



NUS Law Working Paper 2020/025

Artificial Intelligence And The Limits Of Legal Personality

Simon Chesterman

chesterman@nus.edu.sg

[August 2020]

This paper can be downloaded without charge at the National University of Singapore, Faculty of Law Working Paper Series index: <http://law.nus.edu.sg/wps/>

© Copyright is held by the author or authors of each working paper. No part of this paper may be republished, reprinted, or reproduced in any format without the permission of the paper's author or authors.

Note: The views expressed in each paper are those of the author or authors of the paper. They do not necessarily represent or reflect the views of the National University of Singapore.

Citations of this electronic publication should be made in the following manner: Author, "Title," NUS Law Working Paper Series, "Paper Number", Month & Year of publication, <http://law.nus.edu.sg/wps>. For instance, Chan, Bala, "A Legal History of Asia," NUS Law Working Paper 2014/001, January 2014, www.law.nus.edu.sg/wps/

ARTIFICIAL INTELLIGENCE AND THE LIMITS OF LEGAL PERSONALITY

SIMON CHESTERMAN*

Abstract As artificial intelligence (AI) systems become more sophisticated and play a larger role in society, arguments that they should have some form of legal personality gain credence. The arguments are typically framed in instrumental terms, with comparisons to juridical persons such as corporations. Implicit in those arguments, or explicit in their illustrations and examples, is the idea that as AI systems approach the point of indistinguishability from humans they should be entitled to a status comparable to natural persons. This article contends that although most legal systems could create a novel category of legal persons, such arguments are insufficient to show that they should.

Keywords: artificial intelligence, legal personality, natural person, juridical person, corporation, corporate veil, copyright, patent.

In 2021 the Bank of England will complete its transition from paper to polymer with the release of a new fifty pound note. A public selection process saw almost quarter of a million nominations for the face of the new note; the final decision, announced in July 2019, was that Alan Turing would be featured. Turing was a hero for his codebreaking during the Second World War. He also helped establish the discipline of computer science, laying the foundations for what we now call artificial intelligence (AI).¹ Perhaps his best known

* BA/LLB (Hons) (Melbourne), DPhil (Oxford). Dean and Provost's Chair Professor, National University of Singapore Faculty of Law. I am deeply grateful to Damian Chalmers, Hu Ying, Kenneth Khoo, Mark Lim, Lin Lin, Daniel Seng, David Tan, Tan Zhong Xing, Hans Tjio, Umakanth Varottil, Wong Chee Leong, and two anonymous reviewers for their comments on earlier versions of this text. Invaluable research assistance was provided by Violet Huang, Eugene Lau, Ong Kye Jing, and Yap Jia Qing. Errors and omissions are due to the author alone.

¹ For a discussion of attempts to define AI, see SJ Russell and P Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 1-5. Four broad approaches can be identified: acting humanly (the Turing Test), thinking humanly (modelling cognitive behaviour), thinking rationally (building on the logicist tradition), and

contribution, however, is the eponymous test for when true ‘intelligence’ has actually been achieved.

Turing modelled it on a parlour game popular in 1950. A man and a woman sit in a separate room and provide written answers to questions; the other participants have to guess who provided which answer. Turing posited that a similar ‘imitation game’ could be played with a computer. When a machine could fool people into believing that it was human, we might properly say that it was intelligent.²

Early successes along these lines came in the 1960s with programs like Eliza. Users were told that Eliza was a psychotherapist who communicated through words typed into a computer. In fact, ‘she’ was an algorithm using a simple list-processing language. If the user typed in a recognized phrase, it would be reframed as a question. So after entering ‘I’m depressed,’ Eliza might reply ‘Why do you say that you are depressed?’ If it didn’t recognize the phrase, the program would offer something generic, like ‘Can you elaborate on that?’ Even when they were told how it worked, some users insisted that Eliza had ‘understood’ them.³

Parlour games aside, why should it matter if a computer is ‘intelligent’? For several decades, the Turing Test was associated more with the question of whether AI itself was possible, rather than the legal status of such an entity. Yet it is commonly invoked in discussions of legal personality for AI, from Lawrence Solum’s seminal 1992 article onwards.⁴ Though no longer regarded as a serious measure of modern AI in a technical sense, the Turing Test’s longevity as a trope points to a tension in debates over personality that is often overlooked.

As AI systems become more sophisticated and play a larger role in society, there are at least two discrete reasons why they might be recognized as persons before the law. The first is so that there is someone to blame when things go wrong. This is presented as the answer to potential accountability gaps created by their speed, autonomy, and opacity.⁵ A second reason for recognizing personality, however, is to ensure that there is someone to reward

acting rationally (a rational-agent approach favoured by Russell and Norvig as it is not dependent on a specific understanding of human cognition or an exhaustive model of what constitutes rational thought).

² AM Turing, ‘Computing Machinery and Intelligence’ (1950) 59 *Mind* 433.

³ RS Wallace, ‘The Anatomy of ALICE’ in R Epstein, G Roberts, and G Beber (eds), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (Springer 2009) 184-85.

⁴ LB Solum, ‘Legal Personhood for Artificial Intelligences’ (1992) 70 *North Carolina L Rev* 1235-37. Solum himself credits Christopher Stone as first mooted the possibility in a footnote two decades earlier: CD Stone, ‘Should Trees Have Standing? Towards Legal Rights for Natural Objects’ (1972) 45 *Southern California L Rev* 456 n26.

⁵ See S Chesterman, ‘Artificial Intelligence and the Problem of Autonomy’ (2020) 1 *Notre Dame J Emerging Technologies* 210; S Chesterman, ‘Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity’ (2021) *AJCL* forthcoming.

when things go right. A growing body of literature examines ownership of intellectual property created by AI systems, for example.

The tension in these discussions is whether personhood is granted for instrumental or inherent reasons. Arguments are typically framed in instrumental terms, with comparisons to the most common artificial legal person: the corporation. Yet implicit in many of those arguments, or explicit in their illustrations and examples, is the idea that as AI systems approach the point of indistinguishability from humans — that is, when they pass Turing’s Test — they should be entitled to a status comparable to natural persons.

Until recently, such arguments were all speculative. Then in 2017 Saudi Arabia granted ‘citizenship’ to the humanoid robot Sophia⁶ and an online system with the persona of a seven-year-old boy was granted ‘residency’ in Tokyo.⁷ These were gimmicks — Sophia, for example, is essentially a chatbot with a face.⁸ In the same year, however, the European Parliament adopted a resolution calling on its Commission to consider creating ‘a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently.’⁹

This article begins with the most immediate challenge, which is whether some form of juridical personality would fill a responsibility gap or be otherwise advantageous to the legal system. Based on the history of corporations and other artificial legal persons, it does not seem in doubt that most legal systems *could* grant AI systems a form of personality; the more interesting questions are whether they should and what content that personality might have.

The second section then turns to the analogy with natural persons. It might seem self-evident that a machine could never be a natural person. Yet for centuries slaves and women were not recognized as full persons either. If one takes the Turing Test to its logical, *Blade Runner*, conclusion, it is possible that AI systems truly indistinguishable from humans might one day claim the same status. Although arguments about ‘rights for robots’ are presently confined

⁶ O Cuthbert, ‘Saudi Arabia Becomes First Country to Grant Citizenship to a Robot’, *Arab News* (26 October 2017).

⁷ A Cuthbertson, ‘Artificial Intelligence “Boy” Shibuya Mirai Becomes World’s First AI Bot to Be Granted Residency’, *Newsweek*, 6 November 2017.

⁸ D Gershgorin, ‘Inside the Mechanical Brain of the World’s First Robot Citizen’, *Quartz* (12 November 2017).

⁹ European Parliament Resolution with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (European Parliament, 16 February 2017), para 59(f).

to the fringes of the discourse, this possibility is implicit in many of the arguments in favour of AI systems owning the intellectual property that they create.

Taken seriously, moreover, the idea that AI systems could equal humans suggests a third reason for contemplating personality. For once equality is achieved, there is no reason to assume that AI advances would stop there. Though general AI remains science fiction for the present, it invites consideration whether legal status could shape or constrain behaviour if or when humanity is surpassed. Should it ever come to that, of course, the question might not be whether we recognize the rights of a general AI, but whether it recognizes ours.

I. JURIDICAL PERSONALITY: A BODY TO KICK?

Legal personality is fundamental to any system of laws. The question of who can act, who can be the subject of rights and duties, is a precursor to almost every other issue. Yet close examination of these foundations reveals surprising uncertainty and disagreement. Despite this, as John Dewey observed in 1926, ‘courts and legislators do their work without such agreement, sometimes without any conception or theory at all’ regarding the nature of personality. Indeed, he went on, recourse to theory has ‘more than once operated to hinder rather than facilitate the adjudication of a special question of right or obligation.’¹⁰

In practice, the vast majority of legal systems recognize two forms of legal person: natural and juridical. Natural persons are recognized because of the simple fact of being human.¹¹ Juridical persons, by contrast, are non-human entities that are granted certain rights and duties by law. Corporations and other forms of business associations are the most common examples, but many other forms are possible. Religious, governmental, and intergovernmental entities may also act as legal persons at the national and international level.

It is telling that these are all aggregations of human actors, though there are examples of truly non-human entities being granted personhood. In addition to the examples mentioned in the

¹⁰ J Dewey, ‘The Historic Background of Corporate Legal Personality’ (1926) 35 Yale LJ 660.

