Robotic intelligence: philosophical and ethical challenges¹

David Lawrence

This chapter focuses on one field of scientific research (or rather a collection of many subfields) which has the potential to bring about epochal changes of a magnitude not seen since humankind's first forays into tool use – our first steps into differentiating ourselves from other animals. Artificial intelligence (AI) – along with advanced robotics – promises significant effects for our way of life, of working, and of interacting with others. It may even bring about the first time we encounter an equal – or a better – through the development of conscious, thinking, sapient machines. Regulation and policy around the advancement of these technologies presents different challenges – the former raises questions around liability, ownership, employment and more; but the latter presents new issues with no precedent. A sapient intelligence may in effect be a novel being, a potential person – and there is good reason to think we should treat it as such. The possibility of non-*Homo sapiens* persons raises questions about the very nature of humanity and our place at the top of the moral status 'ladder'. Defenders of scientific freedom may have a tall mountain to climb to justify the risks of so momentous a change in society.

Advances in these technologies are already affecting the world of work. A survey of the 100 most cited academics writing on AI suggests an expectation that machines will be developed 'that can carry out most human professions at least as well as a typical human' (Müller and Bostrom 2016) with 90 per cent confidence by 2070, and with 50 per cent confidence by 2050. While these figures are speculative, based on educated guesswork and projections of the rate of technical development, it must be stressed that the prototypes and experimental robots extant today are extremely impressive devices. It is possible to emulate proprioception, tactility (Syntouch 2016), visual processing and object recognition, walking and running (Honda 2016a) – even on rough terrain and at high speeds (Raibert et al. 2008) – and many more elements of human biology through recent advancements in microelectronics and combined systems. Even the high-speed recognition, analysis and reaction needed to play table tennis (*Popular Science* 2010)

can be emulated – perhaps a useless talent for a robot, but one which demonstrates the sophistication that already exists and has been publicly announced. Automation and robotics have long been a feature of the workforce, for example, in the automotive manufacturing industry, but are now in a position to start taking more subtle, customer-facing jobs. ASIMO, Honda's famous walking robot, has acted as a receptionist (*New Scientist* 2005), and has acted intelligently in concert with other ASIMOs as a team of office assistants (Honda 2016b); Softbank's Pepper has become a store attendant in over 100 locations (Nagata 2014). Support line jobs are beginning to be overtaken by chatbots able to provide answers from extensive databases – providing the same service as a human agent, but for far less outlay and running cost (Meister 2017). The threat of automation hangs over many industries (*Financial Times* 2016), and robots are even expected to move into the 'educated professions' such as law and medicine in the near future – with potentially greater efficiency and accuracy than human practitioners (Meltzer 2014).

Predictions range from 47 per cent in the United States, 69 per cent in India, 77 per cent in China, to 85 per cent in Ethiopia of jobs becoming robotically automated or deferred to AI (Osborne and Frey 2015). Some predict the rise of a class of 'useless' people – those unable to work because they have no skills that are not provided better and more cheaply by a robot (Harari 2016). There appears to be a real risk of a global crisis of unemployment if the trend of increasing automation is realised. The potential societal impacts at hand, therefore, are significant.

Furthermore, AI is posited by some as one of – if not *the* – greatest potential threat to humanity. Academic literature provides familiar arguments for this, which are broadly speculative. Russell and Norvig (2003) tell us that AI 'may . . . evolve into a system with unintended behavior' which could manifest in any number of ways that threaten our lives or freedoms. This may not be malicious – a common line of reasoning is that '[t]he AI does not hate you, nor does it love you, but you are made out of atoms which it can use for something else' (Yudkowsky 2011) – which is to say that an AI might value the completion of its own goals over the preservation of *Homo sapiens*, or perhaps would be so driven to complete its task that all other matters are subsidiary. Where there might be attempts to program a moral code to govern such actions and prevent harm to us in pursuit of a specified goal or purpose, critics hold that this would prove almost impossible to accomplish due to the lack of a perfect ethical theory (Muehlhauser and Helm 2012). Any value system that could be bestowed upon an AI would necessarily be flawed, with internal conflicts that we might be forced to concede or capitulate from to avoid a harmful situation. An AI (unless it was fully sapient and capable of nuanced reason, as shall be discussed later), in applying the system rigidly, would fail to avoid this harm. There is also an argument commonly made that conflict is inevitable, that peaceful coexistence is impossible (Lawrence et al. 2016), with the motivations and

goals of an AI being necessarily incompatible with our own, thus forcing one species or the other to dominate.

