of *judgement*. Where once the simplicity of our computers rendered their decision-making capacity completely deterministic, now we are faced with decisions in relation to which we are not only unable to conceive of the breadth of reasoning underlying a decision; but we cannot even predict the outcome (even though, technically, our machines are just as deterministic as ever, albeit far more complex).

Faced with the prospect that valid decision-making can be done without human beings, it seems reasonable to observe both that the 'value compact'[57] underlying the various regulatory schemes governing machines do not account for their new authority; but also, that it does not really answer policy questions we are posing: namely, when is it okay for machines to make decisions that we would otherwise make ourselves? It may be prudent to avoid 're-inventing the intellectual wheel' every time we are faced with something new,[58] but here we cannot be sure, merely by referring to artificial intelligence, whether relevant legal analogies are suitable at all. In other words, this is not a case of

[56] Roger Brownsword, 'The Challenge of Regulatory Connection', *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008).

[57] ibid 166.

[58] Miriam C Buiten, 'Towards Intelligent Regulation of Artificial Intelligence' (2019) 10 European Journal of Risk Regulation 41, 48.

learning about the 'law of the horse' by examining the law of contract.[59] Rather, artificial intelligence is a new technology that (a) betokens a diverse range of behavioural influences (mentioned above) that are important to understand implicitly; and (b) has a range of 'essential qualities' that cannot be ascertained by armchair reference to properties that pepper the discourse of doctrinal legal scholarship.[60] At the very least, a 'working definition of what it means for technology to act, as opposed to inform, is of particular interest to legal analysis'.[61] It may turn out that the unique place of artificially-intelligent technology as somewhere between tool and person, and the function of providing independent judgement, pose problems that we can solve within the context of existing regimes. However, we can't say that for sure until we have a consistent, unambiguous framework for talking about it, and to get that, we need to start from the technology.

CONCLUSION

To understand artificial intelligence, machine decision-making, and the regulatory effect of those technologies, general legal analyses are not appropriate because they tend to anthropomorphise or generalise observations about the technological artifacts they discuss. A better approach is to observe the evolution of a particular technology, in the context of its use, and use that analysis to inform a regulatory approach to mitigating identified harms.

Following these legitimating observations about the properties of artificially-intelligent technology which make it ripe for technological inquiry, it is important to

[59] Easterbrook (n 55).

[60] For the idea that essential qualities distinguish some new technologies (cyberlaw and robotics) from prior technologies and justify a distinct conversation about them, see generally Calo (n 48).

[61] ibid 531.

evidence the basis on which I suggest that existing approaches are inadequate. In short, this is because they engage through the lens of legal doctrine with a technology about which they make assumptions that permit generalisations that limit the utility of their approach. Interestingly, a common objection to technology-specific regulation or bodies of law delimited by the technology they govern is that such approaches rely on distinctions that are arbitrary, creating 'bad taxonomy and bad legal analysis',[62] such as in the case of 'separating telephone harassment laws from computer harassment laws'.[63] This approach cements ephemeral beliefs that obtain from the novelty of the technology or the epoch, rather than attending to the societal conditions that mediate problems of technology and which are 'often well served by traditional legal devices'.[64] However, the disruptive nature of artificially-intelligent technology—and specifically its capacity for non-human, inscrutable decision-making—means that the law's reliance on analogies and history has a similar presumptuous effect: it tends to impute properties such as intent, autonomy, and even intelligence, in an attempt to explain technological complexity in human terms. Since a large part of our intuitive discomfort with machine decision-making arises from the perceived uncompromising, emotionless, *unhuman* nature of machines, this conflation is unhelpful.

[62] Sommer (n 19) 1147–1148.

[63] Moses, 'How to Think about Law, Regulation and Technology' (n 28) 14.

[64] Sommer (n 19) 1148.

A FLAWED LEGAL APPROACH TO ARTIFICIAL INTELLIGENCE

> A.I. is already embedded in our everyday lives: It influences which streets we walk down, which clothes we buy, which articles we read, who we date and where and how we choose to live. It is ubiquitous, yet it remains obscured, invoked all too often as an otherworldly, almost godlike invention, rather than the product of an iterative series of mathematical equations.[65]
>
> — Linda Kinstler, for *The New York Times*

In our collective haste to understand just what is going on with artificial intelligence, we identify perceived similarities or differences between human beings and artificially-intelligent technology, or craft hypothetical scenarios that are reminiscent of the 'lust to define the future' that Sommer complained of in relation to the advent of cyberlaw,[66] and allow the analogies and metaphorical language that emerge to guide and constrain our discussion of consequent legal ramifications, with deleterious effect. This approach is neither universal nor adopted only by legal scholars, but since I am concerned with detailing a prescriptive path for lawyers, errors in the approach of legal scholars are what I discuss here.[67]

The further we abstract, in our discussion of artificial intelligence, from computational techniques or technological artifacts—machine learning, evolutionary

[65] Linda Kinstler, 'Opinion: Can Silicon Valley Find God?' *The New York Times* (16 July 2021) <https://www.nytimes.com/interactive/2021/07/16/opinion/ai-ethics-religion.html> accessed 27 July 2021.

[66] Sommer (n 19) 1147.

[67] For authorship that provides a good contrast with the tendencies I detail here, see Harry Surden, 'Artificial Intelligence and Law: An Overview' (2019) 35 Georgia State University Law Review 1305.

algorithms, autonomous vehicles—the more we must rely on analogies and reductive generalisations that cloak technical ineptitude in universality. This may suffice for rhetorical purposes, but if we want to identify and evaluate the consequences of technological change for society, a more fine-tuned analysis is necessary. Technical complexity militates against meaningful determination of reasonable expectations of technology and its manufacturers and operators. Understanding is the difference between predetermined outcomes and immutable social architectures—*inevitability*—on one hand; and options, debate, and change—*possibility*—on the other. In the case of the nebulous technology that is described as artificial intelligence, upon which we confer *de facto* decision-making authority in matters from sentencing offenders,[68] to establishing voting preferences,[69] to determining the 'quality' of teachers,[70] we need to know: what are its limitations, what are its strengths, when should we rely on it, and when should we reserve decision-making authority for ourselves? Obviously, these questions cannot be answered in the abstract as a general rule for all artificially-intelligent technology. Analyses that do not delve into the complexity of the technology they purport to analyse should thus be treated with suspicion, lest they convince us to take as given something that is not—to treat as certain a mere prediction.

Campolo and Crawford observe that in the case of deep learning, a computational technique which is incompletely understood, technically complex, and inconsistent with

---

[68] Julia Angwin and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 12 April 2021. Note that, presently, in the case of sentencing, machine recommendations constitute only one component of the ultimate judicial determination.

[69] Bartlett (n 2) 20.

[70] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2017) 3–7.

human intuition,[71] the use of metaphorical language 'results in myth-making about deep learning's transcendent, superhuman capacities, especially when it is applied in social settings',[72] with the effect that deep learning is situated 'outside of understanding, outside of regulation, [and] outside of responsibility'.[73] This obscurity manifests in three major forms that I exemplify below: **future-making**, **anthropomorphisation**, and **terminological ambiguity**.

FUTURE-MAKING

Used here, future-making is the tendency to hypothesise about the effect and capacity of artificially-intelligent technology to give substance to a doctrinal legal discussion; such approaches

> involve significant, sometimes unsupported, assumptions about where the technology is headed. That speculative discussion often distracts from the important, but perhaps less exotic, law and policy issues actually raised by AI technology today.[74]

For example, a 2017 paper by Yavar Bathaee takes as a starting point the inexplicability of various hypothetical computer programs to advance his contention that extant doctrines of intent and causation (in the United States), which are 'focused on human conduct', are rendered dysfunctional by the 'black box' nature of artificial intelligence.[75] Although

[71] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1, 2.

[72] Campolo and Crawford (n 49) 1.

[73] ibid 9.

[74] Surden (n 67) 1306.

[75] Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2017) 31 Harvard Journal of Law & Technology 889, 891. For another legal analysis that is similarly speculative, see generally A Michael Froomkin, Ian Kerr and Joelle Pineau, 'When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning' (2019) 61 Arizona Law Review 33.

Bathaee suggests that this precludes the imposition of liability in tort upon designers for the actions of their machines, it is worth pointing out that courts readily infer the intent of human beings from extrinsic evidence despite the 'black box' nature of the brain, so mere complexity cannot be the sole cause of this purportedly confounding legal effect.

Do we have artificially-intelligent technology that is too complex to explain its reasoning fully in human terms? Yes.[76] But the many references to the 'intuitive' learning process of the technologies Bathaee describes—'akin to the way one "knows" how to ride a bike'[77]—are exhibits of the metaphorical language that Campolo and Crawford identify as misleading. This language conveys an independence that outpaces technology we have. Bathaee characterises artificial intelligence as inscrutable and capable of engaging in subtle breaches of the law that might pin down mere humans but which artificial intelligence neatly avoids: 'AI itself has no intent',[78] and it would be impermissibly onerous to impose strict liability on manufacturers for the actions of their machines.[79] Machine decision-making is here construed as so unpredictable as to be divine, and on this construction, it would certainly be unjust—sans guilty mind—to attribute responsibility for its consequences to the manufacturer.

Bathaee correctly observes that in the case of machine decisions that are too complex to be comprehensively reduced to language, simply insisting on more explainability may **(a)** hamper innovation and effectively limit technological innovation;

[76] Consider the complexity of DeepMind's AlphaZero: see David Silver and others, 'Mastering the Game of Go without Human Knowledge' (2017) 550 Nature 354; Will Knight, 'Alpha Zero's "Alien" Chess Shows the Power, and the Peculiarity, of AI' [2017] *MIT Technology Review* <https://www.technologyreview.com/2017/12/08/147199> accessed 15 August 2021.

[77] Bathaee (n 75) 902.

[78] ibid 921.

[79] ibid 931–932.

and **(b)** not always be possible, especially given the potentially unlimited complexity of such decision-making processes. However, the language used connotes impregnable reasoning that is beyond human control, but the reality is far more tractable:

> Machines cannot do anything but execute programs developed by humans, even if those programs enable the machine to reconfigure its program in view of specified machine-readable tasks, and even if humans may develop programs that build new programs.[80]

Thus, the question is less one of how to reconcile doctrinal shortcomings with 'superhuman' complexity, and more a question of what role should be ascribed to artificially-intelligent technology in the contexts in which it is used: we might just as easily characterise the above example as a malfunctioning product as we would an autonomous intelligence whose actions are beyond the control of its designer. In this case, assumptions mislead, and specific examples of extant technological artifacts necessarily ground the discussion in the present.

ANTHROPOMORPHISATION

Anthropomorphisation is the tendency to conflate complexity and technological sophistication with elements of humanness; indeed, we tend to conceive of intelligence as a faculty possessed exclusively by human beings, and characterise intelligent behaviour as human in nature.[81] Bathaee, for example, describes artificial intelligence as 'capable of making decisions experientially or intuitively like humans'.[82] This tendency to humanise

[80] Mireille Hildebrandt, 'The Artificial Intelligence of European Union Law' (2020) 21 German Law Journal 74, 77.

[81] In fact, intelligence is notoriously difficult to characterise and clearly does not reside exclusively in human beings. See generally Dagmar Monett and others, 'Special Issue "On Defining Artificial Intelligence"—Commentaries and Author's Response' (2020) 11 Journal of Artificial General Intelligence 1.

[82] Bathaee (n 75) 891.

the technological subject of analysis has a similar misleading effect. In particular, it connotes a breadth that current technological artifacts do not have: our machines are competent in *narrow* domains, but they are data-crunching computer programs, and their complexity should not be construed as intuition. Metaphors that invoke notions of humanity obscure the reality of artificially-intelligent machines because they do not clarify exactly how the various properties associated with human capabilities are supposed to apply to machines. Machines may 'learn' in some sense, but the concept of learning is extremely broad, and it is misleading to imply that machines can simply be placed in a new environment and induce rules from observation like humans. Some scholars recognise this difficulty and attempt to distinguish between human and machine by reference to 'unpredictability' or some other archetypal machine trait, but this creates an uneasy tension.

For example, Weston Kowert (in a discussion that mirrors Bathaee's on the applicability of legal foreseeability to machines) asserts that 'artificial intelligence software is not unlike the brain of a human child—ready to be molded and shaped by its experiences'.[83] Readers are left with the understanding that a machine learns like a human being, but unlike a human being, a machine 'will teach itself how to solve obstacles in ways that are *unpredictable*'.[84] If artificial intelligence learns like human beings, just why is its learning beyond the realm of legal foreseeability? The law deals quite well with the largely unpredictable nature of human behaviour. Evidently, this characterisation omits some crucial technological aspect that legitimates differentiation between human and machine.

[83] Weston Kowert, 'The Foreseeability of Human-Artificial Intelligence Interactions' (2017) 96 Texas Law Review 181, 183.

[84] ibid (emphasis added).

The metaphorical nature of language means that, to some extent, this tendency is unavoidable; 'machine learning' is a fairly specific, universally accepted term for a range of computational techniques that facilitate increasing accuracy and specificity with exposure to more information. Nonetheless, unexamined use of such terms implies that machines are just doing human things, but better. In actuality, many high-profile issues with artificial intelligence arise from *difference* between humans and machines,[85] and where those differences are subtle, they escape interrogation: for example, sophisticated machine pattern recognition is acausal, and thus although it ostensibly mirrors the human process of learning patterns by acquiring experience, predictions are based on likelihood, not the causal epistemologies on which human logic and rationality are premised. Similarly, some artificially-intelligent technology is described as 'autonomous', which supports the conceptualisation of the technology as beyond human control—free like a human being to make its own decisions.[86] Actually, though, even the most sophisticated deep neural network is unconsciously pursuing a predefined objective, whether it has been imbued with the capacity to respond dynamically to its environment or not.

The use of anthropomorphic metaphor contributes to the 'myth-making' Campolo and Crawford identify because it imports analogies that fit only imperfectly to the technological artifacts it purports to describe. By comparison, referring to specific technological artifacts, in the context of their use, allows us to describe their properties

[85] See, for example, Roger Clarke, 'Regulatory Alternatives for AI' (2019) 35 Computer Law & Security Review 398; Jason Millar and Ian R Kerr, 'Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots' (University of Ottawa 2013) SSRN Scholarly Paper <https://papers.ssrn.com/abstract=2234645> accessed 27 January 2022. See also Andrea Guzman, 'Ontological Boundaries between Humans and Computers and the Implications for Human-Machine Communication' (2020) 1 Human-Machine Communication.

[86] For a good and comprehensive discussion of autonomy (albeit in relation to systems generally), see Hussein A Abbass, Jason Scholz and Darryn J Reid (eds), *Foundations of Trusted Autonomy*, vol 117 (Springer 2018).

with clarifying specificity, which facilitates greater regulatory understanding. This tendency should be avoided.

TERMINOLOGICAL AMBIGUITY

By far the most common source of obscurity in legal discussion of artificial intelligence that I have observed is the tendency to blow past technical definitional tasks to get to the meaty legal issues that seem to lie ahead. Some scholars do not bother to define artificial intelligence, and as I will show, that task is difficult and unhelpful. But the absence of unifying conceptual properties that support a clear definition of the topic of discussion is evidence, in my view, that generalities are unlikely to be broadly applicable; nonetheless, *general* discussions populate legal and medical literature and popular culture. Concomitantly, the use of vague or undefined terms bespeaks a lack of familiarity with the technical subject matter that undermines the authority of the author to draw useful conclusions derived from that source.

Consider the term 'algorithm': it's ubiquitous. We tend to associate algorithms with artificial intelligence; algorithms make our computers smart. Some definitions exist: an algorithm is just a finite set of steps;[87] in mathematics, 'an algorithm is a procedure for computing a function';[88] in other cases, technical discussion results in no definitional agreement.[89] In legal discourse, the term is a stand-in for machines with some capacity for automated decision-making: perhaps that decision-making is construed as reductive;[90] or

[87] See, for example, Eric Dietrich, 'Algorithm' in Robert Andrew Wilson and Frank C Keil (eds), *The MIT Encyclopedia of the Cognitive Sciences* (MIT Press 1999) 11.

[88] Hartley Rogers, *Theory of Recursive Functions and Effective Computability* (1st MIT Press paperback edn, MIT Press 1987) 1.

[89] Andreas Blass and Yuri Gurevich, 'Algorithms: A Quest for Absolute Definitions' (2004) 81 Bulletin of the European Association for Theoretical Computer Science 195.

[90] Pasquale (n 15).

intractably complex—a 'complex interaction of features, classifications, and targets, all of which draw upon a maze of hazy interactive and embedded values'.[91] Importantly, though, it is not used consistently, often not defined at all,[92] and is alluringly but misleadingly technical.

Similarly, descriptions of artificial intelligence that I document here are broad and vague. One paper dealing with the liability consequences of superior machine decision-making describes 'AI' as 'machine-learning-based diagnostic competence',[93] and 'machine learning' as 'simply algorithms designed to draw on data to answer questions'.[94] Another article that straddles the interdisciplinary divide between healthcare and law (and is written by legal scholars) does not give *any* description of the technology it purports to discuss.[95] Some approaches establish categories that appear useful but are really quite arbitrary: how can one decide whether a particular technological artifact can 'predict relationships that humans cannot predict', or merely 'replicate human capabilities, but faster, more reliably, and machine-readably'?[96] What fundamental differences underpin the distinction between these two oddly conceived classes? Other papers pay no heed to the evolution of artificially-intelligent technology whatsoever, apparently on the assumption that only

[91] Sonia K Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 UCLA Law Review 54, 63.

[92] See, for example, Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87 Fordham Law Review 1085. See especially pages 1089–1090.

[93] Froomkin, Kerr and Pineau (n 75) 35.

[94] ibid 44. See also the description of the 'basic components of an ML system' at 44–45.

[95] W Nicholson Price II, Sara Gerke and I Glenn Cohen, 'Potential Liability for Physicians Using Artificial Intelligence' (2019) 322 JAMA 1765.

[96] Andrew D Selbst, 'Negligence and AI's Human Users' (2020) 100 Boston University Law Review 1315, 1333. Note that this paper does not define artificial intelligence, either.

modern technologies require legal analysis. As I will show, history provides critical context that cannot be ignored if a comprehensive technological understanding is to be achieved.

Legal scholars are generally justified in focusing on legal issues and leaving the technical details to be resolved by specialists, but where the legal issues depend so crucially on understanding what a particular type of technology actually does—is the machine just 'another input to a [clinician's] reasoning process',[97] or a fully-fledged 'member of the care team'?[98]—this approach breaks down. Legal reasoning is analogical (a clear strength that permits the law to grapple with and incorporate new scenarios),[99] but analogy depends on the prior identification of relevant similarities, and this is no easy task; as Chopra and White reflect,

> perhaps the hardest problem in devising a theory of liability for artificial agents is deciding which body of analogous case law should be brought to bear in the case of artificial agents.[100]

This process of analogical reasoning cannot be performed in the abstract, since artificially-intelligent technology is novel, difficult to define, and manifests in myriad forms; generalisations are inadequate. Indeed, as Cofone notes in the context of his own taxonomic approach,

---

[97] I Glenn Cohen, 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?' (2020) 108 Georgetown Law Journal 1425, 1442.

[98] ibid 1447.

[99] See generally Emily Sherwin, 'A Defense of Analogical Reasoning in Law' (1999) 66 University of Chicago Law Review 1179.

[100] Samir Chopra and Laurence F White, 'Tort Liability for Artificial Agents', *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press 2011) 119.

> Each A.I. technology presents different levels of embodiment, emergence, and social valence. The real question is, therefore, not finding the proper analog for A.I. in general but finding the most appropriate analog for each A.I. agent.[101]

Not only does this definitional task provide clarity for the purpose of facilitating constructive discourse, but it is a necessary prerequisite to the legal task of determining which legal rules should apply. Ambiguity in the fundamental task of describing what we are talking about does not help regulators decide when and how machine decision-making should be permitted.

CONCLUSION

There is a clear need for specificity in the identification and discussion of artificially-intelligent technology, and although the construction and analysis of hypothetical or general scenarios have their place, they do not provide much insight for regulators who seek to understand the nuanced implications of the use of artificially-intelligent technology in the myriad contexts of its use. The tendency to hypothesise about the future, rely on reductive anthropomorphic metaphor, or avoid the necessary resolution of ambiguous technical terms all contribute to the obscurity that pervades popular and legal discussion of artificial intelligence. Regulators making decisions about the integration of artificially-intelligent technology in society obviously cannot rely on discourse that does not assess the regulatory environment in which such technologies are used, and doctrinal analysis cannot document those environments satisfactorily. In seeking to establish a useful framework for discussing any artificially-intelligent technology, in light of the novelty and regulatory effect that I have referred to above, it is necessary to dispense with legal analogy and engage in an in-depth discussion of a particular domain, its history, and the specific

[101] Cofone (n 54) 175.

technologies that are used to elucidate and substantiate principles that may properly inform regulatory decision-making. To that end, the second part of this thesis advances an increasingly-specific discussion of artificial intelligence as it manifests in clinical medicine.

PART II: ARTIFICIALLY-INTELLIGENT TECHNOLOGY IN CLINICAL MEDICINE

INTRODUCTION

In the first part of this thesis, I posed a question that regulators in different settings are increasingly being asked in one form or another and with various degrees of complexity: just where are the boundaries delimiting the legitimate scope of machine decision-making? As discussed, the novelty of artificially-intelligent technology means that any attempt to answer this question must proceed from an in-depth understanding of particular technological artifacts rather than by analogy drawn from extant bodies of law; and the regulatory effect of such technologies means they should be studied not in a vacuum but in the context of their use. Although artificial intelligence is the topic of much legal scholarship, most does not proceed in the fashion I have described and instead relies on generalisations that conflate technological artifacts, and assumptions that produce an incomplete or misleading picture of the technologies they discuss.