¹¹ This presumes, of course, agreement on the meaning of ‘human’ and terms such as birth and death. See N Naffine, ‘Who Are Law’s Persons? From Cheshire Cats to Responsible Subjects’ (2003) 66 MLR 346.

introduction, these include temples in India,¹² a river in New Zealand,¹³ and the entire ecosystem of Ecuador.¹⁴ There seems little question that a state could attribute some kind of personality to new entities like AI systems;¹⁵ if that happens, recognition would likely be accorded by other states also.¹⁶

A. *Theories of Juridical Personality*

Scholars and law reform bodies have already suggested attributing AI systems with some form of legal personality to help address liability questions, such as an automated driving system entity in the case of driverless cars whose behaviour may not be under the control of their ‘driver’ or predictable by their manufacturer or owner.¹⁷ A few writers have gone further, arguing that procedures need to be put in place to try robot criminals, with provision for ‘punishment’ through reprogramming or, in extreme cases, destruction.¹⁸

These arguments suggest an instrumental approach to personality, but scholarly explanations of the most common form of juridical person — the corporation — show disparate justifications for its status as a separate legal person that may help answer the question of whether that status should be extended to AI systems also.

The aggregate theory, sometimes referred to as the contractarian or symbolist theory, holds that a corporation is a device created by law to allow natural persons who organize themselves as a group to reflect that organization in their legal relations with other parties. Group members could establish individual contractual relations with those other parties limiting liability and so on, but the corporate form enables them to do so collectively at a

¹² See, eg, *Shiromani Gurdwara Prabandhak Committee, Amritsar v Shri Somnath Dass* AIR 2000 SC 1421 (Supreme Court of India).

¹³ Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 (New Zealand), section 14(1). This followed designation of the Te Urewera National Park as ‘a legal entity, [with] all the rights, powers, duties, and liabilities of a legal person’. Te Urewera Act 2014 (New Zealand), section 11(1).

¹⁴ Constitution of the Republic of Ecuador 2008 (Ecuador), art 10.

¹⁵ For a discussion of the limits of what can be a legal person, see VAJ Kurki, *A Theory of Legal Personhood* (OUP 2019) 127-52.

¹⁶ See, eg, *Bumper Development Corp. v Commissioner of Police for the Metropolis* [1991] 1 WLR 1362 (recognizing legal personality of an Indian temple under English law).

¹⁷ See, eg, Changing Driving Laws to Support Automated Vehicles (Policy Paper) (National Transport Commission, May 2018), para 1.5; Automated Vehicles: A Joint Preliminary Consultation Paper (Law Commission, Consultation Paper No. 240; Scottish Law Commission, Discussion Paper No 166, 2018), para 4.107. Cf Chesterman, ‘Autonomy’ (n 5) 225.

¹⁸ See, eg, G Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer 2015); Y Hu, ‘Robot Criminals’ (2019) 52 U Michigan JL Reform 487.

lower cost.¹⁹ The theory has been criticized and is, in any case, the least applicable to AI systems.²⁰

The fiction and concession theories of corporate personality have separate origins²¹ but amount to the same thing: corporations have personality because a legal system chooses to give it to them. As the US Supreme Court observed in 1819, a corporation ‘is an artificial being, invisible, intangible, and existing only in contemplation of law.’²² Personality is granted to achieve policy ends, such as encouraging entrepreneurship, or to contribute to the coherence and stability of the legal system, such as through the perpetuity of certain entities. The purposive aspect used to be more evident when personality was explicitly granted through a charter or legislation; in the course of the twentieth century this became a mere formality.²³ These positivist accounts most closely align with legislative and judicial practice in recognizing personality, and could encompass its extension to AI systems.

The realist theory, by contrast, holds that corporations are neither fictions nor mere symbols but objectively real entities that pre-exist conferral of personality by a legal system. Though they may have members, they act independently and their actions may not be attributable to those members. At its most extreme, it is argued that corporations are not only legal but also moral persons.²⁴ This theory tends to be favoured more by theorists and sociologists than legislators and judges, but echoes the tension highlighted in the introduction: that legal personality is not merely bestowed but deserved. In practice, however, actual recognition as a person before the law remains in the gift of the state.²⁵

The end result is, perhaps, that Dewey was correct a century ago: “‘person” signifies what law makes it signify’.²⁶ Though the question of personality is a binary, however — recognized or not — the content of that status is a spectrum. Setting aside for the moment the idea that an

¹⁹ R Coase, ‘The Nature of the Firm’ (1937) 4 *Economica* 386. Cf V Morawetz, *A Treatise on the Law of Private Corporations* (Little, Brown 1886) 2 (‘the fact remains self-evident that a corporation is not in reality a person or a thing distinct from its constituent parts. The word corporation is but a collective name for the corporators’).

²⁰ N Banteka, ‘Artificially Intelligent Persons’ (2020) 58 *Houston L Rev* forthcoming.

²¹ Dewey (n 10) 665-69.

²² *Trustees of Dartmouth Coll. v Woodward*, 17 US 518, 636 (1819).

²³ See, eg, CE Amsler, RL Bartlett, and CJ Bolton, ‘Thoughts of Some British Economists on Early Limited Liability and Corporate Legislation’ (1981) 13 *History of Political Economy* 774; G Dari-Mattiacci *et al*, ‘The Emergence of the Corporate Form’ (2017) 33 *J of L, Economics and Organization* 193.

²⁴ P French, ‘The Corporation as a Moral Person’ (1979) 16(3) *American Philosophical Q* 207.

²⁵ K Iwai, ‘Persons, Things and Corporations: The Corporate Personality Controversy and Comparative Corporate Governance’ (1999) 47 *AJCL* 583; SM Watson, ‘The Corporate Legal Person’ (2019) 19 *J Corporate L Studies* 137.

²⁶ Dewey (n 10) 655.

AI system might deserve recognition as a person, a state's decision to grant it should be guided by the rights and duties that would be recognized also.

B. The Content of Legal Personality

Legal personality brings with it rights and obligations, but these need not be the same for all persons within a legal system. Even among natural persons, the struggle for equal rights of women, ethnic or religious minorities, and other disadvantaged groups reflects this truth.²⁷

It is possible, for example, to grant only rights without obligations. This has tended to be the approach in giving personhood to nature — both in theory, when it was first advocated in 1972,²⁸ and in practice, such as in the Constitution of Ecuador.²⁹ One could argue that such 'personality' is merely an artifice to avoid problems of standing: enabling human individuals to act on behalf of a non-human rights holder, rather than requiring them to establish standing in their own capacity.³⁰ In any case, it seems inapposite to the reasons for considering personality of AI systems.

On the other hand, AI legal personality could come only with obligations. That might seem superficially attractive, but insofar as those obligations are intended to address accountability gaps there would be some obvious problems. Civil liability typically leads to an award of damages, for example, which can only be paid if the wrongdoer is capable of owning property.³¹ One could imagine scenarios in which those payments were made from a central fund, though this would be more akin to compulsory insurance regimes proposed as an alternative means of addressing the liability question.³² 'Personality' would be a mere formality.

²⁷ JJ Bryson, ME Diamantis, and TD Grant, 'Of, for, and by the People: The Legal Lacuna of Synthetic Persons' (2017) 25 *Artificial Intelligence and Law* 280.

²⁸ Stone (n 4).

²⁹ Constitution of the Republic of Ecuador, arts 71-74.

³⁰ C Rodgers, 'A New Approach to Protecting Ecosystems' (2017) 19 *Environmental L Rev* 266. In New Zealand, by contrast, trustees were established to act on behalf of the environmental features given personality.

³¹ Cf *Liability for Artificial Intelligence and Other Emerging Digital Technologies* (EU Expert Group on Liability and New Technologies, 2019) 38.

³² See, eg, DA Crane, KD Logue, and BC Pilz, 'A Survey of Legal Issues Arising from the Deployment of Autonomous and Connected Vehicles' (2017) 23 *Michigan Telecommunications and Technology L Rev* 256-59. Cf KS Abraham and RL Rabin, 'Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era' (2019) 105 *Virginia L Rev* 127.

In the case of corporations, personality typically means the capacity to sue and be sued, to enter into contracts, to incur debt, to own property, and to be convicted of crimes. On the rights side, the extent to which corporations enjoy constitutional protections comparable to natural persons is the subject of ongoing debate. Though the United States has arguably granted the most protections to corporate entities, even there a line has been drawn at guarantees such as the right against self-incrimination.³³ Typically, juridical persons will have fewer rights than natural ones. (A similar situation obtains in international law, where states enjoy plenary personality and international organizations may have varying degrees of it.³⁴)

1. Private Law

The ability to be sued is one of the primary attractions of personality for AI systems, as the European Parliament acknowledged.³⁵ This presumes, of course, that there are meaningful accountability gaps that can and should be filled. These gaps are often overstated.³⁶ A different reason for wariness about such a remedy is that, even if it did serve a gap-filling function, granting personality to AI systems would also shift responsibility under current laws away from existing legal persons. Indeed, it would create an incentive to transfer risk to such electronic persons in order to shield natural and traditional juridical ones from exposure.³⁷ That is a problem with corporations also, which may be used to protect investors from liability beyond the fixed sum of their investment — indeed, that is often the point of using a corporate vehicle in the first place. The reallocation of risk is justified on the basis that it encourages investment and entrepreneurship.³⁸ Safeguards typically include a requirement that the name of a limited liability entity include that status in its name ('Ltd', 'LLC', etc) and the possibility of piercing the corporate veil in limited circumstances to prevent abuse of the

³³ SA Trainor, 'A Comparative Analysis of a Corporation's Right Against Self-Incrimination' (1994) 18 *Fordham International LJ* 2139. But see *Citizens United v Federal Election Commission*, 558 US 310 (US Supreme Court, 2010).