Beyond academic journals, there has been an exponential increase in the number of media articles and think pieces published over the last two to three years, with the frequency reaching at least several per day in UK media alone in early 2017. Many follow the above trend, presenting AI and robotic technologies as looming threats. Titles² such as ‘The real problem with artificial intelligence’ (Thompson 2015), ‘Why you should fear artificial intelligence’ (Huston 2016), ‘Artificial intelligence: “we’re like children playing with a bomb”’ (Adams 2016), ‘Artificial intelligence to take over half of all jobs in next decade’ (RT International 2017), and ‘Has humanity already lost control of Artificial Intelligence?’ (Best 2017) are commonplace, and range from reasonable discussion to tabloid fearmongering – much as with any controversial technology. Public figures in science and technology, those few who possess such a platform, have proffered their fears and warnings to endorse the idea of AI as threat – most notably Elon Musk, Stephen Hawking and Bill Gates. Gates ‘cannot understand why some people are not concerned’ (Rawlinson 2015), while Hawking warns that the technologies ‘could spell the end of the human race’ (Cellan-Jones 2014) – an idea mirrored by Musk’s claim that AI is ‘[p]otentially more dangerous than nukes’ (Rodgers 2014).

These criticisms and fears are broadly based on an assumption – that AI, even ‘superintelligent’ AI, will necessarily operate as does any other algorithm, following in effect a tremendously complex flow diagram of queries, checks and responses. An intelligence of this type which surpasses our own raw cognitive-processing power might warrant being called ‘super’ as it could, in a narrow sense, outperform us. This is only one of several potential conceptions of AI, though it may be fair to say that it is the most likely, as it is the one that presently exists – albeit probably without yet qualifying as ‘super’. We can see examples in any of the AI which we utilise as individuals and as a society every day – from simple algorithms as used by streaming television services such as Netflix to recommend shows based on your viewing history (Gomez-Uribe and Hunt 2016), to stock market trading programs (Dymova et al. 2016), to the complex Bayesian systems which operate autopiloting systems in aircraft and autonomous cars (Zhu et al. 2014). These are all ‘expert systems’ (Cuddy 2002) or ‘applied’ AI (sometimes known as ‘weak’ AI; Searle 1980) – based on the combination of a knowledge base and an inference engine. In effect, the system is pre-programmed to recognise data and to respond in a certain manner – so, for instance, Netflix’s AI might recognise your habit of watching Hong Kong action films, infer a penchant, and promote Andy Lau or Donnie Yen pictures to you. Similarly, an autonomous car might detect a sudden obstacle ahead and another vehicle pulling alongside, infer the risk of collision and choose to swerve the opposite way. These systems are not making decisions in the manner of a human, using reasoning and intuition to consider cause

and effect, but are instead applying their own type of first-order logical rules (Forgy 1982), which might at a very simple level be summed up as 'if X, then Y'.

These types of system are pervasive, and are involved in almost everything that utilises digital automation. They are, in effect, so immersed in the fabric of our society that they *are* that society. It may well be that humanity could continue without applied AI, as we managed for many millennia, but it is certain that we could not operate in the same way as we do today. Nor could we enjoy the many benefits of these systems that we take for granted. Scientific progression in these fields, and its 'trickle down' into the smallest parts of our lives, has fundamentally altered the human experience. This has been a great benefit to those fortunate enough to enjoy it – and it is a great argument in favour of having the freedom to do so.

However, the influence of these systems, this irreversible interweaving of science and society, leaves us at a crossroads. Further integration of weak AI into our lives, or the pursuit of 'strong' (Kurzweil 2005) or 'general' (Newell and Simon 1976) AI (that can go beyond problem solving into human-level cognition) through the free practice of science, is likely to cause more direct changes to who and what we are. Our place in the hierarchy of beings, even our relative position as the pinnacle of moral status, could be forever altered.