One solution to this obscurity is a detailed, context-specific examination of one type of artificially-intelligent technology with a focus on its history, relevant distinctions between subtypes, and surrounding environment to establish the form and content of an inquiry that goes to meaningfully answering the question I have stated above. This is, in other words, an effort to cut through the proverbial noise. To that end, as I have foreshadowed, the second part of this thesis is a detailed inquiry into the use of artificially-intelligent technology in medicine that advances in three increasingly specific stages and reveals important observations that go to the overarching question of machine decision-making. I start from the ground and work my way up to create a comprehensive picture of artificially-intelligent technology in medicine that is indispensable to a proper understanding of the ethical and legal considerations that underpin answers to that

question. This process necessarily precedes meaningful regulatory discussion and is intended to encourage an interdisciplinary approach to new technologies. An unavoidable consequence of this fine detail is that it can be difficult to appreciate how everything fits together, and I give regular indications of the relevance of my observations at each stage.

Two important themes—which legitimate this approach because they highlight nuance that is not apparent in more general legal discussion—should be noted at the outset. The **first** theme is the fundamental relevance of context and history: in the case of medicine, the fallibility of human clinical judgement means that physicians rely increasingly, in their delivery of care, on information evaluated and disseminated by machines. Thus, although modern discourse on automation and machine decision-making is centred around issues of bias and illegitimate influence which must be protected against, here the costs and benefits are not so clearly divided; and this becomes apparent only by considering the use of artificially-intelligent technology in a particular environment. The **second** theme is the value to be derived from recognising and distinguishing between subtypes of artificially-intelligent technology. In addition to the widespread tendency to refer simply to 'artificial intelligence' as a homogeneous class of technological artifacts, legal scholars commonly make cursory reference to what is cast as the 'first generation' of artificial intelligence, which had its heyday before the age of big data and is now neither novel nor perceived to pose problems requiring legal solutions. The contrast between these machines that reason with human knowledge (knowledge-based) and those that reason from data (data-driven)—which is usefully apparent in medicine—shows us exactly why modern machines are unpredictable and inscrutable, and hints at what this might mean for scientific endeavours and human technological progress generally.

Importantly, in relation to both themes, medicine is inherently complex, and involves the weighing of numerous factors that are omitted by the comparatively simplistic

medical decision-making of experts. This reinforces the view that incomprehensible complexity is sometimes a product of the field in which machine decision-making operates, and machine decision-making can make beneficial use of such complexity despite being impossible to comprehensively understand. In this connection, consider the following comment made by Selbst and Barocas in relation to the utility of explanations of machine decisions:

> Explanations of technical systems are necessary but not sufficient to achieve law and policy goals, most of which are concerned not with explanation for its own sake, but with ensuring that there is a way to evaluate the basis of decision-making against broader normative constraints such as antidiscrimination or due process. It is therefore important to ask how exactly people engage with those machine explanations in order to connect them to the normative questions of interest to law.[102]

In other words, explanations serve the more fundamental purpose of assuring that decisions are made in accordance with established values—such as Bygrave's cognitive sovereignty, discussed above—and this is worth keeping in mind.

In light of these themes, the three stages of my analysis are laid out below. Each of these three stages is intended to elucidate clarificatory observations that are not apparent in less detailed analyses and which are evidence that such scrutiny is warranted in constructing an understanding that informs responses to the introductory interrogatory. In short, each piece of insight that would not otherwise be gained is a testament to the utility of this approach.

First, I discuss artificial intelligence generally to establish the difficulty of working with that term. I note that definitions of artificial intelligence change over time, and that

[102] Selbst and Barocas (n 92) 1088.

both policy and research uses of the term are vague, sometimes contradictory, and generally unhelpful. I also briefly discuss use of the term in modern medical literature, and show that ambiguity persists even within a narrowed field. This section is intended to provide further support for my in-depth focus and show that legal scholarship is not alone in the difficulty it faces in discussing artificial intelligence.

Second, I give an overview of the history of artificial intelligence in medicine (which reflects the development of the field of artificial intelligence more generally). I note the historical distinction between artifacts which rely on knowledge and data, and describe various applications of artificially-intelligent technology in medicine, including the use of clinical decision support systems, which form the basis for much of my inquiry. To give environmental context to the discussion, I describe clinical judgement and recount the history of clinical intuition, including the emergence of evidence-based medicine and increasing reliance on machines to correct the biases inherent in clinical decision-making.

Third, after covering artificial intelligence generally, and introducing the various uses, context, and history of artificially-intelligent technology in medicine, I introduce clinical decision support systems in detail and describe their applications and the various taxonomies by which they may be categorised and better understood. In particular, I focus on the aforementioned division between knowledge and data, since this division permits detailed interrogation of the properties exhibited by artificially-intelligent technology and the basis of the reputed complexity and unpredictability that modern artificially-intelligent technology threatens. I discuss these categories in depth and give examples.

Finally, after conducting this in-depth review, in the third part of this thesis, I reflect on the kinds of normative legal questions that need to be answered and which are served by the specificity afforded by technology-driven inquiries such as the one I undertake here. These questions provide fertile ground for further research.

ARTIFICIAL INTELLIGENCE

> AI is hard to define anyway; most of the time it ends up being viewed as those things which the AI people are doing.[103]
>
> — Edward Feigenbaum, the 'father of expert systems'

There is no universal definition of artificial intelligence.[104] There are significant differences between the meaning of the term in the 20th century and the present, and in usage between artificial intelligence researchers, regulators, medical experts, and the general public. Resolution of this ambiguity serves little practical purpose. Nonetheless, it is useful to understand the component parts of the term and its field, insofar as they relate to this discussion. There are three things to note: **(1)** accepted definitions of artificial intelligence change over time; **(2)** policymakers and artificial intelligence researchers use similar definitions, but emphasise slightly different qualities of artificially-intelligent technology; and **(3)** many medical and artificial intelligence researchers **(a)** speak in broad terms about the field that belie its scope and history; **(b)** use specific terminology to avoid ambiguity; or **(c)** uncritically prescribe a definition that matches popular usage of the term. There is, overall, very little consensus.

[103] Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (A K Peters Ltd 2004) 326.

[104] Pei Wang, 'On Defining Artificial Intelligence' (2019) 10 Journal of Artificial General Intelligence 1, 1; Monett and others (n 81).

THE EVOLUTION OF DEFINITIONS OVER TIME

In the late 20th-century, it was generally thought that symbolic representation of knowledge was the key to intelligence,[105] and artificial intelligence researchers focused their efforts on developing symbolic reasoning programs, such as expert systems.[106] Today, we do not generally view the deterministic nature of symbolic artificial intelligence as constituting real intelligence at all. In fact, with each subsequent advance in artificial intelligence research, previously accepted indicators of intelligence are revealed to be totally inadequate; this is colloquially known as the 'AI effect', or 'Tesler's Theorem'.[107] For example, before IBM's Deep Blue beat chess grandmaster Garry Kasparov at chess in 1996, chess mastery was understood to require intelligence.[108] Afterwards, *The Economist* opined that chess mastery, 'once thought to be the *sine qua non* of intelligent behavior', was not indicative of machine intelligence at all.[109]

DEFINITIONS PROPOSED BY ARTIFICIAL INTELLIGENCE RESEARCHERS

Research papers that focus on specific computational techniques tend to dispense with the mostly academic task of defining artificial intelligence because they describe technical

---

[105] See, for example, Allen Newell and Herbert A Simon, 'Computer Science as Empirical Inquiry: Symbols and Search' (1976) 19 Communications of the ACM 113.

[106] Edward H Shortliffe and Bruce G Buchanan (eds), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison Wesley 1984) 3 <http://www.shortliffe.net> accessed 13 April 2021; Feigenbaum (n 17) 1.

[107] Michael Haenlein and Andreas Kaplan, 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) 61 California Management Review 5, 6; Douglas Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid* (Penguin Books 1980) 601; Larry Tesler, 'Adages and Coinages' (*Larry Tesler Consulting*) <http://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html> accessed 2 May 2021.

[108] McCorduck (n 103) 456.

[109] ibid 419, 433, 480–83.

features in relation to which a broad definition is superfluous. Textbooks and other theoretical discussions more commonly provide such overarching definitions. Peter Norvig and Stuart Russell, in *Artificial Intelligence: A Modern Approach*, describe four distinct approaches to defining artificial intelligence: 'thinking humanly', 'acting humanly', 'acting rationally', and 'thinking rationally'.[110] Other approaches are similarly focused on rationality of a machine's action.[111] These approaches, which import concepts from other disciplines to assess whether a machine can be said to be intelligent, are interesting exploratory frameworks, but because they have a taxonomic rather than outcome-based focus, have limited regulatory utility.

DEFINITIONS USED BY LAWMAKERS

Policy and legislative definitions are numerous and similarly vague.[112] Several representative examples are given here. In the United States, artificial intelligence is given an inclusive working definition in and for the purposes of s 238(g) of the *John S. McCain National Defense Authorization Act for Fiscal Year 2019*,[113] which describes an 'artificial system that performs tasks under varying and unpredictable circumstances without significant

[110] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Pearson Education 2016) 1–5.

[111] Nils Johan Nilsson, *Artificial Intelligence: A New Synthesis* (Morgan Kaufmann 1998) 1; David Poole, Alan Mackworth and Goebel Randy, *Computational Intelligence: A Logical Approach* (Oxford University Press 1998) 1; Stephen M Omohundro, 'The Basic AI Drives', *Proceedings of the First AGI Conference (Frontiers in Artificial Intelligence and Applications)* (2008) 1–2; Stuart Russell, 'Provably Beneficial Artificial Intelligence', *The Next Step: Exponential Life* (BBVA OpenMind 2017).

[112] Some policy documents refer to textbook definitions or do not define artificial intelligence at all. See Committee on Technology, 'Preparing for the Future of Artificial Intelligence' (National Science and Technology Council 2016) <https://obamawhitehouse.archives.gov> accessed 23 August 2021; Networking and Information Technology Research and Development Subcommittee, 'National Artificial Intelligence Research and Development Strategic Plan' (National Science and Technology Council 2016) <https://obamawhitehouse.archives.gov> accessed 23 August 2021.

[113] Pub L No. 115-232, 132 Stat 1636, 1695 (13 August 2018) (codified at 10 USC § 2358).

human oversight' and which 'solves tasks requiring human-like perception'.[114] In the United Kingdom, the House of Lords Select Committee on Artificial Intelligence's report, '*AI in the UK: ready, willing and able?*' recognises that '[t]here is no widely accepted definition of artificial intelligence',[115] but describes artificial intelligence as '[t]echnologies with the ability to perform tasks that would otherwise require human intelligence' which 'usually have the capacity to learn or adapt to new experiences or stimuli'.[116]

The European Commission and other European Union bodies have provided a range of broad definitions of artificial intelligence.[117] In some cases, ambiguity has led to rather curious proposals: the European Parliament adopted a resolution in 2017 suggesting that the Commission examine the possibility of classifying sophisticated robots as 'electronic persons'.[118] In 2018 the Commission defined artificial intelligence as 'systems

[114] Ibid. Other definitions are provided by other US agencies: see Defense Innovation Board, 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense (Supporting Document)' (United States Department of Defense 2019) 8–12 <https://innovation.defense.gov/recommendations> accessed 23 August 2021. Note that although the *John S. McCain National Defense Authorization Act* requires the Secretary of Defense to define artificial intelligence for the Department's internal purposes, no such definition has yet been established: see Congressional Research Service, 'Artificial Intelligence and National Security' (2020) Report R45178 1 <https://crsreports.congress.gov/product/pdf/R/R45178> accessed 23 August 2021.

[115] House of Lords Select Committee on Artificial Intelligence, 'AI in the UK: Ready, Willing and Able? Report of Session 2017-19' (House of Lords Select Committee on Artificial Intelligence 2018) 14.

[116] ibid. This definition was originally proposed by Department for Business, Energy & Industrial Strategy, 'Industrial Strategy: Building a Britain Fit for the Future' (Government of the United Kingdom 2018) Policy Paper 37 <https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future> accessed 23 August 2021.

[117] See European Parliamentary Research Service, 'European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies (European Added Value Assessment)' (European Parliament 2020) Study 12ff <https://doi.org/10.2861/94107> accessed 23 August 2021; European Parliamentary Research Service, 'Civil Liability Regime for Artificial Intelligence: European Added Value Assessment' (European Parliament 2020) Study 3 <https://doi.org/10.2861/737677> accessed 23 August 2021. See also Independent High-Level Expert Group on Artificial Intelligence, 'A Definition of AI: Main Capabilities and Disciplines' (European Commission 2019) <https://digital-strategy.ec.europa.eu/en/library> accessed 23 August 2021.

[118] European Parliament, 'European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics' (European Parliament 2017) P8_TA(2017)0051 <https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html> accessed 18 August 2021. See also Cofone (n 54).

that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals'.[119] The Commission acknowledged in its 2020 artificial intelligence white paper that any definition of artificial intelligence would need to be 'sufficiently flexible to accommodate technical progress'.[120] In April 2021, the European Commission distributed a proposal for an 'Artificial Intelligence Act',[121] which defines an 'artificial intelligence system' as 'software that is developed with one or more of the techniques and approaches listed in Annex I' to the Regulation (which the Commission can amend as necessary). Presently, the listed techniques are machine learning, logic- and knowledge-based, and statistical; this is probably the most specific approach yet.

Artificial intelligence researchers and policymakers tend towards similar broad definitions of artificial intelligence, though the former emphasise rational action, and the latter emphasise human-like intelligence and autonomy. Notably, policy definitions generally incorporate descriptions of artificial intelligence that make reference to the use of data and algorithms; this is likely the result of popular and widespread preoccupation with the perceived bias and unpredictability that modern AI exhibits.

[119] See European Commission, 'Artificial Intelligence for Europe' (2018) Communication COM(2018) 237 final 1; European Commission, 'Coordinated Plan on Artificial Intelligence' (2018) Communication COM(2018) 795 final 1.

[120] European Commission, 'On Artificial Intelligence — A European Approach to Excellence and Trust' (2020) White Paper COM(2020) 65 final 16. In its White Paper, the Commission also noted that the main elements of artificial intelligence are 'data' and 'algorithms', and that 'AI-based products can act autonomously by perceiving their environment and without following a pre-determined set of instructions'.

[121] See European Commission, 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' Proposal for a Regulation of the European Parliament and of the Council COM(2021) 206 final (Procedure 2021/0106/COD) <https://eur-lex.europa.eu/procedure/EN/2021_106> accessed 23 August 2021.

TERMINOLOGY IN MEDICAL LITERATURE

Even in the specific context of scholarly medical literature, references to artificial intelligence do not have a uniform meaning, notwithstanding the considerable and longstanding reliance of the sector on machines; indeed, 'the literature has been replete for 4 decades with studies that present computing systems demonstrating diagnostic accuracy that rivals the performance of expert clinicians'.[122] Mostly, there is little practical use for such a broad term, and when researchers do speak about artificial intelligence, they refer to the subfield of computer science generally (often in the course of an introductory review). Thus, researchers **(a)** speak in extremely broad terms;[123] **(b)** refer to specific computational techniques and have no need to speak of artificial intelligence generally;[124] or **(c)** prescribe definitions that reflect modern popular usage of the term which uncritically conflates artificial intelligence with machine learning techniques (and often deep learning specifically).[125] As an example of the latter, a 2019 paper in *Nature Medicine* simply describes artificial intelligence as the 'development of computer algorithms to perform tasks typically associated with human intelligence', and relies on a textbook on deep learning as its authority for that claim.[126] The most that can be said is that artificial intelligence in medicine is generally understood to be the use of computational techniques

[122] Edward H Shortliffe and Martin J Sepúlveda, 'Clinical Decision Support in the Era of Artificial Intelligence' (2018) 320 JAMA 2199, 2199.

[123] See, for example, Pavel Hamet and Johanne Tremblay, 'Artificial Intelligence in Medicine' (2017) 69 Metabolism S36; AN Ramesh and others, 'Artificial Intelligence in Medicine' (2004) 86 Annals of the Royal College of Surgeons of England 334.

[124] This is true of many biomedical research papers that describe specific computational techniques. See, for example, Andre Esteva and others, 'Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks' (2017) 542 Nature 115.

[125] See, for example, Chayakrit Krittanawong and others, 'Artificial Intelligence in Precision Cardiovascular Medicine' (2017) 69 Journal of the American College of Cardiology 2657.

[126] Jianxing He and others, 'The Practical Implementation of Artificial Intelligence Technologies in Medicine' (2019) 25 Nature Medicine 30.

to extract novel information from large datasets. A brief literature search including terms such as 'artificial intelligence' and 'medicine' will confirm this tendency.

CONCLUSION

The term 'artificial intelligence' describes a multidisciplinary, established field; and a practically limitless range of technological artifacts and software. Not much, apart from the vague notion that intelligence is emulating whatever humans do, seems to unite the field, and various attempts to construct overarching definitions reflect its disparate nature. Consequently, it is necessary to refer to specific subdomains and technological artifacts in order to make meaningful observations that might inform regulators evaluating the effects of machine decision-making. This brief review is intended to legitimate such a specific focus, and having established the ambiguity inherent in such a broad term, I now turn to medicine.

ARTIFICIAL INTELLIGENCE AND MEDICINE

A BRIEF HISTORY

Artificial intelligence in medicine has a rich, complex, and lengthy history, which may be briefly stated in summary form. From the mid- to late-20th century, with increases in computational efficiency and recognition of the mercurial unreliability of clinical judgement, the medical profession incorporated guidelines, statistical methods, and computer systems into clinical practice to systematise decision-making and standardise the quality of patient care.[127] Rapid technological advancement has led to increasing reliance on machines with numerous documented clinical benefits.

The first widely successful systems described as 'artificial intelligence' were used in medicine, because the (ostensibly scientific) profession admitted naturally of the uniformity that computers could ensure.[128] Early artificially-intelligent technology—knowledge-based systems—sought to emulate, and first replace but then merely complement, clinical decision-making.[129] Medical knowledge was extracted from medical experts by knowledge engineers, laboriously converted into a symbolic, computer-readable form, and fed in modular chunks into computer programs that applied rules of formal logic to make inferences from clinician input. In such a fashion, and in a variety of other, less complex ways—including simple linear regressions relating predictive antecedent

[127] For discussion of flowcharts as a method of standardisation, for example, see Peter Szolovits (ed), *Artificial Intelligence in Medicine* (Westview Press 1982) ch 1 <https://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch0.html> accessed 25 April 2021.

[128] See, for example, Vimla L Patel and others, 'The Coming of Age of Artificial Intelligence in Medicine' (2009) 46 Artificial Intelligence in Medicine 5, 6ff.

[129] For an example of a knowledge-based system that sought to effectively replace human clinical decision-making, see Randolph A Miller and FE Masarie Jr., 'The Demise of the "Greek Oracle" Model for Medical Diagnostic Systems' (1990) 29 Methods of Information in Medicine 1; Randolph A Miller, Harry E Pople Jr. and Jack D Myers, 'Internist-I, an Experimental Computer-Based Diagnostic Consultant for General Internal Medicine' (1982) 307 The New England Journal of Medicine 468.

variables to a diagnosis—clinical medicine became less reliant on the exercise of intuitive judgement by experienced physicians, and more reliant upon the corrective mathematical consistency of machines. This trend continued throughout the 1990s alongside the emergence of 'evidence-based medicine', which encouraged deference to tested, rather than traditional, medical practices.

From the mid- to late-2000s, 'technical advances and much faster hardware made it feasible to train neural networks with many layers of feature detectors on large data sets'.[130] The increasing availability of data to which such technical advances could be applied eventually gave rise to deep learning (discussed below), a subset of machine learning that had stagnated in the 1990s in the absence of such technological advances.[131] Since deep learning excels at pattern recognition, a central aspect of clinical medicine, the technology has myriad healthcare applications, and has been shown to be successful (by comparing performance to human experts) in numerous studies.[132]

The history of artificial intelligence in medicine (which reflects the wider field) thus reveals an important distinction between two forms of intelligence that I discuss in this thesis: that based on **knowledge** deduced and provided by humans; and that which is based on **data**. This distinction, though borne out in the literature, is difficult to identify on account of aforementioned terminological ambiguity which pervades even terms with more specific technical meanings that artificial intelligence; for example, machine learning, which is used erroneously as a modern synonym for artificial intelligence (and there is

[130] C David Naylor, 'On the Prospects for a (Deep) Learning Health Care System' (2018) 320 JAMA 1099.

[131] ibid.

[132] See eBox to ibid.

disagreement about the extent to which the two terms overlap), was coined in 1959,[133] and was entirely distinct from knowledge representation systems that were associated with artificial intelligence at the time.[134]

Ambiguous terminology, technical complexity, and a dynamic industry mean that not only is conceptual clarity difficult to achieve for the purpose of transference to different academic disciplines (such as law), but that regulators and policymakers, too, have difficulty understanding the subject matter and expressing their intention in sufficiently precise language; this is especially true for a field in which regulators are under significant pressure to avoiding hampering innovation.[135]

The history of artificial intelligence in medicine reflects the general advancement of the wider field, has a long history, and admits of good specificity in the discussion of specific technological artifacts; for these reasons, medicine forms the broad subject-matter of my inquiry. In the next two sections of this chapter, I discuss **(1)** broad applications of artificially-intelligent technology in medicine; and **(2)** the context and history of clinical medicine and clinical judgement.

---

[133] Arthur L Samuel, 'Some Studies in Machine Learning Using the Game of Checkers' (1959) 3 IBM Journal of Research and Development 210.

[134] John Durkin, 'History and Applications' in Cornelius T Leondes (ed), *Expert Systems: The Technology of Knowledge Management and Decision Making for the 21st Century* (Academic Press 2002) 8–9.

[135] On the difficulty faced by policymakers in regulating artificial intelligence, see, for example, Sankalp Bhatnagar and others, 'Mapping Intelligence: Requirements and Possibilities' in Vincent C Müller (ed), *Philosophy and Theory of Artificial Intelligence 2017* (Springer 2018) 118. See also Cristie Ford, *Innovation and the State: Finance, Regulation, and Justice* (Cambridge University Press 2017).