³⁴ S Chesterman, 'Does ASEAN Exist? The Association of Southeast Asian Nations as an International Legal Person' (2008) XII *Singapore Year Book of International Law* 199.

³⁵ See above n 9.

³⁶ See Chesterman, 'Autonomy' (n 5); Chesterman, 'Opacity' (n 5).

³⁷ Bryson, Diamantis, and Grant (n 27) 287.

³⁸ FH Easterbrook and DR Fischel, 'Limited Liability and the Corporation' (1985) 52 *U Chicago L Rev* 89.

form.³⁹ In the case of AI systems, similar veil piercing mechanisms could be developed — though if a human were manipulating AI in order to protect him- or herself from liability, the ability to do so might suggest that the AI system in question was not deserving of its separate personhood.⁴⁰

Entry into contracts is occasionally posited as a reason to grant AI systems personality.⁴¹ Yet the use of electronic agents to conclude binding agreements is hardly new. High-frequency trading, for example, relies on algorithms concluding agreements with other algorithms on behalf of traditional persons.⁴² Though the autonomy of AI systems may challenge application of existing doctrine to such practices — notably, when something goes wrong, such as a mistake — this is still resolvable without recourse to new legal persons.⁴³

Taking on debt and owning property would be necessary incidents of the ability to be sued and enter into contracts.⁴⁴ The possibility that AI systems could accumulate wealth raises the question of whether or how they might be taxed. Taxation of robots has been proposed as a means of addressing the diminished tax base and displacement of workers anticipated as a result of automation.⁴⁵ Bill Gates, among others, has suggested that such robots — or the companies that own them — should be taxed.⁴⁶ Industry representatives have argued that this would have a negative impact on competitiveness and thus far it has not been adopted.⁴⁷ An alternative is to look not at the machines but at the position of companies abusing market position, with possibilities including more aggressive taxing of profits or requirements for

³⁹ D Millon, 'Piercing the Corporate Veil, Financial Responsibility, and the Limits of Limited Liability' (2007) 56 Emory LJ 1305.

⁴⁰ J Turner, *Robot Rules: Regulating Artificial Intelligence* (Palgrave Macmillan 2019) 193.

⁴¹ See, eg, S Chopra and LF White, *A Legal Theory for Autonomous Artificial Agents* (U Michigan Press 2011) 160.

⁴² See, eg, T Čuk and A van Waeyenberge, 'European Legal Framework for Algorithmic and High Frequency Trading (Mifid 2 and MAR) A Global Approach to Managing the Risks of the Modern Trading Paradigm' (2018) 9 European J Risk Regulation 146.

⁴³ See, eg, *Quoine Pte Ltd v B2C2 Ltd* [2020] SGCA(I) 2 (Singapore Court of Appeal), paras 97-128; Chesterman, 'Autonomy' (n 5) 243-44.

⁴⁴ Some argue that this is the most important function of separate legal personality for corporations: H Hansmann and R Kraakman, 'The Essential Role of Organizational Law' (2000) 110 Yale LJ 387. Cf H Tjio, 'Lifting the Veil on Piercing the Veil' [2014] Lloyd's Maritime and Commercial Law Quarterly 19. It is conceivable that AI systems lacking the ability to own property could still be subject to certain forms of legal process, such as injunctions, and could offset debts through their 'labour'.

⁴⁵ BA King, T Hammond, and J Harrington, 'Disruptive Technology: Economic Consequences of Artificial Intelligence and the Robotics Revolution' (2017) 12(2) J Strategic Innovation and Sustainability 53.

⁴⁶ KJ Delaney, 'The Robot that Takes Your Job Should Pay Taxes, says Bill Gates', *Quartz*, 18 February 2017.

⁴⁷ G Proadhan, 'European Parliament Calls for Robot Law, Rejects Robot Tax', *Reuters* (16 February 2017); L Summers, 'Robots Are Wealth Creators and Taxing Them Is Illogical', *Financial Times* (6 March 2017).

distributed share ownership.⁴⁸ In any case, taxation of AI systems — like the ability to take on debt and own property — would follow rather than justify granting them personality.⁴⁹ (The question of AI systems owning their creations will be considered in the next section.⁵⁰)

In addition to owning property, AI systems might also be called on to manage it. In 2014, for example, it was announced that a Hong Kong venture capital firm had appointed a computer program called Vital to its board of directors.⁵¹ As with the Saudi Arabian government's awarding of citizenship, this was more style than substance — as a matter of Hong Kong law the program was not appointed to anything; in an interview some years later, the managing partner conceded that the company merely treated Vital as a member of the board with observer status.⁵² It is possible that human directors might delegate some responsibility to an AI system, but under most corporate law regimes they cannot absolve themselves of the ultimate responsibility for managing the organization.⁵³ Most jurisdictions require that those directors be natural persons, though in some it is possible for a juridical person — typically another corporation — to serve on the board.⁵⁴ Shawn Bayern has gone further, arguing that loopholes in US business entity law could be used to create limited liability companies, with no human members at all.⁵⁵ This requires a somewhat tortured interpretation of that law — a natural person creates a company, adds an AI system as a member, then resigns⁵⁶ — but suggests the manner in which legal personality might be adapted in the future.

⁴⁸ 'Why Taxing Robots Is not a Good Idea', *Economist* (25 February 2017).

⁴⁹ Cf L Floridi, 'Robots, Jobs, Taxes, and Responsibilities' (2017) 30 *Philosophy & Technology* 1.

⁵⁰ See below section II.B.

⁵¹ R Wile, 'A Venture Capital Firm Just Named an Algorithm to Its Board of Directors', *Business Insider*, 13 May 2014.

⁵² N Burrige, 'AI Takes Its Place in the Boardroom', *Nikkei Asian Review*, 25 May 2017.

⁵³ F Möslin, 'Robots in the Boardroom: Artificial Intelligence and Corporate Law' in W Barfield and U Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar 2018) 658-60.

⁵⁴ See, eg, Personen- und Gesellschaftsrecht (PGR) 1926 (Liechtenstein), art 344; Companies Ordinance 2014 (HK), section 457. This was also possible under English law until 2015. See now Small Business, Enterprise and Employment Act 2015 (UK), section 87.

⁵⁵ S Bayern, 'Of Bitcoins, Independently Wealthy Software, and the Zero-Member LLC' (2014) 108 *Northwestern U L Rev* 1495-500.

⁵⁶ S Bayern, 'The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems' (2015) 19 *Stanford Technology L Rev* 101.

2. Criminal Law

A final quality of legal personality is the most visceral and worthy of some elaboration: the ability to be punished. If given legal personality comparable to a corporation, there seems little reason to argue over whether an AI system could be prosecuted under the criminal law. Provided *actus reus* and *mens rea* are established,⁵⁷ such an entity could be fined or have its property seized; a licence to operate could be suspended or revoked. In some jurisdictions, a winding up order can be made against a juridical person; where that is not available, a fine sufficiently large to bankrupt the entity may have the same effect. In an extreme case, one could imagine a ‘robot criminal’ being destroyed. But would this be desirable and would it be effective?

The most commonly articulated reasons for criminal punishment are retribution, incapacitation, deterrence, and rehabilitation.⁵⁸ Retribution is the oldest reason for punishment, sublimating the victim’s desire for revenge into a societal demonstration that wrongs have consequences.⁵⁹ Calibration of those consequences was at its most literal in the *lex talionis*: an eye for an eye, a tooth for a tooth. The demonstrative effect of fining a corporation — or an electronic ‘person’ — may be preferable to a crime otherwise going unpunished.⁶⁰

The penal system can also be used to incapacitate those convicted of crimes, physically preventing them from reoffending. Typically this is through varying forms of incarceration, but may also include exile, amputation of limbs, castration, and execution. In the case of corporations, it may include withdrawal of a licence to operate or a compulsory winding up order.⁶¹ Here direct analogies with the treatment of dangerous animals and machinery can

⁵⁷ For corporations this has sometimes proved a difficult, but not insurmountable challenge. See VS Khanna, ‘Corporate Criminal Liability: What Purpose Does It Serve?’ (1996) 109 Harvard L Rev 1513; JW Yockey, ‘Beyond Yates: From Engagement to Accountability in Corporate Crime’ (2016) 12 New York U J of L and Business 412-413; SW Buell, ‘Criminally Bad Management’ in J Arlen (ed), *Research Handbook on Corporate Crime and Financial Misdealing* (Edward Elgar 2018) 59. The fact that corporations are capable of violating criminal laws despite lacking free will or moral responsibility presumably dispenses with this as an argument against criminal responsibility of AI systems.

⁵⁸ It is arguable that the symbolic role of criminal law need not require actual punishment — it is not uncommon to have laws that are unenforced in practice. Yet this typically relies on an explicit or implicit decision not to investigate or prosecute specific crimes, rather than acceptance that a class of actors cannot be punished at all.

⁵⁹ Denunciation is sometimes presented as a standalone justification for punishment in its own right. See B Wringle, *An Expressive Theory of Punishment* (Palgrave Macmillan 2016).

⁶⁰ C Mulligan, ‘Revenge Against Robots’ (2018) 69 South Carolina L Rev 579; Hu (n 18) 503-07.

⁶¹ WR Thomas, ‘Incapacitating Criminal Corporations’ (2019) 72 Vanderbilt L Rev 905.

be made, although measures such as putting down a vicious dog or decommissioning a faulty vehicle are administrative rather than penal and do not depend on determinations of 'guilt'.⁶² In some jurisdictions, children and the mentally ill may be deemed incapable of committing crimes, yet they may still be detained by the state if judged to be a danger to themselves or the community.⁶³ Such individuals do not lose their personality; in the case of AI systems, it is not necessary to give them personality in order to impose measures akin to confinement if a product can be recalled or a licence revoked.