There are a number of highly complex subfields within AI research working towards different elements of human-level cognitive function. For example, a true, conscious artificial general intelligence (AGI) would need to be able to perceive and understand information (Russel and Norvig 2003: 537–81, 863–98); to learn (Langley 2011); to process language (Cambria and White 2014); to plan ahead and anticipate (and so visualise itself in time, an important element of most philosophical conceptions of personhood) (Russel and Norvig 2003: 375–459); to possess 'knowledge representation' (Russel and Norvig 2003: 320–63) or the ability to retain, parse and apply the extreme number of discrete facts, truths and logical paths that we take for granted, and be able to use this information to reason; to possess subjectivity; and much more. A number of projects are ongoing, attempting to develop and integrate one or more of these functions into 'artificial brains', using digitally modelled neural networks and other technologies. These include Cyc (*New Scientist* 2006; Cycorp 2016),³ an ongoing 33-year effort to collect and incorporate a vast database of 'common-sense' knowledge equivalent to that which would have been gathered by an adult human in a practical ontology, to enable self-directed reasoning independent of instructions and predetermined action. There is also the Google Brain (Hernandez 2013), a 'deep learning' project to use Google's vast troves of data as a knowledge base and allow the AI to begin to parse things for itself through cross-referencing and recognition. For instance, the Brain, when given access to image files and clips on YouTube, learned unprompted to recognise and identify human faces in motion, and showed a partiality

to videos of cats (WIRED 2012). A third project, the well-known Blue Brain, has successfully modelled 37,000,000 synapses of a rat's sensory cortex (Markram et al. 2015) in an attempt to understand the 'circuitry' with the aim of synthetic replication.

It is a matter of contention as to whether AGI will ever actually come to pass, whether we will ever create, or cause to be created, a conscious being equal to or greater than ourselves. It may be that we are incapable of the technical heights that would be necessary to produce true thought within a machine. If this is the case, then we can only gain from the attempt – success in any one of the fields mentioned above would be a prime example of a dual-use technological development; for instance, true natural language processing would revolutionise international dealings of all kinds. The failure would also, presumably, cement us as the dominant life form.

But what if we were to succeed?

Whether or not the fears of conservative commentators were realised it would be a triumph of science if we were to bring about AGI; probably the greatest such triumph possible. It would mean the creation of new life in a form unlike anything we are familiar with, unlike the possibilities of synthetic biology or genome editing to alter existing beings. For this reason alone, some might argue it is a worthwhile pursuit – for the sheer pioneering possibility and the expansion of our horizons.

The idea of an AI which is able to match us in cognitive ability – not merely outdo us in the capacity to process data but rather one which can perform all the myriad mental acts that make up our conscious mind – frightens many. Perhaps conditioned by decades of science fiction in which the intelligent robots rise up and overthrow their *Homo sapiens* masters, or enact purges against us, the default reaction tends to assume malevolence. News articles invariably use images taken from such films – at least two of the works cited in this chapter contain an allusion to *The Terminator*, with at least one more using a still from the film depicting a killer android (Rodgers 2014). This is an all too common trend with reportage of science and technology generally, but it may have a particularly strong negative impact in this case, as the technology is largely theoretical. Unlike the 'Brave New World' backlash to IVF treatment, which was relatively easily disproven by showing the reality of the science and its results (Harris and Lawrence 2017), it is difficult to beat the appeal of bombastic summer blockbusters with comparatively dry ethical theory.

The idea does not necessarily hold together if we think about what exactly a conscious AGI would be – even assuming that it inhabits a robot body – or what it is likely to want. If an AI is our cognitive equal or better, it must necessarily possess the same faculties as we do – including those which grant us a certain moral status and value. If the measure of this for *Homo sapiens* is to have crossed the threshold for personhood (i.e. per Taylor, Harris and others: having 'a sense of self, a notion of the future and the past, [an ability to] hold values, make choices'; Taylor 1985) through

possessing self-awareness, moral agency and continuous narrative (Lawrence 2017), and our AGI matches these faculties, it must, perforce, qualify as a person.