APPLICATIONS OF ARTIFICIALLY-INTELLIGENT TECHNOLOGY IN MEDICINE

Broadly, there are research,[136] administrative, and clinical applications of artificially-intelligent technology in medicine. Most artifacts either provide decision support or conduct analyses on the basis of which decisions are made, and such support ranges from clinical decision support systems, to hospital management support systems,[137] to retrospective analyses of financial data,[138] to real-time radiographic image interpretation, to give several examples. With the increasing availability of data since the turn of the century, healthcare providers have been better able to analyse aggregate patient information for business purposes,[139] and medicine has generally followed the wider technological trend from simply providing extant knowledge at convenient times to generating new, 'data-centric' information.[140] Many decision support systems have been integrated in clinical practice for long enough that they are no longer perceived as intelligent, and thus, as discussed, many applications of artificially-intelligent technology detailed in recent healthcare literature deal with bioinformatics, computer vision, and other applications of machine learning. There are clinical applications for many forms of artificially-intelligent technology used in biomedical research (indeed, that is ultimately the purpose of such research), but many typical ethical concerns with artificial intelligence—

---

[136] For a comprehensive discussion of various artificial intelligence applications in biomedical research, see Alessandro Blasimme and Effy Vayena, 'The Ethics of AI in Biomedical Research, Patient Care, and Public Health' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (2020) 704–8.

[137] Guisseppi A Forgionne and Rajiv Kohli, 'HMSS: A Management Support System for Concurrent Hospital Decision Making' (1996) 16 Decision Support Systems 209.

[138] Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016) 2.

[139] Kristin B DeGruy, 'Healthcare Applications of Knowledge Discovery in Databases' (2000) 14 Journal of Healthcare Information Management 59, 60. For an early discussion of the economic value of healthcare data, see Ellen M Harper, 'The Economic Value of Health Care Data' (2013) 37 Nursing Administration Quarterly 105.

[140] Blasimme and Vayena (n 136) 704; Geoffrey Hinton, 'Deep Learning—A Technology With the Potential to Transform Health Care' (2018) 320 JAMA 1101.

the quality, representativeness, and inherent biases of data used in developing machine learning algorithms[141]—relate to applications that have not been applied in clinical settings. This fact reflects another important trend: many artificial intelligence applications are studied, but few have been tested in randomised clinical trials,[142] only relatively few have been granted regulatory approval, and even fewer have entered widespread use in clinical practice,[143] though this is changing rapidly.[144]

Of the various subdomains of artificial intelligence in medicine, clinical decision support systems, which are computer systems that support clinical decision-making (but not all constitute artificially-intelligent technology), provide the most meaningful opportunity to assess the implications of use of artificially-intelligent technology in medicine, for several reasons: they are directly related to the patient's experience and the quality of their care; they have been used for a long time and consequent legal and ethical issues have been well canvassed; they are intended to influence clinical decision-making, and thus have prevailing relevance at a time when the influence of decision-making by

[141] Blasimme and Vayena (n 136) 705.

[142] Derek C Angus, 'Randomized Clinical Trials of Artificial Intelligence' (2020) 323 JAMA 1043.

[143] This is probably partially the result of clinicians' objection to the use of machines in decision-making, which has been a topic of discussion in the multidisciplinary field of clinical judgement for decades and is discussed later. See, for example, Ramesh and others (n 123) 337. It could also be the result of poor integration, though this has changed in the last decade as a result of US healthcare industry implementation incentives. See Adam Wright and others, 'Development and Evaluation of a Comprehensive Clinical Decision Support Taxonomy: Comparison of Front-End Tools in Commercial and Internally Developed Electronic Health Record Systems' (2011) 18 Journal of the American Medical Informatics Association 232. On low CDSS implementation levels in the US in the 2010s, see Steven R Simon and others, 'Physicians and Electronic Health Records: A Statewide Survey' (2007) 167 Archives of Internal Medicine 507.

[144] See Eric J Topol, 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25 Nature Medicine 44, 46 (Table 2); Sara Gerke, Timo Minssen and Glenn Cohen, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Adam Bohr and Kaveh Memarzadeh (eds), *Artificial Intelligence in Healthcare* (Academic Press 2020) 298–300; United States Food and Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (*FDA*, 12 January 2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 18 July 2021.

machines has piqued regulatory interest;[145] they are relatively well-defined so that a discussion of their risks and benefits has clear limits that avoids the ambiguity that besets discussion of artificial intelligence more generally; and their division into the two forms of artificial intelligence identified above, and referred to throughout—those based on knowledge and data—is relatively uncontentious and simple to establish.

The following should be noted from the preceding discussion. Artificial intelligence has no universal definition and describes a broad field and range of technologies. Historically, artificial intelligence was associated with knowledge representation, and in modern times, artificial intelligence is associated with machine learning and analysis of large datasets. Artificially-intelligent technology is used in medicine in biomedical research, hospital and healthcare administration, and in clinical practice. The breadth of its applications requires a specific focus in order to make meaningful observations, and it is useful to focus on clinical decision support systems to elucidate specific issues that go to the overarching question of machine decision-making. Before discussing clinical decision support systems, I provide an abridged, recent history of **clinical judgement**, since this is the context in which clinical decision support systems were introduced and are used.

[145] See, for example, European Commission, 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (n 121). Title II of the Act is dedicated to prohibited artificial intelligence practices, including 'practices that have a significant potential to manipulate persons through subliminal techniques beyond their consciousness'.

CLINICAL MEDICINE AND CLINICAL JUDGEMENT

Clinical medicine is the practice of medicine directly upon a patient.[146] It involves taking a patient's history, conducting necessary examinations, making diagnoses, and deciding on a course of action on the basis of that information.[147] It also involves many inherently human components, including the maintenance of interpersonal relationships, trust and fidelity.[148] Clinical judgement, the exercise of judgement in clinical practice by healthcare providers on the basis of their experience, 'lies at the heart of the doctor's connoisseurship, expertise and skills',[149] and is central to the practice of clinical medicine.[150] Certainly, it is necessary to rely on judgement, and not just flowcharts (for example) because the inherent uncertainty of an individual patient's condition regularly requires the application of past experience to novel situations.[151] However, since clinical judgement operates on the basis of intuition—a doctor's 'gut feeling'—its underlying reasoning is largely inexplicable,[152]

---

[146] Indeed, by comparison to medical research, for example, the presence of the patient is the distinctive element. '[U]nlike the biomedical scientist who may be concerned with the general properties of different infectious agents in humans, the clinician is concerned with the properties of particular pathogens in *particular human patients*' (emphasis added): Hugo Tristram Engelhardt, Stuart F Spicker and Bernard Towers, *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14-16, 1977* (Springer 1979) 268.

[147] RRH Lovell and AE Doyle, *An Introduction to Clinical Medicine* (2nd edn, Melbourne University Press 1971) 1–4.

[148] See generally Ian B Wilkinson and others, 'Thinking about Medicine', *Oxford Handbook of Clinical Medicine* (Oxford University Press 2017).

[149] Gunver S Kienle and Helmut Kiene, 'Clinical Judgement and the Medical Profession' (2011) 17 Journal of Evaluation in Clinical Practice 621, 621.

[150] ibid.

[151] This is a commonly cited reason for the fact that doctors have not already been replaced by machines, as commentators predicted in the 1970s: see Jerrold S Maxmen, *The Post-Physician Era: Medicine in the Twenty-First Century* (John Wiley and Sons 1976). In relation to the necessity of judgement in dealing with the inherent uncertainty of medicine, see, for example, Cabitza, Rasoini and Gensini (n 1). That article also refers to an interesting but technical discussion of diagnostic uncertainty in a specific clinical context in Kumar Dharmarajan and others, 'Treatment for Multiple Acute Cardiopulmonary Conditions Among Older Patients Hospitalized with Pneumonia, Chronic Obstructive Pulmonary Disease, or Heart Failure' (2016) 64 Journal of the American Geriatrics Society 1574.

[152] See, for example, Vimla L Patel, José F Arocha and David R Kaufman, 'Expertise and Tacit Knowledge in Medicine', *Tacit Knowledge in Professional Practice* (Psychology Press 1999).

and for that reason, since the mid-20th century, clinical judgement has been subject to criticism for its unreliable and irrational nature.[153] This criticism is supported by numerous studies that establish the consistent underperformance of clinicians against even simple mathematical techniques.[154] As Ofri reflects, '[c]linical judgment is often seen as a smokescreen for not having read this week's *NEJM* or *Lancet*';[155] in other words, the supposed exercise of expert judgement is a basis other than 'hard science' on which to assert unassailable authority.[156]

There are several important things to note about clinical judgement which are relevant to the overarching purpose of my inquiry: **(1)** its evident inconsistency is the reason the medical profession relies on decision support systems; **(2)** it remains a necessary component of clinical medicine and understanding its shortcomings is necessary for a proper appreciation of the effect of clinical decision support systems on clinical decision-making; **(3)** it provides an interesting framework for the incorporation of trusted but unexplained machine decisions into medicine—that is, although artificial intelligence is often complained of for its 'black box' nature, human clinicians are black boxes, too; and **(4)** it offers a convenient lens through which to examine the retention of human oversight and autonomy, since many decisions involving the exercise of clinical judgement may have multiple equally acceptable solutions (such as one of several different treatments), and

---

153 See, for example, Lawrence L Weed and Lincoln Weed, 'Opening the Black Box of Clinical Judgment—an Overview' (1999) 319 BMJ 1279; AS Elstein, 'Clinical Judgment: Psychological Research and Medical Practice' (1976) 194 Science 696; Danielle Ofri, 'The Practice of Medicine: Neither Science nor Art' (2006) 367 The Lancet 807.

154 See Michael Scriven, 'Clinical Judgment' in Hugo Tristram Engelhardt, Stuart F Spicker and Bernard Towers (eds), *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14–16, 1977* (Springer Netherlands 1979) 3; see generally Lewis R Goldberg, 'Simple Models or Simple Processes? Some Research on Clinical Judgments' (1968) 23 American Psychologist 483; see also Meehl (n 6).

155 Ofri (n 153) 807.

156 ibid.

therefore unlike decision-making tasks which have clear answers, human decision-making may be more susceptible to influence. In short, these are all factors that are relevant when considering the utility and permissibility of machine decision-making in medicine but the existence of which cannot be ascertained by looking merely at some technological artifact in isolation, or when considering the field of artificial intelligence generally.

The declining reputation of clinical judgement

Historically, and in stark contrast to the evidence-based medicine ('EBM') that reigns supreme today,[157] clinical medicine was practised in a varying manner according to the clinical reasoning, judgement and intuition (that is, the 'artful, informal, qualitative, or not explicitly quantitative strategies generally employed by clinicians')[158] that developed in individual physicians, hospitals, and localities.[159] In the 1970s and 1980s, a wave of scholars and practitioners identified and detailed cognitive biases and inconsistent or unsubstantiated—but widely used—medical practices that arose from this somewhat haphazard development in clinical settings.[160] The perceived value of clinical judgement, as evidence of the effectiveness of clinical practice, was low,[161] and with the rise of EBM

[157] On the definition of 'evidence-based medicine' see David L Sackett and others, 'Evidence Based Medicine: What It Is and What It Isn't' (1996) 312 BMJ 71.

[158] Elstein (n 153) 696.

[159] Lovell and Doyle (n 147) 1. On the topic of variation in practice between regions, see, for example, John Wennberg and Alan Gittelsohn, 'Small Area Variations in Health Care Delivery: A Population-Based Health Information System Can Guide Planning and Regulatory Decision-Making' (1973) 182 Science 1102.

[160] See generally Elstein (n 153); Alvan Feinstein, *Clinical Judgment* (Williams & Wilkins 1967); David M Eddy and John Billings, 'The Quality of Medical Evidence: Implications for Quality of Care' (1988) 7 Health Affairs 19; David M Eddy, 'Variations in Physician Practice: The Role of Uncertainty' (1984) 3 Health Affairs 74.

[161] Jerry S Wiggins, *Personality and Prediction: Principles of Personality Assessment* (Addison-Wesley 1973) 131; Paul M Spengler and others, 'The Meta-Analysis of Clinical Judgment Project: Effects of Experience on Judgment Accuracy' (2007) 37 The Counseling Psychologist 350, 350–1.

as an alternative, this became unsatisfactory. Archie Cochrane, a Scottish medical reformist in this period, lamented:

> The oldest, and probably still the commonest form of evidence proffered, is clinical opinion. This varies in value with the ability of the clinician and the width of his experience, but its value must be rated low, because there is no quantitative measurement, no attempt to discover what would have happened if the patients had had no treatment, and every possibility of bias affecting the assessment of the result. *It could be described as the simplest (and worst) type of observational evidence*.[162]

In response to these developments, clinicians and commentators proposed the systematic standardisation and formalisation of medical knowledge to reduce error and supplement clinical expertise as a basis for the continuing development of medical practice (though of course, medical reasoning and its rational bases had been a concern long before then).[163] Proposed solutions to the problem of inconsistent or unreliable clinical judgement ranged from the development and promulgation of medical guidelines[164]—which have a 'limited effect on changing physician behavior'[165]—to the reconstruction of clinical

---

[162] Archie Cochrane, *Effectiveness and Efficiency: Random Reflections on Health Services* (Nuffield Provincial Hospitals Trust 1972) 20–21 (emphasis added). Others, such as Michael Scriven, put the matter less delicately:

> The greatest achievements of scientific method in the twentieth century include a large number of demonstrations that accepted scientific or expert practices are, not to put too fine a point on it, bogus.

Scriven (n 154) 4.

[163] Gilbert Blane, *Elements of Medical Logic* (Huntington and Hopkins 1822). Engelhardt et al also refer to Hippocrates' *Precepts* I-III as one of the first sources of 'recognition of the need for greater rigor in clinical reasoning': Engelhardt, Spicker and Towers (n 146) 267.

[164] See David M Eddy, 'Practice Policies—Guidelines for Methods' (1990) 263 JAMA 1839; David M Eddy, *Clinical Decision Making: From Theory to Practice: A Collection of Essays from JAMA* (Jones and Bartlett Publishers 1996).

[165] Michael D Cabana and others, 'Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement' (1999) 282 JAMA 1458, 1458.

reasoning in symbolic form for dissemination by 'artificial intelligence'.[166] In opposition to these developments, many clinicians maintained—in a 'barrage of rationalizations'[167]—that the formalisation of explicit knowledge could not hope to match the 'intuitive decisions of clinicians',[168] especially in novel cases where deviation from standard procedures was necessary.[169] The debate caused significant polarisation,[170] because studies outlining the shortcomings of clinical judgement gave 'a critical, unflattering view of a very highly prized human faculty … The message of the research is often that we are not as adept as we thought we were …'[171]

Clinical decision support systems, and other methods of rationalising judgment, were therefore introduced to 'guard against any use of judgement',[172] in a harsh rebuke of an intrinsically human capacity that was central to the conception of medicine as a finely honed skill (as opposed to an exact science).[173] In this context, it is easy to understand why clinicians maintain their objections to the imposition of automated processes that they perceived as being unable to accommodate the richness and complexity of an individual patient's circumstances. Indeed,

---

166 Engelhardt, Spicker and Towers (n 146) xi.

167 ibid; Scriven (n 154) 4.

168 Engelhardt, Spicker and Towers (n 146) xi.

169 Elstein (n 153) 698.

170 For a highly-cited example of responses to this trend of distrust in clinical judgement, albeit in the field of psychology, see RR Holt, 'Yet Another Look at Clinical and Statistical Prediction: Or, Is Clinical Psychology Worthwhile?' (1970) 25 American Psychologist 337, 348; Spengler and others (n 161) 350–1; see also Elstein (n 153) 696: 'The importance of judgment in medicine has never been doubted'.

171 Elstein (n 153) 698.

172 Kienle and Kiene (n 149) 621.

173 For a discussion of the use of CDSS in EBM, see Ida Sim and others, 'Clinical Decision Support Systems for the Practice of Evidence-Based Medicine' (2001) 8 Journal of the American Medical Informatics Association 527.

> [t]he impossible is expected of doctors because we are socialised to rely on their acquired knowledge and cognitive abilities. But cognitive psychology has shown that the human mind normally functions by oversimplifying and filtering complex information.[174]

Nonetheless, human error remains a cause of much preventable harm,[175] and even simple measures such as automated warnings for prescriptions that exceed guideline limits have significant benefits.[176] Clinical judgement, like the exercise of human intuition generally, can be flawed, and clinical decision support systems are one solution to the errors that result from poor judgement. However, there are obvious benefits of the human capacity to reason by analogy in new situations, and to the simple fact that 'the clinician often knows more about the patient than can be put into the computer'.[177] Thus, as Scriven put it:

> The two most interesting facts about clinical judgment are that it does sometimes turn out to be right even when the judge cannot explicate his or her reasons for the judgment; and that, on the other hand, it is virtually always less reliable than a simple formula …[178]

This brief review of the history of artificially-intelligent technology and clinical judgement highlights the difficult 'tension between the art and science of medicine'[179] and the dependence of modern medicine on the consistency provided by machines and simple

---

[174] Weed and Weed (n 153) 1279.

[175] See generally Institute of Medicine, *To Err Is Human: Building a Safer Health System* (The National Academies Press 2000) <https://www.nap.edu/catalog/9728/to-err-is-human-building-a-safer-health-system>.

[176] David W Bates and others, 'Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors' (1998) 280 JAMA 1311; David W Bates and others, 'The Impact of Computerized Physician Order Entry on Medication Error Prevention' (1999) 6 Journal of the American Medical Informatics Association 313.

[177] Berner (n 138) 4.

[178] Scriven (n 154) 3–4.

[179] Patel, Arocha and Kaufman (n 152) 75.

statistical analyses. In light of this context, it is worth noting that the consequence of the sacrifice of physician autonomy is ostensibly better medicine, and thus to the extent that machines facilitate that outcome, any incursion into the domain of human decision-making is weathered contentedly. Although perhaps obvious, it might also be useful to observe that the permissibility of machine influence is partially determined by the priorities of the domain.

With the benefit of this context, I turn to clinical decision support systems, a convenient and useful subset of medical technological artifacts within which to examine machine decision-making.

CLINICAL DECISION SUPPORT SYSTEMS

INTRODUCTION

Clinical decision support systems ('CDSS') are 'computer systems designed to impact clinician decision making about individual patients at the point in time that these decisions are made'.[180] This is a broad definition; importantly, it is intended to include CDSS that draw conclusions, and thus make decisions, independently of a clinician but where the clinician is responsible for implementing the decision. Machines that perform independent actions without human intervention, such as some ventilators and drug infusion pumps,[181] impact clinical decision-making at an earlier stage in the provision of care (a decision is made to delegate rather than follow a recommendation).[182] I do not consider medical computer systems that impact a patient directly ('outpatient' CDSS).[183] The purpose of this short chapter is to evidence the breadth of technological artifacts that meet this definition, and on the basis of that understanding and the complexity it reveals, introduce the

[180] Berner (n 138) 1. Earlier definitions are narrower, but the widespread use of machine learning necessitates a broader definition such as that one offered here. See Sim and others (n 173) 528.

[181] For a discussion of observed problems arising in the use of autonomous mechanical ventilators, see Nicolas Dufour and others, 'When a Ventilator Takes Autonomous Decisions without Seeking Approbation nor Warning Clinicians: A Case Series' (2020) 13 International Medical Case Reports Journal 521; Shortliffe and Sepúlveda (n 122) 2200. For proof-of-concept discussion of algorithms that might perform these functions independently, see Melanie K Bothe and others, 'The Use of Reinforcement Learning Algorithms to Meet the Challenges of an Artificial Pancreas' (2013) 10 Expert Review of Medical Devices 661; Niranjani Prasad and others, 'A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units' [2017] arXiv:1704.06300 [cs] <http://arxiv.org/abs/1704.06300> accessed 19 June 2021.

[182] For a more strict exclusion of CDSS that 'make' decisions, see the definition given in Robert A Greenes (ed), *Clinical Decision Support: The Road to Broad Adoption* (Elsevier 2014) 8.

[183] Widespread use of fitness trackers and other consumer electronics that provide health information, and their effect on human decision-making, is an interesting topic in itself, though outpatient CDSS is not a new field: see, for example, Amy Berlin, Marco Sorani and Ida Sim, 'Characteristics of Outpatient Clinical Decision Support Systems: A Taxonomic Description' (2004) 107 Studies in Health Technology and Informatics 578.

alternative taxonomic lens (knowledge and data) through which I examine CDSS in more detail.

Examples of CDSS include diagnostic decision support systems, which suggest diagnoses on the basis of patient data; systems that support laboratory test or medication ordering (and provide warnings if medication is over-prescribed or the same test is requested in quick succession);[184] systems that check for harmful interactions between proposed medications ('drug-drug interactions');[185] and systems that automatically notify clinicians of 'serious clinical conditions' via pagers.[186] In many instances, CDSS are integrated into a hospital's electronic health record ('EHR') system and computerised physician order entry ('CPOE') system, so can provide immediate information about a patient or proposed course of treatment. This type of system reflects the supportive nature of modern CDSS, which are generally no longer designed to provide an *answer* to an unresolved clinical question,[187] but instead provide information to assist the clinician to make a more informed decision. This is an important distinction: historically, the United States Food and Drug Administration ('FDA'), which is responsible for regulating medical

[184] William L Galanter, Robert J DiDomenico and Audrius Polikaitis, 'Preventing Exacerbation of an ADE with Automated Decision Support' (2002) 16 Journal of Healthcare Information Management 44; Berner (n 138) 3.

[185] See, for example, Marilyn D Paterno and others, 'Tiering Drug–Drug Interaction Alerts by Severity Increases Compliance Rates' (2009) 16 Journal of the American Medical Informatics Association 40.

[186] This example may seem like it simply notifies a clinician rather than assisting the clinician to make a decision, but actually, the system **(a)** offers suggestions about appropriate responses; and **(b)** uses a medical knowledge base, like diagnostic CDSS, to inform its actions: see Gilad J Kuperman and others, 'Detecting Alerts, Notifying the Physician, and Offering Action Items: A Comprehensive Alerting System.' [1996] Proceedings of the AMIA Annual Fall Symposium 704.