Deterrence is a more recent justification for punishment, premised on the rationality of offenders. By structuring penalties, it imposes costs on behaviour that are intended to outweigh any potential benefits. The ability to reduce criminality to economic analysis may seem particularly applicable to both corporations and AI systems. Yet in the case of the former the incentives are really aimed at human managers who might otherwise act in concert through the corporation for personal as well as corporate gain.⁶⁴ In the case of an AI system, the deterrent effect of a fine would shape behaviour only if its programming sought to maximize economic gain without regard for the underlying criminal law itself.

A final rationale for punishment is rehabilitation. Like incapacitation and deterrence, it is forward-looking and aims to reduce recidivism. Unlike incapacitation, however, it seeks to influence the decision to offend rather than the ability to do so;⁶⁵ unlike deterrence, that influence is intended to operate intrinsically rather than extrinsically.⁶⁶ Rehabilitation in respect of natural persons often appears to be embraced more in theory than in practice; in the United States in particular, it fell from favour in the 1970s.⁶⁷ With respect to corporations, however, the clearer levers of influence have encouraged experimentation with narrowly tailored penalties that aim to encourage good behaviour as well as discourage bad.⁶⁸ Such an approach might seem well suited to AI systems, with violations of the criminal law being

⁶² See, eg, D Legge and S Brooman, *Law Relating to Animals* (Cavendish Publishing 2000).

⁶³ A Loughnan, *Manifest Madness: Mental Incapacity in the Criminal Law* (OUP 2012).

⁶⁴ A Hamdani and A Klement, 'Corporate Crime and Deterrence' (2008) 61 *Stanford L Rev* 271. Cf RA Guttman, 'Effective Compliance Means Imposing Individual Liability' (2018) 5 *Emory Corporate Governance and Accountability Rev* 77.

⁶⁵ J Bentham, 'Panopticon Versus New South Wales' in J Bowring (ed), *The Works of Jeremy Bentham* (William Tait 1843) vol 4, 174.

⁶⁶ See T Ward and S Maruna, *Rehabilitation* (Routledge 2007).

⁶⁷ See, eg, AW Alschuler, 'The Changing Purposes of Criminal Punishment: A Retrospective on the Past Century and Some Thoughts About the Next' (2003) 70 *U Chicago L Rev* 9. Cf FT Cullen and KE Gilbert, *Reaffirming Rehabilitation* (2nd edn, Anderson 2013).

⁶⁸ ME Diamantis, 'Clockwork Corporations: A Character Theory of Corporate Punishment' (2018) 103 *Iowa L Rev* 507.

errors to be debugged rather than sins to be punished.⁶⁹ Indeed, the educative aspect of rehabilitation has been directly analogized to machine learning in a book-length treatise on the topic.⁷⁰ Yet neither legal personality nor the coercive powers of the state should be necessary to ensure that machine learning leads to outputs that do not violate the criminal law.

C. *No Soul to Be Damned*

While arguments justifying liability of corporations tend to be instrumental, it is striking how the emerging literature on ‘robot criminals’ slides into anthropomorphism. The very term suggests a special desire to hold humanoid AI systems to a higher standard than, say, household appliances with varying degrees of autonomy or unembodied AI systems operating in the cloud.⁷¹ There is no principled reason for such a distinction, but it speaks to the tension within arguments for AI personality that blend instrumental and inherent justifications.

Interestingly, arguments over the juridical personality of corporations tend to focus on the opposite problem: their *dissimilarity* to humans, pithily described by the First Baron Thurlow as them having ‘no soul to be damned, and no body to be kicked’.⁷² The lack of a soul has not impeded juridical personality of corporations and poses no principled barrier to treating AI systems similarly. Corporate personality is different from AI personality, however, in that a corporation is made up of human beings, through whom it operates, whereas an AI system is made *by* humans.⁷³

Instrumental reasons could, therefore, justify according legal personality to AI systems. But they do not require it. The implicit anthropomorphism elides further challenges such as defining the threshold of personality when AI systems exist on a spectrum, as well as how personality might apply to distributed systems. It would be possible, then, to create legal persons comparable to corporations — each autonomous vehicle, smart medical device, resume-screening algorithm, and so on could be incorporated.⁷⁴ If there are true liability gaps

⁶⁹ MA Lemley and B Casey, ‘Remedies for Robots’ (2019) 86 U Chicago L Rev 1370.

⁷⁰ Hallevey (n 18) 210-11.

⁷¹ Cf JM Balkin, ‘The Three Laws of Robotics in the Age of Big Data’ (2017) 78 Ohio State LJ 1219.

⁷² MA King, *Public Policy and the Corporation* (Chapman and Hall 1977) 1. See, eg, JC Coffee, Jr, ‘“No Soul to Damn: No Body to Kick”: An Unscandalized Inquiry into the Problem of Corporate Punishment’ (1981) 79 Michigan L Rev 386.

⁷³ SM Solaiman, ‘Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy’ (2017) 25 Artificial Intelligence and Law 174.

⁷⁴ See, eg, Chopra and White (n 41) 161.

then it is possible that such legal forms could fill them. Yet the more likely beneficiaries of such an arrangement would be producers and users, who would thus be insulated from some or all liability.

II. NATURAL PERSONALITY: COGITO, ERGO SUM?

Instrumentalism is not the only reason legal systems recognize personality, however. In the case of natural persons, no Turing Test needs to be passed: the mere fact of being born entitles one to personhood before the law.⁷⁵

It was not always thus. Through much of human history, slaves were bought and sold like property;⁷⁶ indigenous peoples were compared to animals roaming the land, justifying their dispossession;⁷⁷ and for centuries under English law, Blackstone's summary of the position of women held that 'husband and wife are one person, and the husband is that person.'⁷⁸ Even today, natural persons enjoy plenary rights and obligations only if they are adults, of sound mind, and not incarcerated.

As indicated earlier, many of the arguments in favour of AI personality implicitly or explicitly assume that AI systems are approaching human qualities in a manner that would entitle them to comparable recognition before the law. Such arguments have been challenged both for their analysis and their implications. In terms of analysis, Neil Richards and William Smart have termed the tendency to anthropomorphize AI systems the 'android fallacy'.⁷⁹ Experiment after experiment has shown that people are more likely to ascribe human qualities such as moral sensibility to machines on the basis of their humanoid appearance, natural language communication, or the mere fact of having been given a name.⁸⁰ More

⁷⁵ See, eg, Universal Declaration of Human Rights, GA Res 217A(III) (1948), UN Doc A/810 (1948), art 1; International Covenant on Civil and Political Rights (ICCPR), 16 December 1966, 999 UNTS 171, in force 23 March 1976, art 6(1).

⁷⁶ See J Allain (ed), *The Legal Understanding of Slavery: From the Historical to the Contemporary* (OUP 2013).

⁷⁷ S Chesterman, "'Skeletal Legal Principles": The Concept of Law in Australian Land Rights Jurisprudence' (1998) 40 *J Legal Pluralism and Unofficial L* 61.

⁷⁸ L Holcombe, *Wives and Property: Reform of the Married Women's Property Law in Nineteenth-Century England* (Martin Robertson 1983) 18.

⁷⁹ NM Richards and WD Smart, 'How Should the Law Think About Robots?' in R Calo, AM Froomkin, and I Kerr (eds), *Robot Law* (Edward Elgar 2016) 18-21.

⁸⁰ Cf L Damiano and P Dumouchel, 'Anthropomorphism in Human-Robot Co-evolution' (2018) 9 *Frontiers in Psychology* 468.

serious arguments about AI approximating human qualities are challenged due to unexamined assumptions about how those qualities manifest in humans ourselves.⁸¹

In terms of the implications, the 2017 European Parliament resolution prompted hundreds of AI experts from across the continent to warn in an open letter that legal personality for AI would be inappropriate from ‘an ethical and a legal perspective’. Interestingly, such warnings may themselves fall foul of the android fallacy by assuming that legal status based on the natural person model necessarily brings with it all the ‘human’ rights guaranteed under EU law.⁸² Other writers candidly admit that the only basis for denying AI systems personality may ultimately be a form of speciesism — privileging human welfare over robot welfare because we the lawmakers are human.⁸³ If AI systems become so sophisticated that this is our strongest defence, the problem may not be their legal status but our own.⁸⁴

This section nonetheless takes seriously the idea that certain AI systems might have an entitlement to personality due to their inherent qualities. The technical aspects of how those qualities might manifest — and indeed a detailed examination of the human qualities that they mimic — are beyond the scope of this article.⁸⁵ Instead, the focus will be on how and why natural personhood might be extended. A first question to examine is how this has been handled in the past, such as the enfranchisement and empowerment of natural persons long treated as inferior to white men. More recently, activists and scholars have urged further expansion of certain rights to non-human animals such as chimpanzees based on their own inherent qualities. The inquiry then turns to the strongest articulation today of meaningful rights on behalf of AI systems for inherent rather than instrumental reasons: that they should be able to own their creations.

⁸¹ See, eg, E Hildt, ‘Artificial Intelligence: Does Consciousness Matter?’ (2019) 10(1535) *Frontiers in Psychology*, 1-3; G Meissner, ‘Artificial Intelligence: Consciousness and Conscience’ (2020) 35 *AI & Society* 231. For John Searle’s famous ‘Chinese room’ argument, see JR Searle, ‘Minds, Brains, and Programs’ (1980) 3 *Behavioral and Brain Sciences* 417-24.