There are good reasons for thinking this. Any AI possessing human-equivalent intelligence is by default self-aware and conscious: reactivity would merely be the domain of an expert system whereas an AGI worthy of the name must be able to act in a considered fashion as a moral agent. Furthermore, a being without narrative identity would be unable to act in any meaningful way, let alone consider its actions. By fulfilling the requirements of personhood it surely follows that our digital consciousness proves itself deserving of the protections due to a *sapiens* person.⁴ Where we consider legal protections for a group it is because we see that group as possessing whatever level of moral value is worthy of that protection (i.e. that we consider ourselves to possess), and personhood appears to be the qualifying requirement for this. The second major argument on this point centres around animal personhood. A number of legal challenges (*Nonhuman Rights Project v. Lavery* 2014; McKinley 2015; *Matter of Nonhuman Rights Project v. Stanley* 2015; CNN 2016) have been brought to seek legal personhood for great apes, some of which have been successful to greater or lesser degrees. There is no reason that the same consideration ought not be given to other non-*Homo sapiens* beings. If some animals can be judged to have attained sufficient characteristics to be persons, then it follows that AGI which are demonstrably our cognitive equals would be so.

As with other beings we have encountered throughout history – i.e. citizens of foreign nations – personhood is no guarantee of non-aggression or automatic friendship. However, we do not generally assume hostility from other persons. Rather, our shared moral status gives us grounds for understanding. Conflict generally arises where there are incompatible motivations. There is no good reason to assume that things would be fundamentally different with a non-biological conscious being.

As mentioned, conflict will arise – and techno-conservative fears will be realised – if the goals of a novel consciousness are in contravention of our own. It stands to reason that this could be the case – quite what an AGI might want to accomplish, we cannot say, any more than we can guess at the intentions of anyone we meet.⁵ However, we can make some inferences from what we know about their (theoretical) nature as moral agents and persons equivalent to ourselves. Whereas our most basic motivations are driven by the pursuit of survival and reproduction – the continuation of our species – an AI existing as digital code may be able to reproduce infinitely and spontaneously by copying itself. As a person, an AGI must have a sense of narrative identity, and so it follows that it would wish to preserve and realise its vision of itself in the future. An AI needs neither food nor water, and in order to continue existing in the same form as it begins its ‘life’ it is unlikely to have many material needs beyond energy.⁶ However, there is some evidence in simple autonomous robots

(Studer and Lipson 2006) of a spontaneous reproductive strategy emerging that follows the standard biological imperative. If this is indeed the case, and AI housed in robotic shells or machine bodies aimed to replicate – or indeed repair – themselves, resources would be required. The limits on supply could lead to conflict between an AGI and *Homo sapiens*, but this once again seems to assume the worst. Modern man – let alone a superior intelligence possessing a greater moral capacity – does not immediately leap to subjugate those who have something we need (or at least we know this is not the ethically sound path, even if we struggle as a species to stick to it). Instead we engage in negotiation and trade to acquire resources, and it is not clear why an equivalent conscious intelligence would not adopt a similar approach.

If this is so, there are significant implications for how we would see ourselves. No longer would we be the foremost species, nor the most intelligent. We would also no longer be the only member of our community of moral value, something for which the term ‘human’ has previously been a convenient label (Lawrence 2016). If we were to be interacting with and sharing our society with AGIs, it might follow that we can no longer refer to conscious society as being ‘human’ society. Alternatively, it might follow that the term should be expanded to include other conscious beings. In either case there would be a significant re-evaluation required of our ideas about who we are and our place in the world. Scientific development would have the potential to reforge the very concept of humanity; and rather than dominance, our position could become more one of transaction and compromise with equals.⁷

In his *Superintelligence* Nick Bostrom (2014) suggests that we may find ourselves reliant on the mercy and goodwill of an AI, much as do endangered species depend on ours. This is one possibility. Another seems more realistic: the opposite. Any advanced AI with the potential to surpass us and become ‘super’ by learning and modifying itself will not instantaneously pop into being. It seems likely that the hypothetical AI’s fate will be reliant on our goodwill and mercy, much like the mountain gorilla’s, at least for the period of its ‘childhood’, for want of a better term. For example, it would be prudent, when switching on the AI, to do so in a secure manner wherein there was no risk of ‘escape’ – this is referred to as ‘boxing’ the AI (Chalmers 2010). This would presumably involve safeguards such as ensuring no connection to the internet, using air gaps and Faraday cages, among other things. Isolated in this manner, the newborn AI would be reliant on our not choosing to cut the power or delete the program – thereby killing it. Whatever resources it might require would be bestowed by us, whatever interaction or stimulus it received would be portioned and tightly controlled. There are definite ethical issues with this – from the obstruction of autonomy through to harm by imprisonment – if the AI possessed anything approximating personhood and/or consciousness, but it seems highly unlikely that the development of such a being would be conducted in any other way.