[187] By comparison, INTERNIST-I, an early CDSS, did not explain its reasoning process and reduced the physician's role to 'that of a passive observer': Miller and Masarie Jr. (n 129); Miller, Pople Jr. and Myers (n 129).

devices,[188] declined to regulate CDSS on the basis that they were merely assistive tools[189]—'just another special test'[190]—rather than independent decision-makers, but has since recognised that some CDSS actually 'perform complex actions and may in fact represent the "last mile" between the care provider and the patient'.[191]

Given the diversity of this subset of medical technology, how can we begin to understand it in more detail, and how can we extract meaningful observations about artificially-intelligent technology? There are various taxonomies that can be used to categorise CDSS, which I discuss below. What they reveal is that even at this precise level of inquiry—a specific class of technological artifacts within a specific field—there is a great deal of variance. Although it is informative to canvass the topography of technological difference, the overarching purpose of the inquiry necessitates reliance on a higher-level taxonomy, which is useful for understanding the broad themes that underpin the field and are most relevant from a regulatory and legal perspective. Thus, the most useful and widely applicable taxonomy for the purpose of understanding artificial intelligence in the context of CDSS is one that simply divides CDSS into 'knowledge-based' and 'non-knowledge-based', or 'data-driven'. This distinction is uncontentious, technologically significant, and reflects the aforementioned historical transition from knowledge representation to data.

---

[188] See 21 USC § 360c.

[189] W Nicholson Price II, 'Black-Box Medicine' (2015) 28 Harvard Journal of Law & Technology 419, 458; Shortliffe and Sepúlveda (n 122) 2200; Frank E Young, 'Validation of Medical Software: Present Policy of the Food and Drug Administration' (1987) 106 Annals of Internal Medicine 628.

[190] FT de Dombal, 'Ethical Considerations Concerning Computers in Medicine in the 1980s' (1987) 13 Journal of Medical Ethics 179, 179.

[191] Jacob M Reider, 'Impact of National Policies on the Use of Clinical Decision Support' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (Springer 2016) 126; see also US Food and Drug Administration, 'Software as a Medical Device (SAMD): Clinical Evaluation' (8 December 2017) <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation> accessed 3 May 2021.

EXISTING TAXONOMIES OF CLINICAL DECISION SUPPORT SYSTEMS

Since CDSS have been in use, in various forms, in clinical practice for decades, taxonomic descriptions have necessarily evolved to incorporate technological differences that emerge. Thus, in the 1990s, taxonomies focused on divisions within classes of CDSS that might instead be grouped together for regulatory purposes today. Indeed, some descriptions of CDSS simply do not include data-driven systems within their definition of that term.[192] Many taxonomies are designed for the purpose of 'furthering the science of CDSS evaluation'[193]—assessing clinical effectiveness and quality of workflow integration—and these frameworks are less useful for regulatory purposes than for the improvement of clinical practice (though those two objectives are undoubtedly aligned at a high level). For example, a taxonomy devised in 2003 classifies CDSS according to five categories (and various subcategories): 'Context, Knowledge and Data Source, Decision Support, Information Delivery, and Workflow'.[194] This framework, while informative, is nonetheless highly granular and departs significantly from the standard of technological generality apparently preferred by artificial intelligence policymakers.

Critically, since any taxonomy must include modern CDSS that do not have a typical knowledge base, but instead rely on large sets of aggregated data, frameworks such as the one mentioned here are not appropriate: in this case, the 'Knowledge and Data Source' category only makes provision for a 'Clinical Knowledge Source' that is national guidelines or randomised medical trials, and a 'Data Source' that is patient-specific, as

---

[192] David W Bates and others, 'Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-Based Medicine a Reality' (2003) 10 Journal of the American Medical Informatics Association 523.

[193] Amy Berlin, Marco Sorani and Ida Sim, 'A Taxonomic Description of Computer-Based Clinical Decision Support Systems' (2006) 39 Journal of Biomedical Informatics 656, 657.

[194] Ida Sim and Amy Berlin, 'A Framework for Classifying Decision Support Systems' (2003) 2003 AMIA Annual Symposium Proceedings 599, 599.

opposed to aggregated.[195] Other taxonomies divide CDSS according to when they provide decision support and whether they 'intervene' or wait to be activated;[196] what type of knowledge base they rely on; and what kind of support they provide (such as alerts and reminders).[197]

The number of functions provided by computer systems that meet the description of CDSS are vast and vary widely (reflecting the sheer number of variables in clinical practice), as do the circumstances of their implementation and the degree to which they are integrated into existing EHR and other administrative systems in hospitals. Additionally, CDSS may be developed 'in-house' or purchased as a standalone product from a commercial developer, and thus the features as designed do not necessarily match the features that are used.[198] Similarly, 'back-end capabilities'—such as the capacity of a system to access a patient's medication history—may be available but not utilised to provide a 'front-end' tool such as advising on harmful drug-drug interactions,[199] and several back-end capabilities may all facilitate the use of one front-end tool. Consequently, comprehensive taxonomies, many of which are written with the intention of providing a nuanced understanding of the field in order to compare the clinical efficacy of various tools, are similarly complex. In a 2011 study,[200] researchers assessed a range of other

[195] ibid 601.

[196] Jane Metzger and Keith MacDonald, *Clinical Decision Support for the Independent Physician Practice* (California Healthcare Foundation 2003) 11 <https://www.chcf.org/publication/clinical-decision-support-for-the-independent-physician-practice/> accessed 4 May 2021.

[197] See generally Jerome Osheroff and others, *Improving Outcomes with Clinical Decision Support: An Implementer's Guide* (2nd edn, CRC Press 2012).

[198] Wright and others (n 143) 232–4, 239.

[199] ibid 232–4.

[200] Wright and others (n 143).

proposed CDSS classification structures,[201] and developed a descriptive taxonomy divided into six categories comprising 53 qualitatively-determined types of CDSS. Examples range from tools that suggest medication 'order sets' based on patient diagnosis, to expert systems that provide recommendations on dosing and appropriateness of blood transfusions.

These taxonomies are interesting and premised on clear technological differences between CDSS. However, they are primarily descriptive, and although it may be reasonable for regulators to discriminate between CDSS on the basis of small feature differences, the predominant motivation of these researchers is to identify functions and characteristics rather than their practical consequences (which, given the enormous variety of tools and clinical environments, would be difficult to assess).

The preceding discussion reveals a great number of technological distinctions that manifest even within this relatively narrow avenue of inquiry, reinforcing the proposition that more general discussion omits a great deal of important detail. An understanding of that complexity permits me to employ a simpler taxonomic framework—as a means of greater understanding of machine decision-making, rather than reductive generality—which has a clear technological basis; is recognised in the literature; has regulatory utility; and is of epistemological significance: **some CDSS rely on knowledge, and some do not.** This distinction is laid out below.

[201] ibid 233, Table 1.

KNOWLEDGE-BASED AND DATA-DRIVEN CDSS

INTRODUCTION

So far in the second part of this thesis, I have laid the groundwork for this most focused part of my inquiry, in which I substantiate and contrast the differences between knowledge-based and data-driven forms of CDSS. The broad purpose of the foregoing is to document that complexity that inheres in (a) artificial intelligence and the field and technologies that term refers to; (b) medicine; and (c) clinical decision support systems, as a representative subset of intelligent technology. Such complexity militates against reductive general regulatory approaches that assume the applicability of existing legal rules by analogy without critically evaluating the underlying technology they purport to regulate. The specificity with which I can now discuss these two classes of CDSS—afforded by the context provided by the preceding discussion—permits meaningful observations to be made about machine decision-making and legal approaches to understanding artificial intelligence.

The first and most obvious observation is that although there are unambiguous, consequential technical and epistemological differences between these two classes of CDSS, both are described as artificial intelligence during various stages of evolution of the field and according to different metrics. Difference in underlying computational techniques is what sets these classes apart, so unqualified reference to artificial intelligence is simply inadequate.

The second and most important observation arises from the contrast afforded by juxtaposing these two classes of CDSS. As I have discussed, commentators tend to anthropomorphise artificially-intelligent technology while simultaneously adverting to its complexity, unpredictability, and opacity, and this confusing approach results in the 'myth-

making' Campolo and Crawford identified.[202] What makes machine decision-making inscrutable to humans is often not explicitly stated, but this contrast reveals the reason for that inscrutability: knowledge-based CDSS relies on human knowledge that already exists and is merely transmuted into a more convenient, modular format, whereas data-driven CDSS extracts acausal inferences from data that need not correspond to existing human ontologies, and which may be premised on relationships between data that are sensitive to more variables (dimensions) than human cognitive limits permit simultaneous consideration of. Consequently, the opaque and putatively unpredictable nature of data-driven CDSS (and modern artificially-intelligent technology that use machine learning techniques) is a product of its distinctly computational (as opposed to human) configuration. Importantly, this implies (as I discuss below) that useful knowledge about the world need not conform to human knowledge frameworks at all. Thus, regulators thinking about when to permit machine decision-making must appreciate that although machines are perceived as cold and mathematical, this is to characterise machines according to human standards. It is not appropriate to claim that machines are fundamentally inferior because they lack the capacity for intuition or cannot appreciate subtle emotional context; rather, machine decision-making simply incorporates different information.

The differences between CDSS classes may be summarised as follows. Knowledge-based systems rely on a human understanding of causality and preordained knowledge to produce deterministic outputs; are subservient to explicitly stated goals; and, as a consequence of their symbolic representation of knowledge, depend on the prior

[202] Campolo and Crawford (n 49).

exercise of human judgement to synthesise the context from which inputs are derived.[203] By comparison, data-driven CDSS do not make causal inferences;[204] produce reasoning that cannot be reduced to natural language without a loss of detail; rely exclusively on information that can be quantified 'at the expense of other [contextual] elements that are more difficult or impossible to easily describe';[205] and possess agency in the sense that they are practically unpredictable and can be adaptive.[206] This distinction is premised on clear technological differences that are highlighted in relevant literature discussed below.

The primary purpose of this detailed comparison is substantiating the following observation: data-driven machines reason computationally with information that is external to human knowledge frameworks, and they produce useful, valid decisions notwithstanding this fact. Since their reasoning cannot always be subjected to scrutiny, their authority is dependent upon their characterisation in their environment, and recognising that they are machines, reasoning mechanistically, is fundamental to that task.

DISTINCTION IN THE LITERATURE

An explicit distinction between knowledge-based and data-driven CDSS is not novel in CDSS literature,[207] and the transition from knowledge-based to data-driven clinical

[203] Bunyamin Ozaydin, J Michael Hardin and David C Chhieng, 'Data Mining and Clinical Decision Support Systems' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016) 47–8.

[204] Imbuing machines with the capacity to make causal inferences from data is a problem that computer scientists such as Judea Pearl claim hinders the development of artificial *general* intelligence; see generally Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Penguin Books 2019).

[205] Cabitza, Rasoini and Gensini (n 1) 517.

[206] See, for example, the comparison between 'agents' and knowledge-based systems in Rajendra Arvind Akerkar and Priti Srinivas Sajja, *Knowledge-Based Systems* (Jones and Bartlett Publishers 2010) 159.

[207] See most prominently Berner (n 138) chs 1–3. See also Robert Challen and others, 'Artificial Intelligence, Bias and Clinical Safety' (2019) 28 BMJ Quality & Safety 231.

decision support is discussed often. The term 'knowledge-based' is widely used and predates its juxtaposition with the term 'data-driven' (or 'non-knowledge-based'),[208] which is more ambiguous and less commonly used, and which vies with other terms, including 'machine learning', to describe technology that has proliferated with the availability of big data.[209] The explicit use of this distinction in the literature is generally descriptive or taxonomic; for example, Sutton et al. state that

> CDSS are frequently classified as knowledge-based or non-knowledge based. In knowledge-based systems, rules (IF-THEN statements) are created, with the system retrieving data to evaluate the rule, and producing an action or output; Rules can be made using literature-based, practice-based, or patient-directed evidence. CDSS that are non-knowledge based still require a data source, but the decision leverages artificial intelligence (AI), machine learning (ML), or statistical pattern recognition, rather than being programmed to follow expert medical knowledge.[210]

A similar distinction is made by Montani and Striani in a 2019 literature review which classified journal articles according to whether they described knowledge-based or data-driven CDSS.[211] Montani and Striani rely on definitions articulated in the *Barcelona*

[208] See, for example, William B Gevarter, 'Expert Systems: Artificial Intelligence Applied' (1984) 1 Telematics and Informatics 239, 240. For alternative but overlapping use of the term 'data-driven' see generally Hildebrandt (n 3).

[209] For an interesting definition of big data, see generally Emilie Baro and others, 'Toward a Literature-Driven Definition of Big Data in Healthcare' [2015] BioMed Research International <https://doi.org/10.1155/2015/639021> accessed 11 June 2021.

[210] Reed T Sutton and others, 'An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success' (2020) 3 npj Digital Medicine 1, 1.

[211] Stefania Montani and Manuel Striani, 'Artificial Intelligence in Clinical Decision Support: A Focused Literature Survey' (2019) 28 Yearbook of Medical Informatics 120.

Declaration for the Proper Development and usage of Artificial Intelligence in Europe,[212] which describes 'knowledge-based AI' as

> an attempt to model human knowledge in computational terms, starting in a top-down fashion from human self-reporting of what concepts and knowledge individuals use to solve problems or answer queries in a domain of expertise …[213]

'Data-driven AI' is described as

> starting in a bottom-up fashion from large amounts of data of human activity, which are processed with statistical machine learning methods [...] in order to abstract patterns that can then be used to make predictions, complete partial data, or emulate human behaviour in similar conditions in the past.[214]

Other scholars do not premise their discussion of CDSS on this distinction, but comment on a general transition between modes of decision support reflecting background changes in the field:

> While existing approaches to clinical decision support [various methods of distributing knowledge-based content] can already reinforce consistency with best-practices, production of this content is limited in scale by the human-expert, knowledge-based authoring necessary for each intervention. …

---

[212] Although impressively aspirational, the reach and academic relevance of the *Barcelona Declaration* itself is unclear and should not be overstated. For details of the *Barcelona Declaration*, organisers of the event during which it was made, and its signatories, see Artificial Intelligence Research Centre of the Spanish National Research Council, 'Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe' (*IIIA CSIC*, 2017) <http://www.iiia.csic.es/barcelonadeclaration/> accessed 23 June 2021; Luc Steels and Ramon Lopez de Mantaras, 'The Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe' (2018) 31 AI Communications 485; International Center for Scientific Debate, 'Artificial Intelligence: Dreams, Risks, and Reality' (*B-DEBATE*, 2017) <https://www.bdebate.org/en/forum/artificial-intelligence-next-step-evolution> accessed 23 June 2021.

[213] Montani and Striani (n 211) 120.

[214] ibid.

> The meaningful use era of electronic health records (EHR) creates an opportunity for data-driven clinical decision support (CDS) to reduce detrimental practice variability with the collective expertise of many practitioners in a learning health system.[215]

In fact, appreciation for the ways in which insights from big data can complement the distribution of traditionally-derived medical knowledge is unsurprisingly widespread in the medical profession,[216] and the distinction between knowledge-based and data-driven CDSS, at least in clinical contexts, is almost too obvious to note. Nonetheless, this distinction does not receive comparable attention in legal literature.

KNOWLEDGE-BASED CDSS

Knowledge-based CDSS are so called because they rely on a knowledge base from which to 'reason' to an answer. Knowledge is derived by, extracted from, and logically formulated by humans: there is no novel insight, but merely the convenient and timely recapitulation of existing knowledge. There are dozens of types of knowledge-based CDSS: **expert systems** are complex, intelligent, and rely on the explicit representation of large bodies of knowledge; more simple systems rely on knowledge, too, but do not exhibit intelligence. Such simple knowledge-based CDSS include computer-based order sets (mentioned

[215] Jonathan H Chen and others, 'Decaying Relevance of Clinical Data towards Future Decisions in Data-Driven Inpatient Clinical Order Sets' (2017) 102 International Journal of Medical Informatics 71, 72 (footnotes omitted).

[216] See particularly Jennifer Frankovich, Christopher A Longhurst and Scott M Sutherland, 'Evidence-Based Medicine in the EMR Era' (2011) 365 New England Journal of Medicine 1758. See also Dean F Sittig and others, 'Grand Challenges in Clinical Decision Support' (2008) 41 Journal of Biomedical Informatics 387, 389–90; Harlan M Krumholz, 'Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System' (2014) 33 Health Affairs 1163; Greg de Lissovoy, 'Big Data Meets the Electronic Medical Record: A Commentary on "Identifying Patients at Increased Risk for Unplanned Readmission"' (2013) 51 Medical Care 759.

above),[217] systems that hold knowledge about interactions between drugs,[218] or those which sound alerts when certain conditions are satisfied;[219] all are based on preordained medical knowledge, even if merely implicit in their functioning. Most knowledge-based CDSS are these simple applications that are integrated into a hospital's EHR.[220]

Three examples of expert systems

Expert systems were the first form of knowledge-based system,[221] and the first form of artificial intelligence generally considered a success. What ties the following examples together is their reliance on human reasoning and knowledge. **MYCIN**, an early and well-known expert system used to diagnose bacterial infections and recommend antibiotics,[222] was rule-based (although expert systems need not be).[223] MYCIN's rules were structured in an IF-THEN format, where information given by the user (satisfying the IF condition) decided whether the action part of the rule (THEN) was executed. One of MYCIN's rules is as follows:

---

[217] See generally Thomas H Payne and others, 'Preparation and Use of Preconstructed Orders, Order Sets, and Order Menus in a Computerized Provider Order Entry System' (2003) 10 Journal of the American Medical Informatics Association 322.

[218] See, for example, Gilad J Kuperman and others, 'Medication-Related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review' (2007) 14 Journal of the American Medical Informatics Association 29.

[219] Kuperman and others (n 186).

[220] S Andrew Spooner, 'Mathematical Foundations of Decision Support Systems', *Clinical Decision Support Systems: Theory and Practice* (Springer 2016) 19–20.

[221] Durkin (n 134) 11.

[222] See generally Shortliffe and Buchanan (n 106); see specifically ch 5, 'Details of the Consultation System'.

[223] For example, where the knowledge contained in MYCIN's knowledge base comprised IF-THEN rules, as described above, the knowledge contained in the knowledge base of INTERNIST-I, another early expert system, was divided into disease profiles, including symptoms and physical signs of the disease: see Miller, Pople Jr. and Myers (n 129) 469–70. For a discussion of

RULE037

IF: (1) The identity of the organism is not known with certainty, and

(2) The stain of the organism is gramneg, and

(3) The morphology of the organism is rod, and

(4) The aerobicity of the organism is aerobic

THEN: There is strong suggestive evidence (.8) that the class of the organism is enterobacteriaceae.[224]

There are two features of this rule that are immediately apparent: its formulation requires domain-specific knowledge; and there is no mechanism allowing the rule to incorporate new medical insights that affect the validity of its conclusions—if the rule were to no longer reflect sound medical reasoning, it would have to be rewritten.[225] Thus, the rule is the symbolic, linear, local,[226] static expression of the knowledge of a medical expert, intended to capture their reasoning process and the level of certainty this evidence gives them.

MYCIN used rules of thumb—heuristics,[227] or informal reasoning—which are intimately human in nature and which were, at the time, 'an essential key to intelligent problem solving because computationally feasible, mathematically precise methods [were]

[224] Shortliffe and Buchanan (n 106) 82.

[225] Note, however, that the modularity of rule-based systems means that knowledge can be easily added or modified. See generally ibid ch 3, 'The Evolution of MYCIN's Rule Form'.

[226] A local representation is the representation of an entire object by one symbol, as opposed to by distributed component parts, such as in a biological or artificial neural network: see Geoffrey E Hinton, James L McClelland and David E Rumelhart, 'Distributed Representations' in David E Rumelhart, James L McClelland, and PDP Research Group (eds), *Parallel distributed processing: explorations in the microstructure of cognition*, vol 1 (MIT Press 1986).

[227] For the seminal work in the study of heuristics, see Amos Tversky and Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185 Science 1124.

known for only a relatively few classes of problems'.[228] That is to say, expert systems were necessarily modelled on human heuristic reasoning processes, which are potentially suboptimal but incorporate reliable assumptions that exclude technically possible conclusions, because otherwise there would simply be too much to compute.[229]

MYCIN is very clearly deterministic and was programmed according to rules expressed in natural language. However, perhaps obviously, human knowledge need not be expressed in the form of language-based rules; indeed, physicians use probabilistic reasoning in their decision-making, even if that process culminates in a verbal explanation. **DXplain**, another early diagnostic system, expressed relationships between diseases and symptoms in the form of two numbers on an ordinal scale, 'one representing the frequency with which the finding occurs in the disease and the other the degree to which the presence of the finding suggests consideration of the disease'.[230] Thus, although DXplain's knowledge base is a quantification of the relationship between a disease and a symptom, it is human knowledge that originally associates a certain finding with a particular disease and prescribes the strength of the causal relationship between the two.

The final example is a CDSS that uses **fuzzy logic** to diagnose peripheral neuropathy.[231] Although DXplain and MYCIN rely on precise, unchanging logic and probability relationships elicited from the subject-matter experts, they are intended to

[228] Bruce G Buchanan and Richard O Duda, 'Principles of Rule-Based Expert Systems' in Marshall C Yovits (ed), *Advances in Computers*, vol 22 (Academic Press 1983) 165.

[229] For an illustrative example of computational infeasibility, and how that problem led to the creation of DENDRAL, the first expert system, see Durkin (n 134) 11. See also Robert K Lindsay and others, *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project* (McGraw-Hill 1980).

[230] Edward P Hoffer and others, 'DXplain: Patterns of Use of a Mature Expert System' (2005) 2005 AMIA Annual Symposium Proceedings 321.

[231] Reeda Kunhimangalam, Sujith Ovallath and Paul K Joseph, 'A Clinical Decision Support System with an Integrated EMR for Diagnosis of Peripheral Neuropathy' (2014) 38 Journal of Medical Systems 38.

reflect, as closely as possible, the informal, imprecise, protean logic processes used by human experts.[232] In an ideal world, knowledge-based systems would map these human reasoning processes exactly, but in reality,

> experts, in spite of being highly skilled in solving problems in their field, often feel difficulty in stating their knowledge in an orderly and logical manner or sometimes even in understanding their own decision making processes.[233]

This difficulty reflects the uncertain nature of measurement inherent in reality.[234] Fuzzy logic approximately describes the amount of truth of a proposition (unlike binary logic, which is categorical) and permits machines to reason with uncertainty about systems 'where the vast arrays of inputs and outputs could not all possibly be captured analytically or controlled in any conventional sense'.[235] In the case of this fuzzy expert system, a non-expert user simply inputs values describing (a) the results of clinical tests; and (b) an approximation of the severity—expressed by 'fuzzy' variables such as low, normal, or high—of indicative symptoms,[236] and receives a diagnosis. Even in the case of this imprecise representation of knowledge, the underlying reasoning process is based on an expert's classification of certain values as being low, normal, or high; only from these predefined divisions can the CDSS infer a diagnosis.