⁸² Open Letter to the European Commission: Artificial Intelligence and Robotics (April 2018), para 2(b). Cf Turner (n 40) 189-90.

⁸³ See, eg, Bryson, Diamantis, and Grant (n 27) 283. Cf P Singer, ‘Speciesism and Moral Status’ (2009) 40 *Metaphilosophy* 567.

⁸⁴ See below section III.

⁸⁵ See, eg, J-M Fellous and MA Arbib, *Who Needs Emotions? The Brain Meets the Robot* (OUP 2005); W Wallach and C Allen, *Moral Machines: Teaching Robots Right from Wrong* (OUP 2009).

A. *The Extension of Natural Personality*

The arc of the moral universe is long, as Dr Martin Luther King Jr famously intoned, but it bends towards justice. At the time that the United States drafted its Declaration of Independence in 1776, the notion that ‘all men’ [*sic*] were ‘created equal’ was demonstrably untrue. A decade later, the French Declaration on the Rights of Man similarly proclaimed natural and imprescriptible rights for all — ‘nonsense upon stilts’ was Jeremy Bentham’s observation.⁸⁶ Man may well be born free, as Rousseau had opined in the opening lines of *The Social Contract*, but everywhere he remained in chains.⁸⁷

And yet the succeeding centuries did see a progressive realization of those lofty aspirations and the spread of rights. By the middle of the twentieth century, the Universal Declaration of Human Rights could claim that all human beings were ‘born free and equal in dignity and rights’, despite one-third of them living in territories that the UN itself classified as non-self-governing. Decolonization, the end of apartheid, women’s liberation and other movements followed; rights remain a site of contestation, but virtually no state today would seriously contend that human adults are not persons before the law.⁸⁸

Interestingly, some arguments in favour of legal personality for AI draw not on this progressivist narrative of natural personhood but on the darker history of slavery. Andrew Katz and Ugo Pagallo, for example, find analogies with the ancient Roman law mechanism of *peculium*, whereby a slave lacked legal personality and yet could operate as more than a mere agent for his master.⁸⁹ As an example of a creative interpretation of personhood it is interesting, though it relies on instrumental justifications rather than the inherent qualities of slaves. As Pagallo notes, *peculium* was in effect a sort of ‘proto-limited liability company’.⁹⁰ From the previous section, there is no bar on legal systems creating such structures today —

⁸⁶ J Bentham, ‘Anarchical Fallacies’ in J Bowring (ed), *The Works of Jeremy Bentham* (William Tait 1843) vol 2, 501.

⁸⁷ J-J Rousseau, *The Social Contract* (GDH Cole tr, first published 1762, JM Dent 1923) 49.

⁸⁸ For limited exceptions concerning apostates and persons with disabilities, see PM Taylor, *A Commentary on the International Covenant on Civil and Political Rights* (CUP 2020) 449-54. For a discussion of anencephalic infants, see Kurki (n 15) 9.

⁸⁹ A Katz, ‘Intelligent Agents and Internet Commerce in Ancient Rome’ (2008) 20 *Society for Computers and Law* 35; U Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer 2013) 103-06. See also H Ashrafian, ‘Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights’ (2015) 21 *Science and Engineering Ethics* 325; S Nasarre-Aznar, ‘Ownership at Stake (Once Again): Housing, Digital Contents, Animals, and Robots’ (2018) 10 *J Property, Planning, and Environmental L* 78; E Fosch-Villaronga, *Robots, Healthcare, and the Law: Regulating Automation in Personal Care* (Routledge 2019) 152. In 2017, a digital bank called Peculium was established in France — presumably for investors who never studied Latin.

⁹⁰ Pagallo (n 89) 104.

as for whether they *should* do so, reliance upon long discarded laws associated with slavery may not be the strongest case possible.⁹¹

An alternative approach is to consider how the legal system treats animals.⁹² For the most part, they are regarded as property that can be bought and sold, but also as deserving of ‘humane’ treatment.⁹³ Liability of owners for damage caused by animals has limited application to AI systems;⁹⁴ here, the question is whether those animals might ‘own’ themselves.

Various efforts have sought to attribute degrees of personality to nonhuman animals, with little success. In 2013, for example, the Nonhuman Rights Project filed lawsuits on behalf of four captive chimpanzees, arguing that the animals exhibited advanced cognitive abilities, autonomy, and self-awareness. In denying writs of habeas corpus, the New York State Court Appellate Division did not dispute these qualities, but held that extension of rights such as personality had traditionally been linked to the imposition of obligations in the form of a social contract. Since, ‘needless to say’, the chimpanzees could not bear any legal duties, they could not enjoy rights of personality such as the right to liberty.⁹⁵ This was a curious basis on which to dismiss the case, as many humans who lack the capacity to exercise rights or responsibilities — infants, persons in a coma — are nonetheless deemed persons before the law.⁹⁶ A parallel case rejected that argument on the circular basis that it ‘ignores the fact that these are still human beings, members of the human community.’⁹⁷ Leave to appeal to the State Court of Appeals was denied, but one of the judges issued a concurring opinion that ended on a speculative note about the future of such litigation. The issue, Judge Fahey

⁹¹ M Chinen, *Law and Autonomous Machines: The Co-Evolution of Legal Responsibility and Technology* (Edward Elgar 2019) 19.

⁹² See, eg, VAJ Kurki and T Pietrzykowski (eds), *Legal Personhood: Animals, Artificial Intelligence and the Unborn* (Springer 2017); S Stucki, ‘Towards a Theory of Legal Animal Rights: Simple and Fundamental Rights’ (2020) OJLS forthcoming.

⁹³ Cf K Sykes, ‘Human Drama, Animal Trials: What the Medieval Animal Trials Can Teach Us About Justice for Animals’ (2011) 17 *Animal Law* 273.

⁹⁴ In the case of an animal known to belong to a ‘dangerous species’, its keeper is presumed to know that it has a tendency to cause harm and will be held liable for damage that it causes without the need to prove fault on the keeper’s part. For other animals, it must be shown that the keeper knew the specific animal was dangerous. These common law rules on animals are now covered by legislation: Animals Act 1971 (UK). The English courts have shied away from a general doctrine of strict liability for ‘ultra-hazardous activities’. See C Witting, *Street on Torts* (15th edn, OUP 2018) 453.

⁹⁵ *People ex rel Nonhuman Rights Project, Inc v Lavery*, 998 NYS 2d 248 (App Div, 2014).

⁹⁶ RS Abate, *Climate Change and the Voiceless: Protecting Future Generations, Wildlife, and Natural Resources* (CUP 2019) 101-02; Kurki (n 15) 6-10 (distinguishing between passive and active legal personhood).

⁹⁷ *Nonhuman Rights Project, Inc ex rel Tommy v Lavery*, 54 NYS 3d 392, 396 (App Div, 2017).

observed, is profound and far-reaching. ‘Ultimately, we will not be able to ignore it. While it may be arguable that a chimpanzee is not a “person,” there is no doubt that it is not merely a thing.’⁹⁸

Gabriel Hallevy has argued that animals are closer than AI systems to humans when one considers emotionality as opposed to rationality, but that this has not generally led to them being given personhood under the law. Instead, it is AI systems’ *rationality* that provides the basis for personhood.⁹⁹ That may be true with regard to the ability to make out the mental element of a criminal offence, but the fact that it is a crime to torture a chimpanzee but not a computer also points to an important difference in how the legal system values the two types of entity. In fact, a stronger argument may be made to protect embodied AI systems that evoke emotional responses on the part of humans — regardless of the sophistication of their internal processing. It seems probable that laws to protect such ‘social robots’ will at some point be adopted, comparable to animal abuse laws. As in the case of those laws, protection will likely be guided by social mores rather than consistent biological — or technological — standards.¹⁰⁰

The assumption that natural legal personality is limited to human beings is so ingrained in most legal systems that it is not even articulated.¹⁰¹ The failure to extend comparable rights even to our nearest evolutionary cousins bodes ill for advocates of AI personality based on presumed inherent qualities.¹⁰²

B. Rewarding Creativity

A distinct reason for considering whether AI systems should be recognized as persons focuses not on what they are but what they can do. For the most part, this is framed as the question of whether an individual or corporation can claim ownership of work done by an AI system. Implicit or explicit in such discussions, however, is the understanding that if such work had been done by a human then he or she would own it him- or herself.

⁹⁸ *Nonhuman Rights Project, Inc ex rel Tommy v Lavery*, 100 NE 3d 846, 848 (NY, 2018).

⁹⁹ Hallevy (n18) 28.

¹⁰⁰ K Darling, ‘Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects’ in R Calo, AM Froomkin, and I Kerr (eds), *Robot Law* (Edward Elgar 2016) 226-29. Cf R Calo, ‘Robotics and the Lessons of Cyberlaw’ (2015) 103 California L Rev 532.

¹⁰¹ Kurki (n 15) 8.

¹⁰² Cf *ibid* 176-78 (discussing whether AI could be ‘ultimately valuable’ and thus entitled to personhood).

There is, in fact, a long history of questioning whether machine-assisted creation is protectable through copyright.¹⁰³ Early photographs, for example, were not protected because the mere capturing of light through the lens of a *camera obscura* was not regarded as true authorship.¹⁰⁴ It took an iconic picture of Oscar Wilde going all the way to the US Supreme Court before copyright was recognized in mechanically produced creations.¹⁰⁵ The challenge today is distinct: not whether a photographer can ‘own’ the image passively captured by a machine, but who might own new works actively created by one. A computer program like a word processor does not own the text typed on it, any more than a pen owns the words that it writes. But AI systems now write news reports, compose songs, paint pictures — these activities generate value, but can and should they attract the protections of copyright law?

In most jurisdictions, the answer is no.

The US Copyright Office, for example, has stated that legislative protection of ‘original works of authorship’¹⁰⁶ is limited to works ‘created by a human being’. It will not register works ‘produced by a machine or mere mechanical process that operates randomly or automatically without *any* creative input or intervention from a human author.’¹⁰⁷ The word ‘any’ is key and begs the question of what level of human involvement is required to assert authorship.¹⁰⁸

Consider the world’s most famous selfie — of a black crested macaque. David Slater went to Indonesia to photograph the endangered monkeys, which were too nervous to let him take close-ups. So he set up a camera that enabled them to snap their own photos.¹⁰⁹ After the images gained significant publicity, animal rights activists argued that the monkeys had a greater claim to authorship of the photographs than the owner of the camera. Slater did

¹⁰³ J Grimmelmann, ‘There’s No Such Thing as a Computer-Authored Work — and It’s a Good Thing, Too’ (2016) 39 *Columbia J L & the Arts* 403.

¹⁰⁴ M de Cock Buning, ‘Artificial Intelligence and the Creative Industry: New Challenges for the EU Paradigm for Art and Technology by Autonomous Creation’ in W Barfield and U Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar 2018) 524.

¹⁰⁵ *Burrow-Giles Lithographic Co v Sarony*, 111 US 53 (1884). Arguments continued, however, with Germany withholding full copyright of photographs until 1965. A Nordemann, ‘Germany’ in Y Gendreau, A Nordemann, and R Oesch (eds), *Copyright and Photographs. An International Survey* (Kluwer 1999) 135.

¹⁰⁶ 17 USC §102(a).

¹⁰⁷ *Compendium of US Copyright Office Practices* (3rd edn, US Copyright Office 2019), §313.2 (emphasis added).

¹⁰⁸ See DJ Gervais, ‘The Machine as Author’ (2020) 105 *Iowa L Rev* forthcoming (proposing a test of ‘originality causation’).

¹⁰⁹ C Cheesman, ‘Photographer Goes Ape over Monkey Selfie: Who Owns the Copyright?’, *Amateur Photographer*, 7 August 2014.

eventually win, reflecting existing law¹¹⁰ — though as part of a settlement he agreed to donate 25 percent of future royalties from the images to groups protecting crested macaques.¹¹¹ As computers generate more content independently of their human programmers, it is going to be harder and harder for humans to take credit. Instead of training a monkey how to press a button, it may be more like a teacher trying to take credit for the work of his or her student.

Turning to the normative question of whether AI systems themselves *should* have a claim to ownership, the policy behind copyright is often articulated as incentivizing innovation. This has long been seen as unnecessary or inappropriate for computers. ‘All it takes,’ Pamela Samuelson wrote in 1986, ‘is electricity (or some other motive force) to get the machines into production.’¹¹² Here the Turing Test offers a different kind of thought experiment: the more machines are designed to copy human traits, the more important such incentives might become.

Until recently, China followed the orthodoxy that AI-produced work is not entitled to copyright protection.¹¹³ In December 2019, however, a district court in China held that an article produced by an algorithm could not be copied without permission. The article in question was a financial report published by Tencent with a note that it was ‘automatically written’ by Dreamwriter, a news writing program developed by the company in 2015. Shanghai Yingxun Technology Company copied the article without permission and Tencent sued. The article had been taken down, but the infringing company was ordered to pay ¥1,500 (US\$216) for ‘economic losses and rights protection’.¹¹⁴

The Chinese case reflects a distinct reason for recognizing copyright, which is the protection of upfront investment in creative processes. This account presumes that, in the absence of such protection, investment will dry up and there will be a reduced supply of creative

¹¹⁰ *Naruto v Slater*, 888 F 3d 418 (9th Cir, 2018). The Court held that Naruto lacked standing to sue under the US Copyright Act and had no claim to the photographs Slater had published.

¹¹¹ M Haag, ‘Who Owns a Monkey Selfie? Settlement Should Leave Him Smiling’, *New York Times* (11 September 2017).

¹¹² P Samuelson, ‘Allocating Ownership Rights in Computer-Generated Works’ (1986) 47 U Pittsburgh L Rev 1199.

¹¹³ *Beijing Feilin Law Firm v Baidu Corporation (No. 239)* (25 April 2019) (Beijing Internet Court); CHEN Ming, ‘Beijing Internet Court Denies Copyright to Works Created Solely by Artificial Intelligence’ (2019) 14 J Intellectual Property L & Practice 593.

¹¹⁴ 深圳市腾讯计算机系统有限公司 [Shenzhen Tencent Computer System Co Ltd] v 上海盈讯科技有限公司 [Shanghai Yingxun Technology Co Ltd] (24 December 2019) (Shenzhen Nanshan District People’s Court); ZHANG Yangfei, ‘Court Rules AI-Written Article Has Copyright’, *China Daily* (9 January 2020).

works.¹¹⁵ Such an approach to copyright is broadly consistent with common law doctrines concerning work created in the course of employment, known in the United States as work for hire, under which a corporate employer or an individual who commissions a work owns copyright despite the actual ‘author’ being someone else.¹¹⁶ This may not be available in civil law jurisdictions that place a greater emphasis on the moral rights of a human author.¹¹⁷

In Britain, legislation adopted in 1988 does in fact provide copyright protection for ‘computer-generated’ work, the ‘author’ of which is deemed to be the person who undertook ‘the arrangements necessary for the creation of the work’.¹¹⁸ Similar legislation has been adopted in New Zealand,¹¹⁹ India,¹²⁰ Hong Kong,¹²¹ and Ireland.¹²² Though disputes about who took the ‘arrangements necessary’ may arise, ownership by a recognized legal person or by no one at all remain the only possible outcomes.¹²³

The European Parliament in April 2020 issued a draft report arguing that AI-generated works could be regarded as ‘equivalent’ to intellectual works and therefore protected by copyright. It opposed giving personality of any kind to the AI itself, however, proposing that ownership instead vest in ‘the person who prepares and publishes a work lawfully, provided that the technology designer has not expressly reserved the right to use the work in that way’.¹²⁴ The ‘equivalence’ to intellectual work is interesting, justified here on the basis of a proposed shift in recognizing works based on a ‘creative result’ rather than a creative process.¹²⁵

¹¹⁵ K Raustiala and CJ Sprigman, ‘The Second Digital Disruption: Streaming and the Dawn of Datadriven Creativity’ (2019) 94 *New York U L Rev* 1603-04.

¹¹⁶ Cf *Asia Pacific Publishing Pte Ltd v Pioneers & Leaders (Publishers) Pte Ltd* [2011] 4 SLR 381, 398-402 (Singapore Court of Appeal) (distinguishing between authorship and ownership).

¹¹⁷ D Lim, ‘AI & IP: Innovation & Creativity in an Age of Accelerated Change’ (2018) 52 *Akron L Rev* 843-46.

¹¹⁸ Copyright, Designs and Patents Act 1988 (UK), section 9(3). ‘Computer-generated’ is defined in section 178 as meaning that the work was ‘generated by computer in circumstances such that there is no human author of the work’.

¹¹⁹ Copyright Act 1994 (NZ), section 5(2)(a).

¹²⁰ Copyright Amendment Act 1994 (India), section 2.

¹²¹ Copyright Ordinance 1997 (HK), section 11(3).

¹²² Copyright and Related Rights Act 2000 (Ireland), section 21(f).

¹²³ See, eg, *Nova Productions v Mazooma Games* [2007] EWCA Civ 219; A Brown *et al*, *Contemporary Intellectual Property: Law and Policy* (5th edn, OUP 2019) 100-01.

¹²⁴ S Séjourné, Draft Report on Intellectual Property Rights for the Development of Artificial Intelligence Technologies (European Parliament, Committee on Legal Affairs, 2020/2015(INI), 24 April 2020), paras 9-10.

¹²⁵ *Ibid*, Explanatory Statement.

For the time being, then, copyright cannot be owned by AI systems — and does not need to be in order to recognize the creativity of those systems. Nevertheless, reservations as to ownership being claimed by anyone else are evident in the limited rights given for ‘computer-generated’ works. The duration is generally for a shorter period, and the deemed ‘author’ is unable to assert moral rights — such as the right to be identified as the author of the work.¹²⁶ A World Intellectual Property Organization (WIPO) issues paper recognized the dilemma, noting that excluding such works would favour ‘the dignity of human creativity over machine creativity’ at the expense of making the largest number of creative works available to consumers. A middle path, it observed, might be to offer ‘a reduced term of protection and other limitations’.¹²⁷

C. *Protecting Inventors*

Whereas in copyright law the debate is over who owns works produced by AI systems, in patent law the question is whether they can be owned at all. Patent law in most jurisdictions provides or assumes that an ‘inventor’ must be human. In July 2019, Stephen Thaler decided to test those assumptions, filing patents in Britain, the European Union, and the United States that listed an AI system, DABUS, as the ‘inventor’.¹²⁸ The British Intellectual Property Office was willing to accept that DABUS created the inventions, but relevant legislation required that an inventor be a natural person and not a machine.¹²⁹ The European Patent Office (EPO) followed a more circuitous route to the same end, rejecting the applications on the basis that designating a machine as the inventor did not meet the ‘formal requirements’. These included stating the ‘family name, given names and full address of the inventor’.¹³⁰ A name, the EPO observed, does not only identify a person: it enables them to exercise their rights and forms

¹²⁶ Copyright, Designs and Patents Act, section 12(7) (protection for such works is limited to 50 years, rather than 70 years after the death of the author), section 79 (exception to moral rights).