---

[232] See generally Geoff Norman, Meredith Young and Lee Brooks, 'Non-Analytical Models of Clinical Reasoning: The Role of Experience' (2007) 41 Medical Education 1140; Kevin W Eva, 'What Every Teacher Needs to Know about Clinical Reasoning' (2005) 39 Medical Education 98; Geoffrey R Norman and Kevin W Eva, 'Diagnostic Error and Clinical Reasoning' (2010) 44 Medical Education 94.

[233] Kunhimangalam, Ovallath and Joseph (n 231).

[234] On the notion of inherent uncertainty in reality, see the highly-cited essay by Max Black, 'Vagueness: An Exercise in Logical Analysis' (1937) 4 Philosophy of Science 427.

[235] Timothy J Ross, *Fuzzy Logic with Engineering Applications* (2nd edn, John Wiley & Sons 2005) 7.

[236] Fuzzy variables are so-called because the range of values they describe overlap and so their boundaries are not sharp. See Lofti A Zadeh, 'Foreword' in Timothy J Ross, Jane M Booker and WJ Parkinson (eds), *Fuzzy Logic and Probability Applications: Bridging the Gap* (SIAM 2002).

Modelling human knowledge

The genius of expert systems was modelling them on experts, who use domain knowledge to constrain their analyses to a subset of likely possibilities; extracting this knowledge allowed knowledge-based systems to avoid the tedious task of considering *everything*.[237] Modelling human reasoning processes worked well, and justified the description of expert systems as 'intelligent', since the captured knowledge and inferences permitted such systems to 'solve problems that are difficult enough to require significant human expertise for their solution'.[238]

Intelligence was thus framed in terms of human faculties since competent machines relied on human knowledge (expressed, at least implicitly, by reference to causal models) for their reasoning, but this was simply the product of technological limitations; a historical quirk. Consider the illustrative observation of Peter Szolovits in his 1982 *Artificial Intelligence in Medicine*:

> The choice to model the behavior of a computer expert in medicine on the expertise of human consultants is **by no means logically necessary**. If we could understand the functioning in health and in disease of the human body in sufficient depth to model the detailed disease processes which disturb health, then, at least in principle, we could perform diagnosis by fitting our model to the actually observable characteristics of the patient at hand. … Unfortunately, although biomedical research strives for such a depth of understanding, it has not been achieved in virtually any area of medical practice. … most of what we know about the practice of medicine we know from interrogating the

[237] Despite the successes of rule-based systems, the representation of heuristic knowledge in a derivative form (as rules about associations rather than 'underlying disease models') is not without problems. See William J Clancey, 'The Epistemology of a Rule-Based Expert System—a Framework for Explanation' (1983) 20 Artificial Intelligence 215.

[238] Feigenbaum (n 17) 1.

> best human practitioners; therefore, the techniques we tend to build into our programs mimic those used by our clinician informants.[239]

This observation reflects an underlying, interesting, and counter-intuitive point. Intelligence, which Tom Stonier claims is simply 'a property of advanced information systems',[240] is generally perceived through a lens tinged by anthropocentric bias to be the exclusive property of human beings,[241] but there is no reason to impose such a limitation,[242] and thus no need to draw useful medical conclusions exclusively from the reasoning of human experts. Concepts that we describe as 'intelligence' and 'knowledge' are merely the human *interpretation* of underlying information, and there are other methods of reasoning. Indeed, where a physician identifies a relationship between a symptom and a disease as being *causal*, it is either because (a) the underlying mechanism is understood and supports the causal inference;[243] or (b) the physician recognises a pattern consistent with their 'compiled' knowledge and is certain of the empirical association, *despite not necessarily understanding the underlying mechanism*.[244] Both types of reasoning are knowledge, even where deeper understanding could be attained. In light of this distinction, it is misleadingly reductive to refer to artificial intelligence as a general 'attempt to model human

---

[239] Szolovits (n 127) ch 1 (emphasis added).

[240] Tom Stonier, 'The Spectrum of Intelligence' in Tom Stonier (ed), *Beyond Information: The Natural History of Intelligence* (Springer 1992) 19 <https://doi.org/10.1007/978-1-4471-1835-0_2> accessed 24 June 2021.

[241] See, for example, Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014) 84–86; Eliezer Yudkowsky, 'Artificial Intelligence as a Positive and Negative Factor in Global Risk' in Nick Bostrom and Milan M Ćirković (eds), *Global Catastrophic Risks* (Oxford University Press 2008).

[242] See generally Alan Winfield, 'Intelligence Is Not One Thing' (2020) 11 Journal of Artificial General Intelligence 97; Brachman (n 18).

[243] This distinction is described in Peter Szolovits and Stephen G Pauker, 'Categorical and Probabilistic Reasoning in Medical Diagnosis' (1978) 11 Artificial Intelligence 115, 124; Clancey (n 237) 224–226.

[244] See, for example, Eva (n 232) 100–101. For a discussion of the application of 'compiled' knowledge as 'categorical reasoning', see Szolovits and Pauker (n 243) 116–117. For a discussion of the common ignorance of underlying drug mechanisms, see, for example, Editorial, 'Mechanism Matters' (2010) 16 Nature Medicine 347.

intelligence'; in the case of modern, data-driven CDSS, the intention is to model the underlying information.

Conclusion

Knowledge-based systems are so-called because they rely on the explicit formulation and encoding of medical knowledge by experts. Many CDSS, including the simplest systems, are knowledge-based. The informal reasoning processes of experts are used to manipulate and draw conclusions from their knowledge, although both processes and knowledge are represented imperfectly.[245] It is convenient to rely on an expert's reasoning process because this is the framework that imputes causal relevance to medical knowledge, but this is not always necessary. As Szolovits observes, a deeper causal understanding of disease, biology, and health would permit us to dispense with higher-level reasoning processes described by experts as the justification for their beliefs. Consequently, although we have knowledge-based systems, human knowledge need not be the only basis for valid conclusions.

DATA-DRIVEN CDSS

Introduction

From the foregoing it is apparent that describing knowledge-based artificially-intelligent technology as an attempt to model human intelligence is apt. But what about CDSS that are not knowledge-based?

Rather than knowledge derived from human reasoning processes, data-driven CDSS recognise and fit functions to patterns in data, and relevant medical insight is taken

[245] In fact, although experts give explicit reasons to justify their intuitive reasoning process, these may not actually accord with the relevant factors that underpin their intuition. Consequently, any form of knowledge-based computer reasoning is necessarily limited by imperfect human attempts to reconstruct subconscious decision-making processes. See, for example Haidt (n 5); Jonathan Haidt, 'Moral Psychology and the Law: How Intuitions Drive Reasoning, Judgment, and the Search for Evidence' (2012) 64 Alabama Law Review 867.

from identified patterns. As a preliminary observation that will assist to frame the discussion, recall that medical decision-making, like all human decision-making, consists fundamentally of the making of causal inferences,[246] notwithstanding historical scientific (and philosophical) distaste for the concept of causality.[247] This is to say, 'whether or not we call them "causal," we would like to predict the effects of our actions, not just correlations between observable variables'.[248] Data, on the other hand, 'are profoundly dumb about causal relationships':[249] correlation does not equal causation. Thus, here is a significant source of difference between data-driven and knowledge-based CDSS: humans tend to reason causally, but data-driven CDSS need not.

This section on data-driven CDSS is divided into the following two parts: first, a discussion of machine learning, including examples, to clarify the meaning of that term which is central to modern artificially-intelligent technology; and second, examples of data-driven CDSS, which rely on machine learning techniques and exemplify the distinction between knowledge-based and data-driven CDSS.

(1) Machine learning

The classic definition of machine learning was given by Tom Mitchell in 1997:

---

[246] See Vimla L Patel, David A Evans and Guy J Groen, 'Biomedical Knowledge and Clinical Reasoning' in David A Evans and Vimla L Patel (eds), *Cognitive Science in Medicine* (MIT Press 1989) ch 3.2, 'Basic Science in Causal Explanations of Clinical Cases by Medical Students'; Eva (n 232) 100. See generally Michael R Waldmann and York Hagmayer, 'Causal Reasoning', *The Oxford Handbook of Cognitive Psychology* (2013); Clancey (n 237); Szolovits and Pauker (n 243).

[247] Bertrand Russell, for example, stated that 'The law of causation … is a relic of a bygone age': Bertrand Russell, *The Collected Papers of Bertrand Russell Volume 6: Logical and Philosophical Papers 1909-13* (John G Slater ed, Routledge 1992) 193. See also David Hume's observations described in Waldmann and Hagmayer (n 246) 734. See also Jonathan Schaffer, 'The Metaphysics of Causation' in Edwards N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Fall 2016 edition, Metaphysics Research Lab, Stanford University 2016) <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics> accessed 15 July 2021.

[248] Domingos (n 14) 87.

[249] Pearl and Mackenzie (n 204) 14.

> A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.[250]

Arthur Samuel described machine learning more informally in his seminal work as 'the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning'.[251] Similarly, Nils Nilsson says that

> a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves.[252]

In a lay sense, then, and for regulatory and policy purposes, machine learning is the capacity of a machine to adapt its behaviour in light of new information without being told *how* to change in explicit, granular terms. A change in behaviour might be **(a)** on an ongoing basis in a real-world environment, such as where new data acquired during use in relation to an underrepresented population leads to refinement of the machine's behaviour in relation to members of that population; or **(b)** during the typical, pre-deployment process of training a machine learning model, where training data is input so the model can learn the function that maps the input to the desired output. The former is a current problem for regulators in relation to medical devices that require regulatory approval, since

[250] Thomas M Mitchell, *Machine Learning* (McGraw-Hill 1997) 1.

[251] Samuel (n 133) 210.

[252] Nils J Nilsson, *Introduction to Machine Learning: An Early Draft of a Proposed Textbook* (Robotics Laboratory, Department of Computer Science, Stanford University 1998) 1 <https://robotics.stanford.edu/people/nilsson/mlbook.html> accessed 10 July 2021.

the device, its efficacy, and its reliability might change in an unpredictable manner after approval has been granted.[253] Two types of machine learning are described below.

Two examples of machine learning

A **support vector machine** (or support vector network, 'SVM') attempts to distinguish between two classes in a dataset by demarcating a decision boundary, or hyperplane, that clearly separates the classes (which clump together on either side of the boundary because they share characteristics that make them identifiable as belonging to their class).[254] The boundary should be as far as possible from the most extreme examples of each class (the support vectors) because this allows the greatest room on each side of the boundary within which unseen examples may fall and be correctly classified. New data may include more extreme examples of each class, and the boundary will consequently need to be redefined to accommodate newly observed class limits. Thus, conceptually, an SVM is rather straightforward: even though it can classify non-linear data, it is, functionally speaking, just a method for establishing a dividing line between classes. Nonetheless, an SVM is an example of machine learning because it can use new information to refine its output.

Artificial neural networks can also learn and are prominent in biomedical research, popular culture, and legal scholarship. An artificial neural network is so-called because it is composed (conceptually) of artificial neurons that are typically organised into layers. Neurons are tied to other neurons by connections, each connection of a certain

[253] Angus (n 142); United States Food and Drug Administration, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)' (2019) Discussion Paper <https://www.fda.gov/media/122535/download> accessed 20 August 2021.

[254] See the seminal paper on support-vector networks: Corinna Cortes and Vladimir Vapnik, 'Support-Vector Networks' (1995) 20 Machine Learning 273. See also Arvin Agah (ed), *Medical Applications of Artificial Intelligence* (CRC Press 2013) 11–12.

weight that is varied in the process of learning to produce a desired output. Here is another example of misleading anthropomorphic conceptualisations: although neural networks are often represented symbolically by dots and lines, and are described using language that reinforces the perception that the neural network *literally* emulates brain functioning, they are really just a series of mathematical functions 'connected in a chain' (each function corresponding to a layer). The length of the chain of functions is the relative depth of the neural network,[255] and more depth—which means more functions—means more complexity,[256] and thus more capacity for nuanced behaviour.[257] A deep neural network can therefore perform 'deep' learning, and can learn a function that incorporates a greater range of features observed in data than a neural network with less layers; a 'typical' deep neural network may have hundreds of millions of adjustable weights.[258] Importantly, the 'key advantage of deep learning' is that relevant features in a dataset are identified automatically, rather than selected and engineered by hand, as was historically necessary.[259]

Consequences of the complexity of machine learning

Several observations about machine learning are relevant to my overarching inquiry: **first**, since representation learning is an automated process, the features identified by a neural network or other machine as relevant to the task are not controlled by human beings;[260] **second**, given the sheer complexity of some machine learning models, such as deep neural networks, a good model can easily 'learn more complicated and subtle patterns in the input

[255] This explanation is given in extremely useful terms in Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning* (MIT Press 2016) 164–165 <https://www.deeplearningbook.org/>.

[256] See, for example, ibid 163.

[257] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, 'Deep Learning' (2015) 521 Nature 436, 437–8.

[258] ibid 436.

[259] ibid 438.

[260] ibid.

data than the average human observer is able to';[261] **third**, in the case of high-dimensional data,[262] 'our [human] intuitions, which come from a three-dimensional world, often do not apply … in high dimensions it is difficult to understand what is happening';[263] and **fourth**, machine learning is used to identify features, describe and summarise patterns within data, and generate models to represent that information, and thus, as Issam El Naqa points out, it is directed towards the generation of hypotheses that did not previously exist, rather than the testing of hypotheses already formulated.[264]

It is therefore data, not human beings, that dictate the intricate details of machine learning that operate on patterns that exist beyond the realm of practical comprehension by human beings, and those patterns inform and guide the discovery and consolidation of new knowledge, 'not suspected *a priori*',[265] into new hypotheses. Thus, in the provocative words of Chris Anderson, even while we are poised to 'lose the tether of data as something that can be visualized in its totality', we can proceed well enough on the basis of data-driven science 'even without coherent models, unified theories, or really any mechanistic explanation at all'.[266] In other words, the world admits of a complexity that exceeds the

[261] Issam El Naqa and Martin J Murphy, 'What Is Machine Learning?' in Issam El Naqa, Ruijiang Li and Martin J Murphy (eds), *Machine Learning in Radiation Oncology: Theory and Applications* (Springer 2015) 5.

[262] High-dimensional data refers to data in which a single observation has a high number of variables (or features) that specify its properties, and thus a high number of dimensions are necessary to describe differences and relationships between datapoints. Since human beings cannot conceive visually or graphically of more than three spatial dimensions, a function describing the relationship between data with four or more variables can only be expressed with mathematical symbols and is consequently difficult to parse.

[263] Domingos (n 14) 82.

[264] Issam El Naqa, 'Computational Learning Theory' in Issam El Naqa, Ruijiang Li and Martin J Murphy (eds), *Machine Learning in Radiation Oncology: Theory and Applications* (Springer 2015) 14.

[265] MA Oquendo and others, 'Machine Learning and Data Mining: Strategies for Hypothesis Generation' (2012) 17 Molecular Psychiatry 956. See also DJ Hand, 'Mining Medical Data' (2000) 9 Statistical Methods in Medical Research 305.

[266] Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete' [2008] *Wired* <https://www.wired.com/2008/06/pb-theory/> accessed 1 May 2021.

human capacity to model and understand it, but machines can ostensibly deduce meaningful patterns from this noise. Although this circumvents the tedious, incremental bottleneck of traditional empirical observation and subsequent hypothesis,[267] we are potentially left only with access to outcomes, not explicable reasoning—a novel prospect:

> [W]hat we know depends upon the output of machines the functioning of which we cannot follow, explain, or understand.
>
> Since we first started carving notches in sticks, we have used things in the world to help us to know that world. But never before have we relied on things that did not mirror human patterns of reasoning—we knew what each notch represented—and that we could not later check to see how our non-sentient partners in knowing came up with those answers.[268]

These comments are particularly applicable in medicine, where the unwieldy complexity of interconnected biological processes—and the consequent 'inability of physicians to accurately predict patient outcomes despite advanced monitoring and knowledge of physiology'[269]—necessitated historical reliance on reductionist physiological models.[270] But machine learning facilitates integration of far more variables than human models can, as I show in representative examples of data-driven CDSS described below.

---

[267] See Oquendo and others (n 265) 956.

[268] David Weinberger, 'Our Machines Now Have Knowledge We'll Never Understand' [2017] *Wired* <https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/> accessed 17 May 2021.

[269] Michael R Pinsky, 'Complexity Modeling: Identify Instability Early' (2010) 38 Critical Care Medicine S649, S649.

[270] See generally Pinsky (n 269). See also Tim Wilson, Tim Holt and Trisha Greenhalgh, 'Complexity and Clinical Care' (2001) 323 BMJ 685.

(2) Examples of data-driven CDSS

The following examples emphasise the use of machine learning to recognise patterns in large datasets, and in which outputs are dependent on such patterns rather than explicit human reasoning.

Example: early warning of intraoperative hypotension CDSS

Description of the study

A pertinent example, and one of few clinical trials of data-driven CDSS,[271] is given by Wijnberge et al in a 2020 paper.[272] This RCT evaluated the use of a 'machine-learning derived early warning system' that predicts the onset of hypotension (low blood pressure) during surgery. The method underpinning development of the proprietary computer software is briefly described below.

The equipment used in the study measures, processes, and graphs the fluctuating blood pressure of a patient; 166 features are extracted directly from the arterial pressure waveform, including information such as pulse rate, standard deviation of the waveform, and the quantified regularity of 'fluctuations in cardiac measurements'. However, since these 'basic' features 'do not undergo changes until late in the hypotension development process'[273]—that is to say, they have no predictive power when consulted individually—researchers calculated derivative features, such as the variability of each basic feature and 'relationships between compensatory physiological processes' (for example, blood

[271] Angus (n 142).

[272] Marije Wijnberge and others, 'Effect of a Machine Learning–Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial' (2020) 323 JAMA 1052.

[273] Feras Hatib and others, 'Machine-Learning Algorithm to Predict Hypotension Based on High-Fidelity Arterial Pressure Waveform Analysis' (2018) 129 Anesthesiology 663, 11 (Supplemental Digital Content).

pressure increasing and heartrate decreasing),[274] to obtain over 3,000 additional features which more accurately describe the interconnected patterns of bodily processes that cause changes in blood pressure. The 51 most predictive features were combined to produce '2.6 million combinatorial features' that capture 'linear and nonlinear interactions'. Finally, researchers used logistic regression to model the probabilistic relationship between features and hypotension, identified by analysis of 545,959 minutes of arterial pressure waveform data (the training data) from 293 patients.[275]

In the RCT, an alarm sounded when the computer software determined that the probability of hypotension within the next 15 minutes was greater than 85%. The anaesthetist was encouraged by the alarm to act in accordance with treatment recommendations prepared by the authors of the RCT displayed on the monitor of the CDSS.[276] The result of the RCT was a reduction in the length of 'hypotensive events', and fewer hypotensive events overall, where the early warning system was used.[277] Of particular note is the fact that in 19 cases (of 377) when the early warning system sounded an alarm, the anaesthetist, suffering alarm fatigue,[278] 'refused to treat because of the frequency of the alarms'.[279]

[274] See pages 3–9 of the Supplemental Digital Content to Hatib and others (n 273).

[275] ibid 668.

[276] See M Wijnberge and others, 'The Use of a Machine-Learning Algorithm That Predicts Hypotension during Surgery in Combination with Personalized Treatment Guidance: Study Protocol for a Randomized Clinical Trial' (2019) 20 Trials 582.

[277] Wijnberge and others (n 272) 1056–1058.

[278] Marc Wilken and others, 'Alarm Fatigue: Causes and Effects' [2017] Studies in Health Technology and Informatics 107.

[279] Wijnberge and others (n 272) 1058.

Discussion

In this example, the data-driven nature of the CDSS is evident: features of the arterial waveform are extracted from a huge dataset on the basis of their associative predictive capacity, and the relevance of each feature is determined exclusively by its complicated, quantified relationship with other features. By comparison, there is no predictive insight to be gained from analysis of basic features whose details fall within human cognitive limits, and thus any attempt at causal inference is lost in the morass of mathematical intricacy that underpins the model's predictive ability.

Although the redeeming virtue of clinical intuition is said to be its capacity for integration of rich context into decision-making—that is, a doctor's 'thorough appreciation of the network of pathophysiologic causality'[280]—it is evident that a clinician cannot hope to deploy that judgement on an acceptably systematic basis to properly evaluate the veracity of this model, given its sheer complexity.[281] On the assumption that this model outperforms any predictive faculties a clinician may have honed during clinical practice—and in this small-scale study, that assumption is shown to be true—advanced training in the interpretation of a machine learning model's output does nothing to reassert the human physician's primacy. Rather, the rightness or wrongness of such a model, as clinically validated, has transcended any human ability to interrogate and contest it. All that is left, in this case, is to refuse to follow the directions of the CDSS on the irrational basis that it has sounded too many alarms. One immediate question is whether the clinical autonomy exercised in the case of alarm fatigue—the only time when the clinician's

[280] G Octo Barnett, 'The Computer and Clinical Judgment' (1982) 307 New England Journal of Medicine 493, 494. See also Randolph A Miller, 'Why the Standard View Is Standard: People, Not Machines, Understand Patients' Problems' (1990) 15 The Journal of Medicine and Philosophy 581.

[281] As discussed, clinical judgement is error-prone; 'the likelihood is high that an AI algorithm could make more accurate predictions'. See Angus (n 142) 1044. See also Norman and Eva (n 232).

decision does not follow the direction of the CDSS—is legitimate and should be preserved. Clearly, the law should not countenance an irrational failure to abide by the advice of a machine which is provably accurate—even on the basis that that advice is inexplicable—if the cost of doing so is a less optimal medical outcome.[282]

Other examples follow.