¹²⁷ Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence (World Intellectual Property Organisation, WIPO/IP/AI/2/GE/20/1 REV, 21 May 2020), para 23. See also M du Sautoy, *The Creativity Code: Art and Innovation in the Age of AI* (Harvard UP 2019) 102; R Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* (CUP 2020) 71-91.

¹²⁸ The patents were for a ‘food container’ and ‘devices and methods for attracting enhanced attention’. DABUS is an acronym for Device for the Autonomous Bootstrapping of Unified Sentience.

¹²⁹ Patents Act 1977 (UK), sections 7, 13. See *Whether the Requirements of Section 7 and 13 Concerning the Naming of Inventor and the Right to Apply for a Patent Have Been Satisfied in Respect of GB1816909.4 and GB1818161.0 (BL O/741/19)* (4 December 2019) (UK Intellectual Property Office), paras 14-20. The tribunal went on to observe that Thaler could not have acquired ownership from DABUS ‘as the inventor cannot itself hold property’ (para 23).

¹³⁰ European Patent Convention, done at Munich, 5 October 1973, in force 7 October 1977, art 81, rule 19(1).

part of their personality. ‘Things’, by contrast, ‘have no rights which a name would allow them to exercise.’¹³¹

The US application was also rejected, based in part on the fact that relevant statutes repeatedly referred to inventors using ‘pronouns specific to natural persons’ such as ‘himself’ and ‘herself’. The Patent and Trade Office (USPTO) cited cases holding that conception — ‘the touchstone of inventorship’ — is a ‘mental act’ that takes place in ‘the mind of the inventor’. Those cases concluded that invention in this sense is limited to natural persons and not corporations. The USPTO concluded that an application listing an AI as an ‘inventor’ was therefore incomplete, but was careful to avoid making any determination concerning ‘who or what’ actually created the inventions in question.¹³²

These decisions were consistent with case law and the practice of patent offices around the world, none of which — yet — allows for an AI system to be recognized as an inventor. Analogous to copyright law, one purpose of the patent system is to encourage innovation by granting a time-limited monopoly in exchange for public disclosure. As even the creators of DABUS acknowledged, an AI system is unlikely to be motivated to innovate by the prospect of patent protection. Any such motivation would be found in its programming: it must be instructed to innovate.¹³³

As for whether a human ‘inventor’ could be credited for work done by such a system, there is no equivalent of the work for hire doctrine. To be an inventor, the human must have actually conceived of the invention.¹³⁴ Joint inventions are possible and contributions do not need to be identical, but in the absence of a natural person making a significant conceptual contribution then an invention would be, on current law, ineligible for patent protection.¹³⁵

¹³¹ *Grounds for the EPO Decision on Application EP 18 275 163* (27 January 2020) (European Patent Office), para 22.

¹³² *In re Application No.: 16/524,350 (Decision on Petition)* (22 April 2020) (US Patent and Trademark Office).

¹³³ *Whether the Requirements of Section 7 and 13 Concerning the Naming of Inventor and the Right to Apply for a Patent Have Been Satisfied in Respect of GB1816909.4 and GB1818161.0* (n 129), para 28.

¹³⁴ *Manual of Patent Examining Procedure (MPEP)* (9th edn, US Patent and Trademark Office 2017), §2137.01.

¹³⁵ JA Cubert and RGA Bone, ‘The Law of Intellectual Property Created by Artificial Intelligence’ in W Barfield and U Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar 2018) 418. There is a tenuous argument that scope for interpretation may lie in the fact that ‘inventor’ is defined in US law as the person who ‘invents or discovers’ the subject matter of the invention. 35 USC § 101. See, eg, R Abbott, ‘I Think, Therefore I Invent: Creative Computers and the Future of Patent Law’ (2016) 57 Boston College L Rev 1098; WM Schuster, ‘Artificial Intelligence and Patent Ownership’ (2018) 75 Washington & Lee L Rev 1977; K Hartung, ‘Dear USPTO: Patents for Inventions by AI Must Be Allowed’, *IP Watchdog*, 21 May 2020.

Among the interesting aspects of these recent developments are the means by which the same conclusion was reached. As in the case of copyright, there appeared to be no serious doubt that an AI system was capable of creating things that were patentable inventions. The British IP Office explicitly accepted that DABUS had done just that;¹³⁶ the USPTO was at pains to avoid making such a conclusion explicit. The EPO, for its part, dodged the issue by holding that, because they have no legal personality, ‘AI systems and machines cannot have rights that come from being an inventor’.¹³⁷ The EPO was the most blatant, but all three decisions relied on formalism — language in the relevant statute that provided or implied that the rights in question were limited to natural persons. The USPTO diligently dusted off a copy of *Merriam-Webster’s Collegiate Dictionary* to conclude that the use of ““whoever” suggests a natural person.’¹³⁸

Legal tribunals routinely face choices between substantive justice and procedural regularity. Statutes of limitations are intended to provide certainty in legal relations; the courts of equity emerged to temper that certainty with justice. In both copyright and patent law, continuing to privilege human creativity over its machine equivalent may ultimately need to be justified by the kind of speciesism mentioned earlier.¹³⁹ At times, however, the language used to engage in such rationalizations of the status quo echoes older legal forms that kept property relations in their rightful place. Among its reasons denying that AI systems like DABUS could hold or transfer rights to a patent, for example, the EPO dismissed analogies between machines and employees: ‘Rather than being employed,’ the EPO concluded, ‘they are owned.’¹⁴⁰

Such statements are accurate for the time being. But if the boosters of general AI are correct and some form of sentience is achieved, the more appropriate analogy between legal personality and slavery may not be the limited economic rights slaves held in ancient Rome, but the constraints imposed on AI systems today.

¹³⁶ *Whether the Requirements of Section 7 and 13 Concerning the Naming of Inventor and the Right to Apply for a Patent Have Been Satisfied in Respect of GB1816909.4 and GB1818161.0* (n 129), para 15.

¹³⁷ *Grounds for the EPO Decision on Application EP 18 275 163* (n 131), para 27.

¹³⁸ *In re Application No.: 16/524,350* (n 132) 4.

¹³⁹ See above n 83.

¹⁴⁰ *Grounds for the EPO Decision on Application EP 18 275 163* (n 131), para 31.

III. CONSTRAINING SUPERINTELLIGENCE

If AI systems do eventually match human intelligence it seems unlikely that they would stop there. The prospect of AI surpassing human capabilities has long dominated a popular sub-genre of science fiction.¹⁴¹ Though most serious researchers do not presently see a pathway even to general AI in the near future, there is a rich history of science fiction presaging real world scientific innovation.¹⁴² Taking Nick Bostrom's definition of superintelligence as an intellect that greatly exceeds human cognitive performance in virtually all relevant domains,¹⁴³ it is at least conceivable that such an entity could be created within the next century.¹⁴⁴

The risks associated with that development are hard to quantify.¹⁴⁵ Though a malevolent superintelligence bent on extermination or enslavement of the human race is the most dramatic scenario, more plausible ones include a misalignment of values, such that the ends desired by the superintelligence conflict with those of humanity, or a desire for self-preservation, which could lead such an entity to prevent humans from being able to switch it off or otherwise impair its ability to function. An emerging literature examines these questions of what final and instrumental goals a superintelligence might have,¹⁴⁶ though the discourse was long dominated by voices far removed from traditional academia.¹⁴⁷ Visa Kurki's recent book length discussion of legal personality, for example, includes a chapter

¹⁴¹ See, eg, H Harrison, *War with the Robots* (Grafton 1962); PK Dick, *Do Androids Dream of Electric Sheep?* (Doubleday 1968); AC Clarke, *2001: A Space Odyssey* (Hutchinson 1968).

¹⁴² P Jordan *et al*, 'Exploring the Referral and Usage of Science Fiction in HCI Literature' (2018) arXiv 1803.08395v2.

¹⁴³ N Bostrom, *Superintelligence: Paths, Dangers, Strategies* (OUP 2014) 22. Early speculation on superintelligence is typically traced to IJ Good, 'Speculations Concerning the First Ultra-intelligent Machine' in FL Alt and M Rubinoff (eds), *Advances in Computers* (Academic 1965) vol 6, 31. Turing himself raised the possibility in a talk in 1951, later published as AM Turing, 'Intelligent Machinery, A Heretical Theory' (1996) 4 *Philosophia Mathematica* 259-60; he in turn credited a yet earlier source — Samuel Butler's 1872 novel *Erewhon*.

¹⁴⁴ See, eg, O Etzioni, 'No, the Experts Don't Think Superintelligent AI is a Threat to Humanity', *MIT Technology Review* (20 September 2016). He cites a survey of 80 fellows of the American Association for Artificial Intelligence (AAAI) on their views of when they thought superintelligence (as defined by Bostrom) would be achieved. None said in the next 10 years, 7.5 per cent said in 10-25 years; 67.5 per cent said in more than 25 years; 25 per cent said it would never be achieved.

¹⁴⁵ P Bradley, 'Risk Management Standards and the Active Management of Malicious Intent in Artificial Superintelligence' (2019) 35 *AI & Society* 319; A Turchin and D Denkenberger, 'Classification of Global Catastrophic Risks Connected with Artificial Intelligence' (2020) 35 *AI & Society* 147.

¹⁴⁶ See, eg, O Häggström, 'Challenges to the Omohundro-Bostrom Framework for AI Motivations' (2019) 21 *Foresight* 153.