Intervention and informational data-driven CDSS

Data-driven CDSS might be classed as one of two subsets: intervention (of which the above early warning system is one example); and informational.[283] Members of the informational subset include, for example, artificial neural networks trained to classify images; two well-known studies validated convolutional neural networks (which excel particularly at image recognition) that could classify images of skin cancer and identify diabetic retinopathy in photographs of patients' retinas.[284] In those examples, one anticipated purpose of using the technology as a CDSS is to obtain information for the practitioner, and depending on who uses the information (either an expert or a clinician without the expertise that the neural network deploys), these analyses guide or affirm decision-making. However, this is not always the case, even for image classification CDSS.

[282] This is the commonly accepted logical outcome of increasing machine competence. See Price II, Gerke and Cohen (n 95); Froomkin, Kerr and Pineau (n 75).

[283] This distinction is made by other scholars: see Froomkin, Kerr and Pineau (n 75) 45; Cohen (n 97) 1426 (fn. 3).

[284] Varun Gulshan and others, 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs' (2016) 316 JAMA 2402; Esteva and others (n 124).

IDx-DR

IDx-DR,[285] a device that is also used to detect diabetic retinopathy in retinal photographs (and a commonly-cited example of artificially-intelligent technology in medicine),[286] was approved by the FDA in 2018 to be marketed as a device that can operate *independently* of expert review.[287] In its news release, the FDA said:

> IDx-DR is the first device authorized for marketing that provides a screening decision *without the need for a clinician to also interpret the image or results*, which makes it usable by health care providers who may not normally be involved in eye care.[288]

Since IDx-DR provides a screening function, it is an *intervention* data-driven CDSS: whereas image classification CDSS might otherwise be used to confirm a decision that rests with an expert, IDx-DR, an 'autonomous AI diagnostic system', makes the screening decision itself.[289]

[285] See, for example, Michael D Abràmoff and others, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 npj Digital Medicine 39.

[286] Cohen (n 97) 1429.

[287] United States Food and Drug Administration, 'FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems' (*FDA*, 11 April 2018) <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> accessed 15 July 2021. See also IDx LLC, 'A Multi-Center Study to Evaluate Performance of an Automated Device for the Detection of Diabetic Retinopathy' (ClinicalTrials.gov 2016) Clinical trial registration NCT02963441 <https://clinicaltrials.gov/ct2/show/study/NCT02963441> accessed 14 July 2021.

[288] United States Food and Drug Administration, 'FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems' (n 287) (emphasis added).

[289] Abràmoff and others (n 285) 1.

The computer software used in the device is complex.[290] Detectors for different types of tissue damage—in most cases, convolutional neural networks[291]—each look at the image, and their outputs are 'fused' using a separate machine learning algorithm that classifies the images.[292] Researchers describe this computational technique as 'clinically inspired' and 'physiologically plausible' in the sense that the diagnostic process reacts to the same visual cues as an expert, and follows the same steps from identification of biomarkers to a 'higher level clinical decision'.[293]

At first glance, this characterisation may appear to undermine the distinction I have made between knowledge and data. That is to say, here, as in the case of many computational models that perform classificatory functions after supervised learning, data is used to hone the recognition of objects that fit human ontological frameworks. However, such a system is trained on a large dataset (in this case, 'over 1 million lesion patches') to fine-tune distributed representations[294]—where the constituent properties of an object are represented by an assortment of activated units in a network, rather than one symbol—of relevant objects to give an aggregated result that is emergent from the collective behaviour of the network. The reason such machines are described as black boxes is because any explanation of their reasoning is contained in the ostensibly random activation of units that contribute in a nonlinear way to the eventual output. Although I have discouraged analogies, here I make an exception: this is akin to a human being

[290] The following description is based on ibid 4–5.

[291] For a simplified but intuitive description of convolutional neural networks, see Lawrence Carin and Michael J Pencina, 'On Deep Learning for Medical Image Analysis' (2018) 320 JAMA 1192.

[292] For the study developing and validating the fusion of detector outputs, see Meindert Niemeijer, Michael D Abramoff and Bram van Ginneken, 'Information Fusion for Diabetic Retinopathy CAD in Digital Color Fundus Photographs' (2009) 28 IEEE Transactions on Medical Imaging 775.

[293] Abràmoff and others (n 285) 4.

[294] See Hinton, McClelland and Rumelhart (n 226).

attempting to justify a decision by describing the activation and relationships between all the biological neurons engaged to process information pertinent to that decision; that description is unintelligible and completely divorced from the moral, ethical, or procedural purposes for which such justifications are sought. Although human beings have the cognitive faculties necessary to reason in language, machine reasoning (in the case of a system like IDx-DR) is split into units that do not correspond to such units of meaning.

Several other scholars have provided useful lists of studies that validate or describe data-driven applications that may soon be used in clinical practice.[295] Data-driven CDSS are increasingly being cleared by regulators for use in clinical settings,[296] and technological advancement is facilitating evolution

> from reactive systems, trained to classify patients from gold standard cases, with a measurable degree of accuracy, to proactive autonomous systems which continuously learn from experience, whose performance is judged on outcome.[297]

Thus, the regulatory approval of systems like IDx-DR typifies a trend towards outcome-based evaluation of CDSS that underpins the field of artificial intelligence broadly: reasoning is dissociated entirely from human knowledge because, as it turns out, human knowledge frameworks, limited by human cognition, hinder the making of relevant observations. Useful information emerges spontaneously from such systems, and is no

[295] Topol (n 144); Montani and Striani (n 211); Ozaydin, Hardin and Chhieng (n 203) 47–48; eBox Supplement, 'A Sampling of Health-related Applications of Deep Learning', in Naylor (n 130).

[296] Topol (n 144); Gerke, Minssen and Cohen (n 144) 297–298.

[297] Challen and others (n 207) 232.

longer obtained only at the conclusion of a causal reasoning process that admits of explanation.[298]

CONCLUSION

I have reviewed and compared typical examples of knowledge-based and data-driven CDSS and described the concepts and techniques that underpin them. This comparison reveals a critical contrast between knowledge and data that is not apparent in less detailed discussions, and highlights the overtly mechanical nature of artificially-intelligent technology to which commentators misleadingly impute human faculties. The conclusion to be drawn may be simply stated: we now have machines that reason according to nuanced information that cannot be completely expressed in human language, but which deliver verifiably accurate decisions that, in domains such as medicine, improve on human cognitive frailties even where they undermine the authority of experts within the ambit of their expertise. In such cases, causal explanations are dispensed with (being necessary only for *human* reasoning), with the consequence that machine decisions cannot be subjected to rational interrogation and must instead stand or be overturned according exclusively to the authority conferred upon them by their environment.

Such an observation, and conclusions that obtain from it, cannot be easily derived from discourse that relies on analogical descriptions of machine decision-making in human terms. Complexity does not arise from artificial 'intuition', or from the nebulous, cerebral, learned cogitation of an artificial neural network: it's just mathematics and data, and this

[298] It would be remiss not to mention the use of machine learning in **precision medicine** (loosely speaking, disease understood from the perspective of molecular biology) in the context of outcome-based practices, but unfortunately, I do not have space here to do it justice. See generally Price II (n 189); National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (The National Academies Press 2011); Euan A Ashley, 'Towards Precision Medicine' (2016) 17 Nature Reviews Genetics 507.

detailed analysis serves to disarm ambiguity and reframe machine decision-making in more realistic terms, ready to be evaluated and consciously ascribed a role that accurately reflects this characterisation.

PART III: CONSEQUENCES & QUESTIONS

In the first part of this thesis, I justified my technological inquiry as being relevant to decisions about artificially-intelligent technology that regulators and policymakers must make, and exhibited the unhelpful ambiguity that pervades existing discussion of artificial intelligence. In the second part, I conducted an extensive contextual review of artificial intelligence, the use of artificially-intelligent technology in medicine, clinical autonomy and judgement, and CDSS. I made observations arising from the implications of contrasting knowledge-based and data-drive CDSS, the most important being the implicitly mechanical nature of machine decision-making, notwithstanding aforementioned anthropomorphism; and the prevailing value of such decisions, notwithstanding the computational, inscrutable method of their determination. In short, we are dealing with complex machines that have neither substance nor personality, but which reason with insight that we should not ignore even if we cannot understand it.

It remains to consider what this means, in the context of clinical medicine, for attempts to answer the introductory interrogatory: **what happens when we give decision-making authority to artificially-intelligent technology?** The suggestion I gave as a preliminary response to this question is two-pronged: where human autonomy is the priority, such as in political deliberation, machines cannot be substituted for human decision-makers; but where the priority is some defined outcome, such as in medicine, we *should* rely on increasingly-superior machine decision-making. Since some machine decision-making is unintelligible for human beings, this necessarily implies deference to machine decisions in predefined situations. The gravity of this implication, and regulatory decisions that incorporate it, cannot be articulated by reference to anthropomorphic characterisations of artificially-intelligent technology, nor without a complete understanding of how people will relate to a particular artifact in its context. The idea of

useful decisions that are effectively beyond reproach emanating from an unaccountable entity without legal personhood is completely novel, and the law must obviously respond in novel ways, not least on account of the fact that the giving of reasons is ordinarily central to the lawful exercise of decision-making authority.

The primary objective of this thesis has been to establish the utility of looking closely at technological artifacts in their context to make meaningful observations that are not apparent on less in-depth, technology-neutral accounts. Having drawn useful conclusions, I can attempt to discuss them in terms of their consequences, although I cannot answer the secondary question—what do these observations mean?—in its totality. I sketch some parts of the existing scholarly discussion and suggest some answers in the context of liability, professional relationships, and autonomy.

LEGAL & ETHICAL ISSUES IN THE CONTEXT OF MEDICINE AND CDSS

LIABILITY

In what circumstances should we hold a physician liable for consequences arising from the use of a machine in their decision-making process? This depends crucially on the machine, its function, and its prescribed role: we can expect doctors to remain abreast of human knowledge explained in causal terms and contained in knowledge-based CDSS; but we cannot expect them to interrogate the decisions of data-driven CDSS except in limited circumstances where those decisions are demonstrably contrary to fundamental pieces of extant medical knowledge. Since data-driven CDSS are increasingly used to generate insight, rather than simply streamline or automate human processes, demonstrable error is increasingly unlikely. There are numerous components of this problem to consider—for example, is the CDSS providing a service, or is it product; and is the patient using the device, or the physician?—and I cannot do them all justice; I provide only an overview.

Issues of liability remain the arguably pre-eminent doctrinal concern of legal scholars in their consideration of artificially-intelligent technology and have been a topic of extensive discussion for over fifty years.[299] Admittedly, the legal consensus is well settled: physicians will, at some point, have a legal duty to consult diagnostic computers. In 1967, Roy Freed opined:

[299] See, for example, William B Schwartz, 'Medicine and the Computer — The Promise and Problems of Change' (1970) 283 New England Journal of Medicine 1257; Roy N Freed, 'Legal Aspects of Computer Use in Medicine' (1967) 32 Law and Contemporary Problems 674; Gerke, Minssen and Cohen (n 144); Randolph A Miller and Steven H Brown, 'Legal and Regulatory Issues Related to the Use of Clinical Software in Health Care Delivery' in Robert A Greenes (ed), *Clinical Decision Support: The Road to Broad Adoption* (2nd edn, Elsevier 2014).

> During the time in which the output of a particular laboratory diagnosis system is accepted by the profession on a par with the opinions of medical specialists, it could be an exercise of due professional care for a doctor to rely upon an erroneous opinion produced by that system, if the error was not obvious to a person with the proper degree of skill.[300]

By 2019, not much had changed:

> What happens if medical practice reaches a point where AI becomes part of the standard of care, the consensus view of good medical practice? If and when that happens, [the status quo] may change substantially: physicians may incur liability for *rejecting* correct but nonstandard AI recommendations and may conversely avoid liability for injury if they were *following* incorrect AI recommendations.[301]

This is consistent with the oft-cited opinion of Judge Learned Hand,[302] delivering the decision of the United States Court of Appeals for the Second Circuit in 1932 in *The TJ Hooper*:

> Courts must in the end say what is required; … [and] when some have thought a device necessary, at least we may say that they were right, and the others too slack.[303]

In this sense, then, the law is perfectly capable of prescribing a standard of care that mandates the use of novel technology; but the normalisation of use of a technology is obviously distinct from those upon whom liability is imposed for harm. de Dombal posed this as a 'frequent and obvious' question in 1987: 'if one uses a computer and the computer

[300] Freed (n 299) 681–2.

[301] Price II, Gerke and Cohen (n 95) 1765–1766 (footnotes omitted).

[302] This case is mentioned in Berner (n 138) 9–10; see also the electronic Data Supplement to Weed and Weed (n 153).

[303] *The TJ Hooper* (1932) 60 F2d 737 (2nd Cir).

is wrong who is to blame?'[304] Suppose that we must decide between the physician and the manufacturer. In the case of knowledge-based CDSS, the physician is clearly a learned intermediary (in the doctrinal sense of that term used in the United States),[305] capable of interrogating the output of the machine; such systems are merely 'tool[s] to support clinicians make the final decision'.[306] In the case of data-driven CDSS, and especially in systems used in precision medicine,[307] such as to tailor therapeutic decisions according to a patient's genetic profile,[308] the complexity of the system and the patient's genome may be beyond the reasonable comprehension of the physician, and only the manufacturer can ensure the reliability of the machine by clinical validation.

The clear normative answer in the latter case is that physicians should not be liable for harm arising from decisions they cannot interrogate, especially where, as in the case of precision medicine, better care can be provided by reliance on data-driven CDSS than according to ordinary, human medical knowledge. This is consistent with the understanding of the medical context that I have developed above: 'AI can provide data *and judgment*',[309] and reliance on external knowledge sources is widely expected.[310] Indeed,

---

[304] de Dombal (n 190) 182.

[305] See, for example, *Sterling Drug v Cornish* (1966) 370 F2d 82 (8th Cir).

[306] Gerke, Minssen and Cohen (n 144) 314.

[307] Price II (n 189).

[308] For a study examining the relationship between susceptibility to adverse drug reactions and a particular genotype, see Ann K Daly and others, 'HLA-B*5701 Genotype Is a Major Determinant of Drug-Induced Liver Injury Due to Flucloxacillin' (2009) 41 Nature Genetics 816. See generally Joshua C Denny, Laura K Wiley and Josh F Peterson, 'Use of Clinical Decision Support to Tailor Drug Therapy Based on Genomics' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (Springer 2016).

[309] Angus (n 142) 1043 (emphasis added).

[310] Greenes (n 182) 114; Robert R Weaver, 'Informatics Tools and Medical Communication: Patient Perspectives of "Knowledge Coupling" in Primary Care' (2003) 15 Health Communication 59; Jane Ogden and others, 'Doctors Expressions of Uncertainty and Patient Confidence' (2002) 48 Patient Education and Counseling 171.

machines were introduced for the purpose of improving unreliable decision-making and restraining clinical freedom.[311]

The legal solution facilitating this normative answer is more difficult. Should strict liability for harm be imposed upon manufacturers,[312] notwithstanding the historical reluctance of US courts in applying product liability standards to computer software? [313] Should the law mandate an ethical framework for computer scientists?[314] These issues require resolution, but are beyond the scope of this thesis.

FIDUCIARY AND PROFESSIONAL RELATIONSHIPS

With whom is the patient in a relationship for the purposes of their clinical care?

An appreciation of the mechanical nature of data-driven CDSS revealed by the taxonomy I have advanced above suggests that although it may be reasonable for such machines to make decisions, they almost certainly cannot fill the interpersonal, human role a doctor performs in addition to their strict medical functions. In this connection, Kenneth Goodman eloquently reflects:

---

[311] See, eg, JR Hampton, 'The End of Clinical Freedom' (1983) 287 British Medical Journal (Clinical research ed.) 1237.

[312] A number of articles in the 1980s discussed the possibility of strict liability being imposed on medical software developers; see Randolph A Miller, Kenneth F Schaffner and Alan Meisel, 'Ethical and Legal Issues Related to the Use of Computer Programs in Clinical Medicine' (1985) 102 Annals of Internal Medicine 529, 530–32; Vincent M Brannigan and Ruth E Dayhoff, 'Liability for Personal Injuries Caused by Defective Medical Computer Programs' (1981) 7 American Journal of Law & Medicine 123; Susan Lanoue, 'Computer Software and Strict Products Liability' (1983) 20 San Diego Law Review 439. However, In relation to the imposition of strict liability in the UK for product defects and the consequences for medical practitioners, see C Dyer, 'Product Liability Comes Closer' (1986) 293 British Medical Journal (Clinical research ed.) 1489.

[313] Miller and Brown (n 299) 717; W Nicholson Price II, 'Artificial Intelligence in Health Care: Applications and Legal Issues' (2017) 14 SciTech Lawyer 10, 13; Gerke, Minssen and Cohen (n 144) 314.

[314] Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 Nature Machine Intelligence 501.

> As long as the healing professions are practiced in a matrix of scientific uncertainty and patient values, we err if we appoint computational decision support as a surrogate for compassionate communication, shared decisions, and quality care by competent humans.[315]

Arguably, the twin offerings of medical authority and emotional support that have historically emanated from the physician may be disunited in situations where data-driven CDSS provide superior and untestable judgement. Here, again, a contextual understanding of medicine supports the view that a single doctor is no longer the omniscient arbiter. The specialisation and large-scale nature of clinical medicine practised in hospitals already requires that patients rely on multiple practitioners for their medical care, and entrust sensitive information to multiple staff; in this sense, the patient has not really been entitled to doctor-patient confidentiality of information shared with one primary physician for decades.[316] This issue is exacerbated by the advent of EHRs and national health record databases.

Confidentiality aside, the distributed nature of medical care facilitated by the use EHRs means that doctors are now less likely to follow a patient 'from their presentation through their ultimate outcome',[317] so responsibility for a patient's safety might conceivably shift from an individual to the 'health care system' of a hospital as a whole.[318] In jurisdictions that recognise the doctor-patient relationship as fiduciary in nature, such a

---

[315] Kenneth W Goodman, 'Ethical and Legal Issues in Decision Support' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016) 138.

[316] See, for example, Mark Siegler, 'Confidentiality in Medicine — A Decrepit Concept' (1982) 307 New England Journal of Medicine 1518.

[317] Danton S Char, Nigam H Shah and David Magnus, 'Implementing Machine Learning in Health Care — Addressing Ethical Challenges' (2018) 378 The New England journal of medicine 981, 982.

[318] ibid 983.

situation makes the imposition of fiduciary obligations unworkable.[319] Additionally, since a physician may not fully understand a course of action recommended by data-driven CDSS, their capacity to obtain informed consent from their patient is constrained to the information they are given themselves.[320]

Machines are not people, cannot take the Hippocratic oath, and have no intrinsic appreciation for consequences, but where the primary responsibility of the physician trends away from exercising judgement in the administration of medical care and towards providing support, their professional obligations invariably change, and it is consequently necessary to reimagine their role and responsibilities in light of machine decision-making. In light of the trend towards the consultation of external materials, the availability of large databases of medical knowledge, and the 'increasingly rigorous scientific basis of medical practice';[321] and the advent of shared decision-making in which the physician plays a more facilitative, rather than decisive role,[322] the suggestion that medical decision-making might incorporate deference to machines seems quite ordinary.

AUTONOMY AND INDEPENDENCE: THE STATUS OF MACHINES

Finally, how does the role of machines interact with human autonomy, independence, and oversight? I have suggested that the unintelligible (but useful) nature of some machine decisions made by data-driven CDSS leaves medical practitioners with

[319] ibid.

[320] Cohen (n 97). For relevant background, see generally Jonathan Pugh, 'Rational Autonomy and Decision-Making Capacity', *Autonomy, Rationality, and Contemporary Bioethics* (Oxford University Press 2020).

[321] Lachlan Forrow, Steven A Wartman and Dan W Brock, 'Science, Ethics, and the Making of Clinical Decisions: Implications for Risk Factor Intervention' (1988) 259 JAMA 3161, 3161.

[322] See generally G Elwyn and others, 'Shared Decision Making and the Concept of Equipoise: The Competences of Involving Patients in Healthcare Choices.' (2000) 50 British Journal of General Practice 892; AM Stiggelbout and others, 'Shared Decision Making: Really Putting Patients at the Centre of Healthcare' (2012) 344 BMJ e256.

little choice but to accept them. Some commentators speculate that this kind of reliance on machines—and the consequent creation of a 'collective surgical consciousness'[323] that transcends the traditional, individual expertise of medical practitioners—might 'give such tools unintended power and authority'.[324] The implications of such an eventuality are unclear, although critics have long warned of the specialist skill loss that I mentioned in the introduction to this thesis. Artificially-intelligent technology has been elevated from the mere tool typified by knowledge-based CDSS to an effectively independent third party in a decision-making process; this might serve to undermine the kind of cognitive sovereignty envisioned by references to the ethical imperative of 'human agency and oversight' which have proliferated over the last decade in regulatory treatises on 'trustworthy AI'.[325]

In the case of technologies like IDx-DR (discussed above), which are intended to operate without human oversight, it will be necessary to decide what kind of ancillary medical responsibilities should be delegated; maybe physicians should defer to data-driven CDSS in medical decision-making where machines make provably better decisions, but what about mortality risk predictions or triage decisions? In the former case, intensive care specialists already apply illness severity classification systems, such as APACHE,[326] to newly-admitted patients on the understanding that, in some cases, 'we now create

[323] Daniel A Hashimoto and others, 'Artificial Intelligence in Surgery: Promises and Perils' (2018) 268 Annals of Surgery 70, 74.

[324] Char, Shah and Magnus (n 317) 982.

[325] Consider, for example, that the first of the European Commission High-Level Expert Group on Artificial Intelligence's seven key requirements for trustworthy AI is 'human agency and oversight': see High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (European Commission 2019) 15–16.

[326] WA Knaus and others, 'APACHE II: A Severity of Disease Classification System' (1985) 13 Critical care medicine 818.

outcomes that are worse than death';[327] it is simply not desirable to 'do everything for everyone'.[328] Should unintelligible machines be imbued with the authority to decide who is better off dead?

In relation to triage decisions, consider the following example: in a 2015 study (unrelated to artificial intelligence),[329] researchers found that social interventions by clinicians were less effective for patients with severe alcohol-related problems than for patients with comparatively moderate alcohol-related problems; for the former group, 'the expensive intervention was unlikely warranted'.[330] This kind of data amounts to actionable information that data-driven CDSS may use to facilitate a cost-effective 'differential treatment response',[331] but it is not difficult to imagine the dystopian scenario in which similar prognostic analysis by data-driven CDSS results in a complete denial of care on probabilistic grounds.[332] In this scenario, unlike cases that mask algorithmic bias with proxy variables or erroneous assumptions,[333] there is no doubt of the accuracy of the statistical analysis, but being obliged to accept machine decisions at face value requires that we be satisfied with the functions that are delegated to them. This requires regulators to

[327] William Knaus, 'Ethical Implications of Risk Stratification in the Acute Care Setting' (1993) 2 Cambridge Quarterly of Healthcare Ethics 193, 193.

[328] ibid.

[329] Jesse M Hinde and others, 'The Impact of a Mandated Trauma Center Alcohol Intervention on Readmission and Cost per Readmission in Arizona' (2015) 53 Medical Care 639.

[330] Christopher J Lindsell, William W Stead and Kevin B Johnson, 'Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem' (2020) 323 JAMA 2141, 2142.

[331] ibid 2141.

[332] Triage by data-driven CDSS is prospective but likely inevitable: see Challen and others (n 207) 235.

[333] See Rich Caruana and others, 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission', *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery 2015) 1721; Ziad Obermeyer and others, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' (2019) 366 Science 447; Crystel M Gijsberts and others, 'Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events' (2015) 10 PLOS ONE <https://doi.org/10.1371/journal.pone.0132321> accessed 15 June 2021. See generally O'Neil (n 70).

balance the 'wisdom and common sense that derive from the understanding and experience of everyday human existence',[334] on one hand, against human limits: clinicians simply 'cannot work efficiently with all [they] know about a problem or all the data that could be collected'.[335] Recognising machines as capable, but neither intuitive nor human, is critical to making these kinds of informed regulatory decisions.

[334] Barnett (n 280) 493.

[335] Arthur S Elstein, 'Human Factors in Clinical Judgment: Discussion of Scriven's "Clinical Judgment"' in Hugo Tristram Engelhardt, Stuart F Spicker and Bernard Towers (eds), *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14-16, 1977* (Springer 1979) 18.

CONCLUSION

I have covered a lot of ground in the three parts of this thesis. First and foremost, I have shown that it is impractical and imprecise to refer generally to artificial intelligence, which in fact does not describe anything more specific than a general, multidisciplinary field of study and perhaps the complex machines it produces. I have shown that features (such as clinical judgement) of particular environments (medicine) in which artificially-intelligent technology is used are relevant to regulation of machine decision-making, and I have shown why and how data-driven artificially-intelligent technology reasons differently from ordinary human knowledge in such a way as to render its reasoning unintelligible, but nonetheless useful. Although the historical legal approach to regulating new technology has been to import analogous bodies of law and apply general principles in a technology-neutral manner, here that approach is inappropriate. In fact, data-driven artificially-intelligent technology, and its capacity for making useful but inexplicable decisions, has no precedent, and it is necessary to examine the feature of complex technological artifacts closely to identify just which rules should apply.

I began with a discussion of technology regulation, in which I noted the consensus of regulation scholars that 'regulation' encompasses not only the historical command and control of the state, but also the network of interactions propagated by actors in society and influenced by the environment within which activity is contained. Such a conception legitimises the close inspection of artificially-intelligent technology which, for its complexity and ubiquity, exerts a regulatory effect. I also established that the nature of artificially-intelligent technology—as capable of acting independently and producing accurate but unverifiable judgement—renders it a technological novelty, and these characteristics commend such technology to legal scholars for close inquiry prior to deciding how the law should deal with its manifestations.

In contrast to these observations, I documented a tendency among legal scholars to use general and otherwise inaccurate language to speak about technological artifacts and thus contribute to an air of mystery that precludes meaningful normative discussion of the implications of use of artificially-intelligent technology. The absence of clarity in this discourse, and the regulatory effect and novelty of artificially-intelligent technology, justify the close, circumstance-sensitive inquiry that I undertook in the second part of this thesis to elucidate relevant principles and establish a blueprint for similar analyses in other contexts.

My in-depth inquiry advanced in three increasingly focused stages, beginning with a general discussion of artificial intelligence—highlighting its inherent ambiguity—and examining the use of that term by lawmakers and in medical literature. I discussed the use of artificially-intelligent technology in medicine, and the evolution of medical practice from the primacy of individual judgement to modern reliance on evidence-based medicine, statistical methods, and machines. This context provided a backdrop for my discussion of clinical decision support systems, which included a review of various uses and related taxonomies of the technology, and the subsequent establishment of a simple and informative distinction between knowledge-based and data-driven systems. By contrasting these two types of clinical decision support systems, and with an understanding of clinical judgement, I showed that not only does this distinction lend further credence to my contention that uncritical generality is confusing and unwarranted, but that context is critical to a proper understanding of the implications of use of artificially-intelligent technology, and a technology-specific approach to regulation of artificially-intelligent technology is necessary. In particular, I discussed historical reliance on machines, and the prevailing value of deference to machine decision-making, which support my contention

that the law may need to adapt to the possibility of useful but inexplicable machine knowledge, and that such a conclusion cannot be reached from a less precise review.

Finally, I discussed the various ways in which the understanding of data-driven clinical decision support systems that I advanced—as complex but clearly non-human machines—lend insight into legal and ethical problems that we are likely to face in the context of medicine, including who should be liable for machine decision-making when doctors cannot assess the accuracy of outputs; how the ethical and legal relationship between patient and physician may need to change to incorporate machine decision-making; and what implications reliance on non-human decision-makers may have for human autonomy. This discussion reveals clear avenues for further research.

In domains in which we value accuracy, scientific progress, and the attainment of specific, process-agnostic outcomes, I suggest that we may need to come to terms with the limits of human knowledge and cognition, and embrace the supplemental value of machine decision-making, which can be accurate, reliable, and more complex than the reasoning of human beings. Reliance on machines requires the clear prior establishment of values that will guide our interaction with them, and in the ascertainment of those values, looking first to the law before understanding the target of regulation will not work. Thus, the answer to the introductory interrogatory that I tentatively propose in light of the foregoing is this: we are liable to benefit from machine insight, but only if we maintain an implicit understanding of exactly what it is that we have conferred decision-making authority upon (even if we cannot understand its reasoning). In short, we need a framework to ascribe trust to machine decision-making outcomes, and before we can trust our machines, we need to know what they are.

BIBLIOGRAPHY

Abbass HA, Scholz J and Reid DJ (eds), *Foundations of Trusted Autonomy*, vol 117 (Springer 2018)

Abràmoff MD and others, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 npj Digital Medicine 39

Agah A (ed), *Medical Applications of Artificial Intelligence* (CRC Press 2013)

Akerkar RA and Sajja PS, *Knowledge-Based Systems* (Jones and Bartlett Publishers 2010)

Anderson C, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete' [2008] *Wired* <https://www.wired.com/2008/06/pb-theory/> accessed 1 May 2021

Angus DC, 'Randomized Clinical Trials of Artificial Intelligence' (2020) 323 JAMA 1043

Angwin J and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 12 April 2021

Artificial Intelligence Research Centre of the Spanish National Research Council, 'Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe' (*IIIA CSIC*, 2017) <http://www.iiia.csic.es/barcelonadeclaration/> accessed 23 June 2021

Ashley EA, 'Towards Precision Medicine' (2016) 17 Nature Reviews Genetics 507

Baldwin R, Cave M and Lodge M, *Understanding Regulation: Theory, Strategy, and Practice* (Oxford University Press 2011)

Bargh JA and Chartrand TL, 'The Unbearable Automaticity of Being' (1999) 54 American Psychologist 462

Barnett GO, 'The Computer and Clinical Judgment' (1982) 307 New England Journal of Medicine 493

Baro E and others, 'Toward a Literature-Driven Definition of Big Data in Healthcare' [2015] BioMed Research International <https://doi.org/10.1155/2015/639021> accessed 11 June 2021

Bartlett J, *The People vs Tech: How the Internet Is Killing Democracy (and How We Save It)* (Ebury Press 2018)

Bates DW and others, 'Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors' (1998) 280 JAMA 1311

——, 'The Impact of Computerized Physician Order Entry on Medication Error Prevention' (1999) 6 Journal of the American Medical Informatics Association 313

——, 'Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-Based Medicine a Reality' (2003) 10 Journal of the American Medical Informatics Association 523

Bathaee Y, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2017) 31 Harvard Journal of Law & Technology 889

Berlin A, Sorani M and Sim I, 'Characteristics of Outpatient Clinical Decision Support Systems: A Taxonomic Description' (2004) 107 Studies in Health Technology and Informatics 578

——, 'A Taxonomic Description of Computer-Based Clinical Decision Support Systems' (2006) 39 Journal of Biomedical Informatics 656

Berner ES (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016)

Bhatnagar S and others, 'Mapping Intelligence: Requirements and Possibilities' in Vincent C Müller (ed), *Philosophy and Theory of Artificial Intelligence 2017* (Springer 2018)

Black J, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103

——, 'Critical Reflections on Regulation' (Centre for Analysis of Risk and Regulation, London School of Economics and Political Science 2002) CARR Discussion Paper 4 <http://eprints.lse.ac.uk/35985> accessed 9 August 2021

Black J, Lodge M and Thatcher M (eds), *Regulatory Innovation: A Comparative Analysis* (Edward Elgar 2005)

Black M, 'Vagueness: An Exercise in Logical Analysis' (1937) 4 Philosophy of Science 427

Blane G, *Elements of Medical Logic* (Huntington and Hopkins 1822)

Blasimme A and Vayena E, 'The Ethics of AI in Biomedical Research, Patient Care, and Public Health' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (2020)

Blass A and Gurevich Y, 'Algorithms: A Quest for Absolute Definitions' (2004) 81 Bulletin of the European Association for Theoretical Computer Science 195

Bostrom N, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014)

Bothe MK and others, 'The Use of Reinforcement Learning Algorithms to Meet the Challenges of an Artificial Pancreas' (2013) 10 Expert Review of Medical Devices 661

Brachman RJ, '(AA)AI More than the Sum of Its Parts' (2006) 27 AI Magazine 19

Brannigan VM and Dayhoff RE, 'Liability for Personal Injuries Caused by Defective Medical Computer Programs' (1981) 7 American Journal of Law & Medicine 123

Brownsword R, 'Code, Control, and Choice: Why East Is East and West Is West' (2005) 25 Legal Studies 1

——, 'The Challenge of Regulatory Connection', *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008)

——, 'The Technologies of the 21st Century: Regulatory Challenge and Regulatory Opportunity', *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008)

Buchanan BG and Duda RO, 'Principles of Rule-Based Expert Systems' in Marshall C Yovits (ed), *Advances in Computers*, vol 22 (Academic Press 1983)

Buiten MC, 'Towards Intelligent Regulation of Artificial Intelligence' (2019) 10 European Journal of Risk Regulation 41

Burrell J, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1

Bygrave LA, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' (University of Oslo Faculty of Law 2020) Research Paper 2020–35 <https://papers.ssrn.com/abstract=3721118> accessed 28 July 2021

Cabana MD and others, 'Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement' (1999) 282 JAMA 1458

Cabitza F, Rasoini R and Gensini GF, 'Unintended Consequences of Machine Learning in Medicine' (2017) 318 JAMA 517

Calo R, 'Robotics and the Lessons of Cyberlaw' (2015) 103 California Law Review 513

Campolo A and Crawford K, 'Enchanted Determinism: Power without Responsibility in Artificial Intelligence' (2020) 6 Engaging Science, Technology, and Society 1

Carin L and Pencina MJ, 'On Deep Learning for Medical Image Analysis' (2018) 320 JAMA 1192

Caruana R and others, 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission', *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery 2015)

Challen R and others, 'Artificial Intelligence, Bias and Clinical Safety' (2019) 28 BMJ Quality & Safety 231

Char DS, Shah NH and Magnus D, 'Implementing Machine Learning in Health Care — Addressing Ethical Challenges' (2018) 378 The New England journal of medicine 981

Chen JH and others, 'Decaying Relevance of Clinical Data towards Future Decisions in Data-Driven Inpatient Clinical Order Sets' (2017) 102 International Journal of Medical Informatics 71

Chopra S and White LF, 'Tort Liability for Artificial Agents', *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press 2011)

Clancey WJ, 'The Epistemology of a Rule-Based Expert System—a Framework for Explanation' (1983) 20 Artificial Intelligence 215

Clarke R, 'Regulatory Alternatives for AI' (2019) 35 Computer Law & Security Review 398

Cochrane A, *Effectiveness and Efficiency: Random Reflections on Health Services* (Nuffield Provincial Hospitals Trust 1972)

Cofone IN, 'Servers and Waiters: What Matters in the Law of A. I.' (2018) 21 Stanford Technology Law Review 167

Cohen IG, 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?' (2020) 108 Georgetown Law Journal 1425

Collingridge D, *The Social Control of Technology* (Pinter 1980)

Committee on Technology, 'Preparing for the Future of Artificial Intelligence' (National Science and Technology Council 2016) <https://obamawhitehouse.archives.gov> accessed 23 August 2021

Congressional Research Service, 'Artificial Intelligence and National Security' (2020) Report R45178 <https://crsreports.congress.gov/product/pdf/R/R45178> accessed 23 August 2021

Cortes C and Vapnik V, 'Support-Vector Networks' (1995) 20 Machine Learning 273

Daly AK and others, 'HLA-B*5701 Genotype Is a Major Determinant of Drug-Induced Liver Injury Due to Flucloxacillin' (2009) 41 Nature Genetics 816

Danaher J, 'The Threat of Algocracy: Reality, Resistance and Accommodation' (2016) 29 Philosophy & Technology 245

——, 'The Ethics of Algorithmic Outsourcing in Everyday Life' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019)

de Dombal FT, 'Ethical Considerations Concerning Computers in Medicine in the 1980s' (1987) 13 Journal of Medical Ethics 179

de Lissovoy G, 'Big Data Meets the Electronic Medical Record: A Commentary on "Identifying Patients at Increased Risk for Unplanned Readmission"' (2013) 51 Medical Care 759

Defense Innovation Board, 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense (Supporting Document)' (United States Department of Defense 2019) <https://innovation.defense.gov/recommendations> accessed 23 August 2021

DeGruy KB, 'Healthcare Applications of Knowledge Discovery in Databases' (2000) 14 Journal of Healthcare Information Management 59

Denny JC, Wiley LK and Peterson JF, 'Use of Clinical Decision Support to Tailor Drug Therapy Based on Genomics' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (Springer 2016)

Department for Business, Energy & Industrial Strategy, 'Industrial Strategy: Building a Britain Fit for the Future' (Government of the United Kingdom 2018) Policy Paper

<https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future> accessed 23 August 2021

Dharmarajan K and others, 'Treatment for Multiple Acute Cardiopulmonary Conditions Among Older Patients Hospitalized with Pneumonia, Chronic Obstructive Pulmonary Disease, or Heart Failure' (2016) 64 Journal of the American Geriatrics Society 1574

Dietrich E, 'Algorithm' in Robert Andrew Wilson and Frank C Keil (eds), *The MIT Encyclopedia of the Cognitive Sciences* (MIT Press 1999) 11

Domingos P, 'A Few Useful Things to Know about Machine Learning' (2012) 55 Communications of the ACM 78

Dufour N and others, 'When a Ventilator Takes Autonomous Decisions without Seeking Approbation nor Warning Clinicians: A Case Series' (2020) 13 International Medical Case Reports Journal 521

Durkin J, 'History and Applications' in Cornelius T Leondes (ed), *Expert Systems: The Technology of Knowledge Management and Decision Making for the 21st Century* (Academic Press 2002)

Dyer C, 'Product Liability Comes Closer' (1986) 293 British Medical Journal (Clinical research ed.) 1489

Easterbrook FH, 'Cyberspace and the Law of the Horse' (1996) 1996 University of Chicago Legal Forum 207

Eddy DM, 'Variations in Physician Practice: The Role of Uncertainty' (1984) 3 Health Affairs 74

——, 'Practice Policies—Guidelines for Methods' (1990) 263 JAMA 1839

——, *Clinical Decision Making: From Theory to Practice: A Collection of Essays from JAMA* (Jones and Bartlett Publishers 1996)

Eddy DM and Billings J, 'The Quality of Medical Evidence: Implications for Quality of Care' (1988) 7 Health Affairs 19

Editorial, 'Mechanism Matters' (2010) 16 Nature Medicine 347

El Naqa I, 'Computational Learning Theory' in Issam El Naqa, Ruijiang Li and Martin J Murphy (eds), *Machine Learning in Radiation Oncology: Theory and Applications* (Springer 2015)

El Naqa I and Murphy MJ, 'What Is Machine Learning?' in Issam El Naqa, Ruijiang Li and Martin J Murphy (eds), *Machine Learning in Radiation Oncology: Theory and Applications* (Springer 2015)

Elstein AS, 'Clinical Judgment: Psychological Research and Medical Practice' (1976) 194 Science 696

Elstein AS, 'Human Factors in Clinical Judgment: Discussion of Scriven's "Clinical Judgment"' in Hugo Tristram Engelhardt, Stuart F Spicker and Bernard Towers (eds),

Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14-16, 1977 (Springer 1979)

Elwyn G and others, 'Shared Decision Making and the Concept of Equipoise: The Competences of Involving Patients in Healthcare Choices.' (2000) 50 British Journal of General Practice 892

Engelhardt HT, Spicker SF and Towers B, *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14-16, 1977* (Springer 1979)

Esteva A and others, 'Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks' (2017) 542 Nature 115

European Commission, 'Artificial Intelligence for Europe' (2018) Communication COM(2018) 237 final

——, 'Coordinated Plan on Artificial Intelligence' (2018) Communication COM(2018) 795 final

——, 'On Artificial Intelligence — A European Approach to Excellence and Trust' (2020) White Paper COM(2020) 65 final

——, 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' Proposal for a Regulation of the European Parliament and of the Council COM(2021) 206 final (Procedure 2021/0106/COD) <https://eur-lex.europa.eu/procedure/EN/2021_106> accessed 23 August 2021

European Parliament, 'European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics' (European Parliament 2017) P8_TA(2017)0051 <https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html> accessed 18 August 2021

European Parliamentary Research Service, 'Civil Liability Regime for Artificial Intelligence: European Added Value Assessment' (European Parliament 2020) Study <https://doi.org/10.2861/737677> accessed 23 August 2021

——, 'European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies (European Added Value Assessment)' (European Parliament 2020) Study <https://doi.org/10.2861/94107> accessed 23 August 2021

Eva KW, 'What Every Teacher Needs to Know about Clinical Reasoning' (2005) 39 Medical Education 98

Feigenbaum E, 'Knowledge Engineering For The 1980s' (1982) <https://exhibits.stanford.edu/feigenbaum/catalog/bf210yv9971> accessed 19 April 2021

Feinstein A, *Clinical Judgment* (Williams & Wilkins 1967)

Ford C, *Innovation and the State: Finance, Regulation, and Justice* (Cambridge University Press 2017)

Forgionne GA and Kohli R, 'HMSS: A Management Support System for Concurrent Hospital Decision Making' (1996) 16 Decision Support Systems 209

Forrow L, Wartman SA and Brock DW, 'Science, Ethics, and the Making of Clinical Decisions: Implications for Risk Factor Intervention' (1988) 259 JAMA 3161

Frankovich J, Longhurst CA and Sutherland SM, 'Evidence-Based Medicine in the EMR Era' (2011) 365 New England Journal of Medicine 1758

Freed RN, 'Legal Aspects of Computer Use in Medicine' (1967) 32 Law and Contemporary Problems 674

Froomkin AM, Kerr I and Pineau J, 'When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning' (2019) 61 Arizona Law Review 33

Galanter WL, DiDomenico RJ and Polikaitis A, 'Preventing Exacerbation of an ADE with Automated Decision Support' (2002) 16 Journal of Healthcare Information Management 44

Gerke S, Minssen T and Cohen G, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Adam Bohr and Kaveh Memarzadeh (eds), *Artificial Intelligence in Healthcare* (Academic Press 2020)

Gevarter WB, 'Expert Systems: Artificial Intelligence Applied' (1984) 1 Telematics and Informatics 239

Gijsberts CM and others, 'Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events' (2015) 10 PLOS ONE <https://doi.org/10.1371/journal.pone.0132321> accessed 15 June 2021

Goldberg LR, 'Simple Models or Simple Processes? Some Research on Clinical Judgments' (1968) 23 American Psychologist 483

Goodfellow I, Bengio Y and Courville A, *Deep Learning* (MIT Press 2016) <https://www.deeplearningbook.org/>

Goodman KW, 'Ethical and Legal Issues in Decision Support' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016)

Goodman RL, 'Commentary: Health Care Technology and Medical Education: Putting Physical Diagnosis in Its Proper Place' (2010) 85 Academic Medicine 945

Greenes RA (ed), *Clinical Decision Support: The Road to Broad Adoption* (Elsevier 2014)

Gulshan V and others, 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs' (2016) 316 JAMA 2402

Guzman A, 'Ontological Boundaries between Humans and Computers and the Implications for Human-Machine Communication' (2020) 1 Human-Machine Communication

Haenlein M and Kaplan A, 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) 61 California Management Review 5

Haidt J, 'The New Synthesis in Moral Psychology' (2007) 316 Science 998

——, 'Moral Psychology and the Law: How Intuitions Drive Reasoning, Judgment, and the Search for Evidence' (2012) 64 Alabama Law Review 867

Hamet P and Tremblay J, 'Artificial Intelligence in Medicine' (2017) 69 Metabolism S36

Hampton JR, 'The End of Clinical Freedom' (1983) 287 British Medical Journal (Clinical research ed.) 1237

Hand DJ, 'Mining Medical Data' (2000) 9 Statistical Methods in Medical Research 305

Harper EM, 'The Economic Value of Health Care Data' (2013) 37 Nursing Administration Quarterly 105

Hashimoto DA and others, 'Artificial Intelligence in Surgery: Promises and Perils' (2018) 268 Annals of Surgery 70