¹⁴⁷ See DJ Chalmers, 'The Singularity: A Philosophical Analysis' (2010) 17(9-10) *J Consciousness Studies* 7.

specifically on personality of AI that concludes with the statement: ‘Of course, if some AIs ever become sentient, many of the questions addressed in this chapter will have to be reconsidered.’¹⁴⁸

In the face of many unknown unknowns, two broad strategies have been proposed to mitigate the risk. The first is to ensure any such entity can be controlled, either by limiting its capacities to interact with the world or ensuring our ability to contain it, including a means of stopping it functioning: a kill switch.¹⁴⁹ Assuming that the system has some kind of purpose, however, that purpose would most likely be best served by continuing to function. In a now classic thought experiment, a superintelligence tasked with making paperclips could take its instructions literally and prioritize that above all else. Humans who might decide to turn it off would need to be eliminated, their atoms deployed to making ever more paperclips.¹⁵⁰

Arguments that no true superintelligence would do anything quite so daft rely on common sense and anthropomorphism, neither of which should be presumed to be part of its code. A true superintelligence would, moreover, have the ability to predict and avoid human interventions or deceive us into not making them.¹⁵¹ It is entirely possible that efforts focused on controlling such an entity may bring about the catastrophe that they are intended to prevent.¹⁵²

For this reason, many writers prioritize the second strategy, which is to ensure that any superintelligence is aligned with our own values — emphasizing not what it *could* do, but what it might *want* to do. This question has also fascinated science fiction writers, most prominently Isaac Asimov, whose three laws of robotics are a leitmotif in the literature.¹⁵³ Here the narrower focus is on whether granting AI systems legal personality in the near term might serve as a hedge against the risks of superintelligence emerging in the future.

This is, in effect, another instrumental reason for granting personality. Of course, there is no reason to assume that including AI systems within human social structures and treating them ‘well’ would necessarily lead to them reciprocating the favour should they assume

¹⁴⁸ Kurki (n 15) 189.

¹⁴⁹ See generally Bostrom (n 143) 127-44.

¹⁵⁰ N Bostrom, ‘Ethical Issues in Advanced Artificial Intelligence’ in I Smit and GE Lasker (eds), *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence* (2003) vol 2, 12.

¹⁵¹ J Danaher, ‘Why AI Doomsayers Are Like Sceptical Theists and Why It Matters’ (2015) 25 *Mind and Machines* 231.

¹⁵² W Totschnig, ‘The Problem of Superintelligence: Political, not Technological’ (2019) 34 *AI & Society* 907.

¹⁵³ I Asimov, ‘Runaround’, *Astounding Science Fiction*, March 1942.

dominance.¹⁵⁴ Nevertheless, presuming rationality on the part of a general AI, various authors have proposed approaches that amount to socializing AI systems to human behaviour.¹⁵⁵ To avoid the sorcerer's apprentice problem of a machine simply told to 'make paperclips', for example, its goals could be tied to human preferences and experiences. This might be done by embedding those values within the code of such systems prior to them achieving superintelligence. Goals would thereby be articulated not as mere optimization — the number of paperclips produced, for example — but as fuzzier objectives such as maximizing the realization of human preferences,¹⁵⁶ or inculcating a moral framework and a reflective equilibrium that would match the progressive development of human morality itself.¹⁵⁷

To a lawyer, that sounds a lot like embedding these new entities within a legal system.¹⁵⁸ If one of the functions of a legal system is the moral education of its subjects, it is conceivable that including AI in this way could contribute to a reflective equilibrium that might encourage an eventual superintelligence to embrace values compatible with our own.

For the present, this is proposed more as a thought experiment than a policy prescription. If a realistic path to superintelligence emerges, it might become a more urgent concern.¹⁵⁹ There is no guarantee that the approach would be effective, but some small comfort might be taken from the fact that, for the most part, the categories of legal persons recognized in most jurisdictions, along with the rights they enjoy, have tended to expand over time rather than contract. Those regimes that have taken such legal recognition away have tended to be among the most despicable. Apocalyptic scenarios aside, positioning ourselves and our silicon

¹⁵⁴ Turner (n 40) 164.

¹⁵⁵ R Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking 2005) 424 ('Our primary strategy in this area should be to optimize the likelihood that future nonbiological intelligence will reflect our values of liberty, tolerance, and respect for knowledge and diversity. The best way to accomplish this is to foster those values in our society today and going forward.');

N Soares and B Fallenstein, 'Agent Foundations for Aligning Machine Intelligence with Human Interests: A Technical Research Agenda' in V Callaghan et al (eds), *The Technological Singularity: Managing the Journey* (Springer 2017) 117-20.

¹⁵⁶ SJ Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking 2019). Cf Bostrom's suggestion that the goal for a superintelligence might be expressed as 'achieve that which we would have wished the AI to achieve if we had thought about the matter long and hard'. Bostrom (n 143) 141.

¹⁵⁷ E Yudkowsky, 'Complex Value Systems in Friendly AI' in J Schmidhuber, KR Thórisson, and M Looks (eds), *Artificial General Intelligence* (Springer 2011) 388.

¹⁵⁸ Cf S Omohundro, 'Autonomous Technology and the Greater Human Good' (2014) 26 *J Experimental & Theoretical Artificial Intelligence* 308.

¹⁵⁹ In the event that that path lies through augmentation of humans rather than purely artificial entities, those humans are likely to remain subjects of the law.

siblings as equals might serve the goal of reinforcing a normative regime in which our interests are aligned, or at least not opposed, if or when we are surpassed.¹⁶⁰

In the alternative, like the chimpanzees in their New York cages, humanity's greatest hope may not be to be treated as peers, but at least to be seen as more than things.

IV. CONCLUSION: THE LIMITS OF PERSONALITY

In 1991 a prize was established to encourage more serious attempts at the Turing Test. One of the first winners succeeded in part by tricking people — the program made spelling mistakes that testers assumed must have been the result of human fallibility.¹⁶¹ Though the Turing Test remains a cultural touchstone, it is far from the best measure of AI research today. As a leading textbook notes, the quest for flight succeeded when the Wright brothers stopped trying to imitate birds and started learning about aerodynamics.¹⁶² Aeronautical engineers today don't define the goal of their field as making machines that fly so exactly like pigeons that they can fool other pigeons.

In the same way, most arguments in favour of AI legal personality suffer from being both too simple and too complex. They are too simple in that AI systems exist on a spectrum with blurred edges. There is as yet no meaningful category that could be identified for such recognition; if instrumental reasons required recognition in specific cases then this could be achieved using existing legal forms. The arguments are too complex in that many are variations on the android fallacy, based on unstated assumptions about the future development of AI systems for which personality would not only be useful but deserved. At least for the foreseeable future, the better solution is to rely on existing categories, with responsibility for wrongdoing tied to users, owners, or manufacturers rather than the AI systems themselves. Driverless cars are following that path, for example, with a likely shift from insuring drivers to insuring vehicles.¹⁶³

This may change. It is conceivable that synthetic beings of comparable moral worth to humans may one day emerge. Failing to recognize that worth may reveal us to be either an 'autistic

¹⁶⁰ Cf S Livingston and M Risse, 'The Future Impact of Artificial Intelligence on Humans and Human Rights' (2019) 33 *Ethics and International Affairs* 141.

¹⁶¹ 'Artificial Stupidity', *The Economist* (1 August 1992).

¹⁶² Russell and Norvig (n 1) 3.

¹⁶³ Chesterman, 'Autonomy' (n 5) 220.

species', unable to comprehend the minds of other types of beings,¹⁶⁴ or merely prejudiced against those different from ourselves. If this happens, as Turing hypothesized in 1951, 'it seems probable that, once the machine thinking method had started, it would not take long to outstrip our feeble powers.'¹⁶⁵

Turing himself never lived to see a computer even attempt his test in practice. Prosecuted for homosexual acts in 1952, he chose chemical castration as an alternative to prison. He died two years later at the age of 41, apparently after committing suicide by eating a cyanide-laced apple. The announcement that Turing would grace the new 50 pound note followed an official pardon, signed by the Queen in 2013.¹⁶⁶

Yet the more fitting tribute may be Ian McEwan's novel, *Machines Like Me*, which imagines an alternative timeline in which Turing lived and was rewarded with the career and the knighthood he deserved. The novel takes seriously the prospect of true AI, in the form of a brooding synthetic Adam, who expresses his love for the human Miranda by writing thousands upon thousands of haikus. Ultimately, however, consciousness is a burden for the machines — struggling to find their place in the world; so pure that they are unable to reconcile human virtues and human vices.

It also offers Turing a chance to rethink his test. 'In those days,' the fictional Turing says at age 70, referring to his younger self, 'I had a highly mechanistic view of what a person was. The body was a machine, an extraordinary one, and the mind I thought of mostly in terms of intelligence, which was best modelled by reference to chess or maths.'¹⁶⁷

The reality, of course, is that chess is not a representation of life. Life is an open system; it is messy. It is also unpredictable. In the novel, the first priority of the AI robots is to disable the kill switch that might shut them down. Yet most of them ultimately destroy themselves — as the real Turing did — unable to reconcile their innate nature with the injustices of the world around them. Before asking whether we could create such thinking machines, McEwan reminds us, we might want to pause and ask whether we should.

¹⁶⁴ Chopra and White (n 41) 191.

¹⁶⁵ AM Turing, *Intelligent Machinery, a Heretical Theory* (lecture given to the '51 Society' at Manchester) (Turing Digital Archive, AMT/B/4, 1951) 6.

¹⁶⁶ See generally D Turing, *Prof Alan Turing Decoded: A Biography* (History Press 2015).

¹⁶⁷ I McEwan, *Machines Like Me* (Vintage 2019) 300.