Hatib F and others, 'Machine-Learning Algorithm to Predict Hypotension Based on High-Fidelity Arterial Pressure Waveform Analysis' (2018) 129 Anesthesiology 663

He J and others, 'The Practical Implementation of Artificial Intelligence Technologies in Medicine' (2019) 25 Nature Medicine 30

High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (European Commission 2019)

Hildebrandt M, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar 2016)

——, 'The Artificial Intelligence of European Union Law' (2020) 21 German Law Journal 74

Hinde JM and others, 'The Impact of a Mandated Trauma Center Alcohol Intervention on Readmission and Cost per Readmission in Arizona' (2015) 53 Medical Care 639

Hinton G, 'Deep Learning—A Technology With the Potential to Transform Health Care' (2018) 320 JAMA 1101

Hinton GE, McClelland JL and Rumelhart DE, 'Distributed Representations' in David E Rumelhart, James L McClelland, and PDP Research Group (eds), *Parallel distributed processing: explorations in the microstructure of cognition*, vol 1 (MIT Press 1986)

Hoffer EP and others, 'DXplain: Patterns of Use of a Mature Expert System' (2005) 2005 AMIA Annual Symposium Proceedings 321

Hofstadter D, *Gödel, Escher, Bach: An Eternal Golden Braid* (Penguin Books 1980)

Holt RR, 'Yet Another Look at Clinical and Statistical Prediction: Or, Is Clinical Psychology Worthwhile?' (1970) 25 American Psychologist 337

House of Lords Select Committee on Artificial Intelligence, 'AI in the UK: Ready, Willing and Able? Report of Session 2017-19' (House of Lords Select Committee on Artificial Intelligence 2018)

IDx LLC, 'A Multi-Center Study to Evaluate Performance of an Automated Device for the Detection of Diabetic Retinopathy' (ClinicalTrials.gov 2016) Clinical trial registration NCT02963441 <https://clinicaltrials.gov/ct2/show/study/NCT02963441> accessed 14 July 2021

Independent High-Level Expert Group on Artificial Intelligence, 'A Definition of AI: Main Capabilities and Disciplines' (European Commission 2019) <https://digital-strategy.ec.europa.eu/en/library> accessed 23 August 2021

Institute of Medicine, *To Err Is Human: Building a Safer Health System* (The National Academies Press 2000) <https://www.nap.edu/catalog/9728/to-err-is-human-building-a-safer-health-system>

International Center for Scientific Debate, 'Artificial Intelligence: Dreams, Risks, and Reality' (*B-DEBATE*, 2017) <https://www.bdebate.org/en/forum/artificial-intelligence-next-step-evolution> accessed 23 June 2021

Kahn PH and others, 'The New Ontological Category Hypothesis in Human-Robot Interaction', *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2011) <https://doi.org/10.1145/1957656.1957710> accessed 5 August 2021

Kamarinou D, Millard C and Singh J, 'Machine Learning with Personal Data' (Queen Mary University of London, School of Law 2016) Legal Studies Research Paper 247/2016 <https://papers.ssrn.com/abstract=2865811> accessed 29 July 2021

Katyal SK, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 UCLA Law Review 54

Kienle GS and Kiene H, 'Clinical Judgement and the Medical Profession' (2011) 17 Journal of Evaluation in Clinical Practice 621

Kinstler L, 'Opinion: Can Silicon Valley Find God?' *The New York Times* (16 July 2021) <https://www.nytimes.com/interactive/2021/07/16/opinion/ai-ethics-religion.html> accessed 27 July 2021

Knaus W, 'Ethical Implications of Risk Stratification in the Acute Care Setting' (1993) 2 Cambridge Quarterly of Healthcare Ethics 193

Knaus WA and others, 'APACHE II: A Severity of Disease Classification System' (1985) 13 Critical care medicine 818

Knight W, 'Alpha Zero's "Alien" Chess Shows the Power, and the Peculiarity, of AI' [2017] *MIT Technology Review* <https://www.technologyreview.com/2017/12/08/147199> accessed 15 August 2021

Koops B-J, 'Ten Dimensions of Technology Regulation: Finding Your Bearings in the Research Space of an Emerging Discipline' in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Technology Regulation* (Wolf Legal Publishers 2010)

Kowert W, 'The Foreseeability of Human-Artificial Intelligence Interactions' (2017) 96 Texas Law Review 181

Krittanawong C and others, 'Artificial Intelligence in Precision Cardiovascular Medicine' (2017) 69 Journal of the American College of Cardiology 2657

Krumholz HM, 'Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System' (2014) 33 Health Affairs 1163

Kunhimangalam R, Ovallath S and Joseph PK, 'A Clinical Decision Support System with an Integrated EMR for Diagnosis of Peripheral Neuropathy' (2014) 38 Journal of Medical Systems 38

Kuperman GJ and others, 'Detecting Alerts, Notifying the Physician, and Offering Action Items: A Comprehensive Alerting System.' [1996] Proceedings of the AMIA Annual Fall Symposium 704

——, 'Medication-Related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review' (2007) 14 Journal of the American Medical Informatics Association 29

Lanoue S, 'Computer Software and Strict Products Liability' (1983) 20 San Diego Law Review 439

LeCun Y, Bengio Y and Hinton G, 'Deep Learning' (2015) 521 Nature 436

Lessig L, *Code and Other Laws of Cyberspace* (Basic Books 1999)

Lindsay RK and others, *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project* (McGraw-Hill 1980)

Lindsell CJ, Stead WW and Johnson KB, 'Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem' (2020) 323 JAMA 2141

Lovell RRH and Doyle AE, *An Introduction to Clinical Medicine* (2nd edn, Melbourne University Press 1971)

Lum K and Isaac W, 'To Predict and Serve?' (2016) 13 Significance 14

Maxmen JS, *The Post-Physician Era: Medicine in the Twenty-First Century* (John Wiley and Sons 1976)

McCorduck P, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (A K Peters Ltd 2004)

Meehl PE, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press 1954)

Metzger J and MacDonald K, *Clinical Decision Support for the Independent Physician Practice* (California Healthcare Foundation 2003) <https://www.chcf.org/publication/clinical-decision-support-for-the-independent-physician-practice/> accessed 4 May 2021

Millar J and Kerr IR, 'Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots' (University of Ottawa 2013) SSRN Scholarly Paper <https://papers.ssrn.com/abstract=2234645> accessed 27 January 2022

Miller AP, 'Want Less-Biased Decisions? Use Algorithms.' [2018] *Harvard Business Review* <https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms> accessed 3 August 2021

Miller RA, 'Why the Standard View Is Standard: People, Not Machines, Understand Patients' Problems' (1990) 15 The Journal of Medicine and Philosophy 581

Miller RA and Brown SH, 'Legal and Regulatory Issues Related to the Use of Clinical Software in Health Care Delivery' in Robert A Greenes (ed), *Clinical Decision Support: The Road to Broad Adoption* (2nd edn, Elsevier 2014)

Miller RA and Masarie Jr. FE, 'The Demise of the "Greek Oracle" Model for Medical Diagnostic Systems' (1990) 29 Methods of Information in Medicine 1

Miller RA, Pople Jr. HE and Myers JD, 'Internist-I, an Experimental Computer-Based Diagnostic Consultant for General Internal Medicine' (1982) 307 The New England Journal of Medicine 468

Miller RA, Schaffner KF and Meisel A, 'Ethical and Legal Issues Related to the Use of Computer Programs in Clinical Medicine' (1985) 102 Annals of Internal Medicine 529

Mitchell TM, *Machine Learning* (McGraw-Hill 1997)

Mittelstadt B, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 Nature Machine Intelligence 501

Monett D and others, 'Special Issue "On Defining Artificial Intelligence"—Commentaries and Author's Response' (2020) 11 Journal of Artificial General Intelligence 1

Montani S and Striani M, 'Artificial Intelligence in Clinical Decision Support: A Focused Literature Survey' (2019) 28 Yearbook of Medical Informatics 120

Morgan B and Yeung K, *An Introduction to Law and Regulation: Text and Materials* (Cambridge University Press 2007)

Moses LB, 'Regulating beyond Nanotechnology: Do Nano-Specific Problems Require Nano-Specific Solutions?', *2010 IEEE International Symposium on Technology and Society* (2010)

——, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5 Law, Innovation and Technology 1

Murray A and Scott C, 'Controlling the New Media: Hybrid Responses to New Forms of Power' (2002) 65 The Modern Law Review 491

National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (The National Academies Press 2011)

Naylor CD, 'On the Prospects for a (Deep) Learning Health Care System' (2018) 320 JAMA 1099

Networking and Information Technology Research and Development Subcommittee, 'National Artificial Intelligence Research and Development Strategic Plan' (National Science and Technology Council 2016) <https://obamawhitehouse.archives.gov> accessed 23 August 2021

Newell A and Simon HA, 'Computer Science as Empirical Inquiry: Symbols and Search' (1976) 19 Communications of the ACM 113

Niemeijer M, Abramoff MD and van Ginneken B, 'Information Fusion for Diabetic Retinopathy CAD in Digital Color Fundus Photographs' (2009) 28 IEEE Transactions on Medical Imaging 775

Nilsson NJ, *Introduction to Machine Learning: An Early Draft of a Proposed Textbook* (Robotics Laboratory, Department of Computer Science, Stanford University 1998) <https://robotics.stanford.edu/people/nilsson/mlbook.html> accessed 10 July 2021

Nilsson NJ, *Artificial Intelligence: A New Synthesis* (Morgan Kaufmann 1998)

Norman G, Young M and Brooks L, 'Non-Analytical Models of Clinical Reasoning: The Role of Experience' (2007) 41 Medical Education 1140

Norman GR and Eva KW, 'Diagnostic Error and Clinical Reasoning' (2010) 44 Medical Education 94

Obermeyer Z and others, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' (2019) 366 Science 447

Ofri D, 'The Practice of Medicine: Neither Science nor Art' (2006) 367 The Lancet 807

Ogden J and others, 'Doctors Expressions of Uncertainty and Patient Confidence' (2002) 48 Patient Education and Counseling 171

Omohundro SM, 'The Basic AI Drives', *Proceedings of the First AGI Conference (Frontiers in Artificial Intelligence and Applications)* (2008)

O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2017)

Oquendo MA and others, 'Machine Learning and Data Mining: Strategies for Hypothesis Generation' (2012) 17 Molecular Psychiatry 956

Osborne D and Gaebler T, *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* (Addison-Wesley 1992)

Osheroff J and others, *Improving Outcomes with Clinical Decision Support: An Implementer's Guide* (2nd edn, CRC Press 2012)

Ozaydin B, Hardin JM and Chhieng DC, 'Data Mining and Clinical Decision Support Systems' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (3rd edn, Springer 2016)

Pasquale F, 'A Rule of Persons, Not Machines: The Limits of Legal Automation' (2019) 87 George Washington Law Review 1

Patel VL and others, 'The Coming of Age of Artificial Intelligence in Medicine' (2009) 46 Artificial Intelligence in Medicine 5

Patel VL, Arocha JF and Kaufman DR, 'Expertise and Tacit Knowledge in Medicine', *Tacit Knowledge in Professional Practice* (Psychology Press 1999)

Patel VL, Evans DA and Groen GJ, 'Biomedical Knowledge and Clinical Reasoning' in David A Evans and Vimla L Patel (eds), *Cognitive Science in Medicine* (MIT Press 1989)

Paterno MD and others, 'Tiering Drug–Drug Interaction Alerts by Severity Increases Compliance Rates' (2009) 16 Journal of the American Medical Informatics Association 40

Payne TH and others, 'Preparation and Use of Preconstructed Orders, Order Sets, and Order Menus in a Computerized Provider Order Entry System' (2003) 10 Journal of the American Medical Informatics Association 322

Pearl J and Mackenzie D, *The Book of Why: The New Science of Cause and Effect* (Penguin Books 2019)

Pierre J and Peters BG, *Governing Complex Societies: Trajectories and Scenarios* (Palgrave Macmillan 2005)

Pinsky MR, 'Complexity Modeling: Identify Instability Early' (2010) 38 Critical Care Medicine S649

Poole D, Mackworth A and Randy G, *Computational Intelligence: A Logical Approach* (Oxford University Press 1998)

Prasad N and others, 'A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units' [2017] arXiv:1704.06300 [cs] <http://arxiv.org/abs/1704.06300> accessed 19 June 2021

Price II WN, 'Black-Box Medicine' (2015) 28 Harvard Journal of Law & Technology 419

——, 'Artificial Intelligence in Health Care: Applications and Legal Issues' (2017) 14 SciTech Lawyer 10

Price II WN, Gerke S and Cohen IG, 'Potential Liability for Physicians Using Artificial Intelligence' (2019) 322 JAMA 1765

Pugh J, 'Rational Autonomy and Decision-Making Capacity', *Autonomy, Rationality, and Contemporary Bioethics* (Oxford University Press 2020)

Rambachan A and Roth J, 'Bias In, Bias Out? Evaluating the Folk Wisdom' [2020] arXiv:1909.08518 [cs.LG] <http://arxiv.org/abs/1909.08518> accessed 4 August 2021

Ramesh AN and others, 'Artificial Intelligence in Medicine' (2004) 86 Annals of the Royal College of Surgeons of England 334

Reider JM, 'Impact of National Policies on the Use of Clinical Decision Support' in Eta S Berner (ed), *Clinical Decision Support Systems: Theory and Practice* (Springer 2016)

Rogers H, *Theory of Recursive Functions and Effective Computability* (1st MIT Press paperback edn, MIT Press 1987)

Rose N, *Powers of Freedom: Reframing Political Thought* (Cambridge University Press 1999)

Ross TJ, *Fuzzy Logic with Engineering Applications* (2nd edn, John Wiley & Sons 2005)

Russell B, *The Collected Papers of Bertrand Russell Volume 6: Logical and Philosophical Papers 1909-13* (John G Slater ed, Routledge 1992)

Russell S, 'Provably Beneficial Artificial Intelligence', *The Next Step: Exponential Life* (BBVA OpenMind 2017)

Russell S and Norvig P, *Artificial Intelligence: A Modern Approach* (3rd edn, Pearson Education 2016)

Sackett DL and others, 'Evidence Based Medicine: What It Is and What It Isn't' (1996) 312 BMJ 71

Samuel AL, 'Some Studies in Machine Learning Using the Game of Checkers' (1959) 3 IBM Journal of Research and Development 210

Schaffer J, 'The Metaphysics of Causation' in Edwards N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Fall 2016 edition, Metaphysics Research Lab, Stanford University 2016) <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics> accessed 15 July 2021

Scharre P, *Army of None: Autonomous Weapons and the Future of War* (WW Norton & Company 2019)

Schwartz WB, 'Medicine and the Computer — The Promise and Problems of Change' (1970) 283 New England Journal of Medicine 1257

Scott C, 'Regulation in the Age of Governance: The Rise of the Post-Regulatory State' in Jacint Jordana and David Levi-Faur (eds), *The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance* (Edward Elgar 2004)

Scriven M, 'Clinical Judgment' in Hugo Tristram Engelhardt, Stuart F Spicker and Bernard Towers (eds), *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14–16, 1977* (Springer Netherlands 1979)

Selbst AD, 'Negligence and AI's Human Users' (2020) 100 Boston University Law Review 1315

Selbst AD and Barocas S, 'The Intuitive Appeal of Explainable Machines' (2018) 87 Fordham Law Review 1085

Sherwin E, 'A Defense of Analogical Reasoning in Law' (1999) 66 University of Chicago Law Review 1179

Shortliffe EH and Buchanan BG (eds), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison Wesley 1984) <http://www.shortliffe.net> accessed 13 April 2021

Shortliffe EH and Sepúlveda MJ, 'Clinical Decision Support in the Era of Artificial Intelligence' (2018) 320 JAMA 2199

Siegler M, 'Confidentiality in Medicine — A Decrepit Concept' (1982) 307 New England Journal of Medicine 1518

Silver D and others, 'Mastering the Game of Go without Human Knowledge' (2017) 550 Nature 354

Sim I and others, 'Clinical Decision Support Systems for the Practice of Evidence-Based Medicine' (2001) 8 Journal of the American Medical Informatics Association 527

Sim I and Berlin A, 'A Framework for Classifying Decision Support Systems' (2003) 2003 AMIA Annual Symposium Proceedings 599

Simon SR and others, 'Physicians and Electronic Health Records: A Statewide Survey' (2007) 167 Archives of Internal Medicine 507

Sittig DF and others, 'Grand Challenges in Clinical Decision Support' (2008) 41 Journal of Biomedical Informatics 387

Sommer JH, 'Against Cyberlaw' (2000) 15 Berkeley Technology Law Journal 1145

Spengler PM and others, 'The Meta-Analysis of Clinical Judgment Project: Effects of Experience on Judgment Accuracy' (2007) 37 The Counseling Psychologist 350

Spooner SA, 'Mathematical Foundations of Decision Support Systems', *Clinical Decision Support Systems: Theory and Practice* (Springer 2016)

Steels L and Lopez de Mantaras R, 'The Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe' (2018) 31 AI Communications 485

Stiggelbout AM and others, 'Shared Decision Making: Really Putting Patients at the Centre of Healthcare' (2012) 344 BMJ e256

Stonier T, 'Machine Intelligence and the Long-Term Future of the Human Species' (1988) 2 AI & Society 133

——, 'The Spectrum of Intelligence' in Tom Stonier (ed), *Beyond Information: The Natural History of Intelligence* (Springer 1992) <https://doi.org/10.1007/978-1-4471-1835-0_2> accessed 24 June 2021

Surden H, 'Artificial Intelligence and Law: An Overview' (2019) 35 Georgia State University Law Review 1305

Sutton RT and others, 'An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success' (2020) 3 npj Digital Medicine 1

Szolovits P (ed), *Artificial Intelligence in Medicine* (Westview Press 1982) <https://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch0.html> accessed 25 April 2021

Szolovits P and Pauker SG, 'Categorical and Probabilistic Reasoning in Medical Diagnosis' (1978) 11 Artificial Intelligence 115

Taplin J, *Move Fast and Break Things: How Facebook, Google and Amazon Have Cornered Culture and What It Means for All of Us* (Macmillan 2017)

Tesler L, 'Adages and Coinages' (*Larry Tesler Consulting*) <http://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html> accessed 2 May 2021

Topol EJ, 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25 Nature Medicine 44

Tversky A and Kahneman D, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185 Science 1124

United States Food and Drug Administration, 'FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems' (*FDA*, 11 April 2018) <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> accessed 15 July 2021

——, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)' (2019) Discussion Paper <https://www.fda.gov/media/122535/download> accessed 20 August 2021

——, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (*FDA*, 12 January 2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 18 July 2021

US Food and Drug Administration, 'Software as a Medical Device (SAMD): Clinical Evaluation' (8 December 2017) <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation> accessed 3 May 2021

Waldmann MR and Hagmayer Y, 'Causal Reasoning', *The Oxford Handbook of Cognitive Psychology* (2013)

Wang P, 'On Defining Artificial Intelligence' (2019) 10 Journal of Artificial General Intelligence 1

Weaver RR, 'Informatics Tools and Medical Communication: Patient Perspectives of "Knowledge Coupling" in Primary Care' (2003) 15 Health Communication 59

Weed LL and Weed L, 'Opening the Black Box of Clinical Judgment—an Overview' (1999) 319 BMJ 1279

Weinberger D, 'Our Machines Now Have Knowledge We'll Never Understand' [2017] *Wired* <https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/> accessed 17 May 2021

Wennberg J and Gittelsohn A, 'Small Area Variations in Health Care Delivery: A Population-Based Health Information System Can Guide Planning and Regulatory Decision-Making' (1973) 182 Science 1102

Wiggins JS, *Personality and Prediction: Principles of Personality Assessment* (Addison-Wesley 1973)

Wijnberge M and others, 'The Use of a Machine-Learning Algorithm That Predicts Hypotension during Surgery in Combination with Personalized Treatment Guidance: Study Protocol for a Randomized Clinical Trial' (2019) 20 Trials 582

Wijnberge M and others, 'Effect of a Machine Learning–Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial' (2020) 323 JAMA 1052

Wilken M and others, 'Alarm Fatigue: Causes and Effects' [2017] Studies in Health Technology and Informatics 107

Wilkinson IB and others, 'Thinking about Medicine', *Oxford Handbook of Clinical Medicine* (Oxford University Press 2017)

Wilson T, Holt T and Greenhalgh T, 'Complexity and Clinical Care' (2001) 323 BMJ 685

Winfield A, 'Intelligence Is Not One Thing' (2020) 11 Journal of Artificial General Intelligence 97

Wright A and others, 'Development and Evaluation of a Comprehensive Clinical Decision Support Taxonomy: Comparison of Front-End Tools in Commercial and Internally Developed Electronic Health Record Systems' (2011) 18 Journal of the American Medical Informatics Association 232

Wu T, 'The Tyranny of Convenience' *The New York Times* (16 February 2018) <https://www.nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html> accessed 4 August 2021

Yeung K, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118

——, 'Algorithmic Regulation: A Critical Interrogation' (2018) 12 Regulation & Governance 505

Yeung K and Lodge M, 'Algorithmic Regulation: An Introduction' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019)

Young FE, 'Validation of Medical Software: Present Policy of the Food and Drug Administration' (1987) 106 Annals of Internal Medicine 628

Yudkowsky E, 'Artificial Intelligence as a Positive and Negative Factor in Global Risk' in Nick Bostrom and Milan M Ćirković (eds), *Global Catastrophic Risks* (Oxford University Press 2008)

——, 'Cognitive Biases Potentially Affecting Judgement of Global Risks' in Nick Bostrom and Milan M Cirkovic (eds), *Global Catastrophic Risks* (Oxford University Press 2008)

Zadeh LA, 'Foreword' in Timothy J Ross, Jane M Booker and WJ Parkinson (eds), *Fuzzy Logic and Probability Applications: Bridging the Gap* (SIAM 2002)

Zuboff S, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019)

Sterling Drug v Cornish (1966) 370 F2d 82 (8th Cir)

The TJ Hooper (1932) 60 F2d 737 (2nd Cir)

Pub L No. 115-232, 132 Stat 1636, 1695 (13 August 2018) (codified at 10 USC § 2358)

21 USC § 360c