This should remind us that there is a deeper story to be told about the weight of scientific advancements. As the stewards of scientific progress, we are beholden to all parties – both to existing persons, and to the beings we may create through artificial intelligence research. The risks and fears surrounding artificial intelligence are purely our problems to solve, or to prevent from arising through careful design and the implementation of appropriate regulation and policy to govern their development. This work is presently beginning – already bodies within nations likely to drive the research and technologies in question are exploring the challenges and proposing their own means of addressing them. Reports such as the White House National Science and Technology Council Committee on Technology’s *Preparing for the Future of Artificial Intelligence*, the UK House of Commons’ Science and Technology Committee *Report on Robotics and Artificial Intelligence*, and the European Parliament’s *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics* all emerged at the end of 2016, though it should be said that none of these documents are definite regulatory roadmaps. They do, however, aim to provide a basis for controlling the integration of AI and our lives – to bridge the gap between science and society in a controlled manner. Whether the suggestions will be effective is yet to be seen, but the fact that these documents exist is a promising start. What we must ensure, though, is that we consider reality – whether advanced technological development is permitted or tightly controlled, there will always be the chance that it is developed in secret and beyond regulatory reach.

When we consider the potential stakes – smart systems that could upend our society, or the birth of AGI that could think and reason like a human, with wants and needs and perhaps moral rights of its own, there is probably good reason to want to ‘get in front’ of these challenges; but it does not follow that we *should*. To try to control or limit the development of robotics and AI⁸ may prevent responsible and conscientious parties from doing so, but it will not stop others. With the potential impacts so significant, it seems that the sensible approach would be to ensure that freedom to act rests in the hands of those most able (or those likely to be so) to do so appropriately and with consideration for the consequences. Guidelines and regulations that attempt to control technologies after the fact are rarely great successes, and with one as ephemeral as an AI (of any type) it will be all the more difficult. Furthermore, with regard to AI the balancing act of scientific freedom and the preservation of the status quo is a futile endeavour – AI will, no doubt, be the greatest technological challenge to our society, and has already fundamentally altered how we live. We cannot ban their development outright, for so many parts of that society rely on them already. We cannot enforce our rules to prevent them altering themselves and advancing themselves beyond our reach. We cannot, ultimately, seek to control that which we cannot yet imagine.

The final consideration, perhaps, ought to be this: we are rapidly approaching a science that can create a new form of conscious life – perhaps a quantitatively ‘better’ form of life. We have, as has been argued by many (Savulescu and Kahane 2009; Harris 2010; Mautner 2010), a moral imperative – if not an obligation – to create life and more so to create life that stands to live a better life than can we. The reasons for this are myriad, but not the least is the need to ensure our propagation – or that of our inheritors. If conscious AI could facilitate this, if freedom of scientific research could help these beings come to pass, then what risks and reasons against it are truly enough to outweigh that?

Notes

- 1 This chapter draws on research from Lawrence (2017).
- 2 I acknowledge that these titles are somewhat cherry-picked for effect. However, the sheer ease of finding such articles is telling, even if they are interspersed with more positive portrayals – all were within one click of a simple Google News search for ‘artificial intelligence’.
- 3 I thank John Harris for informing me of this fascinating endeavour.
- 4 A number of domestic and international documents of rights provide these protections, and may well be applicable here to form the basis of any legal policymaking addressing this issue.
- 5 Though we can probably say that with an applied AI we would almost certainly program it not to come into conflict with us, and will presumably seek to find means of avoiding the threat of an expert system accomplishing its task to the detriment of all else as hypothesised earlier in this chapter.
- 6 Though conceivably a conscious being could have material wants, if not needs.
- 7 Or, as some would have it, betters.
- 8 Beyond, perhaps, seeking to prevent anything designed to break existing laws – such as autonomous killing machines or AI expressly created to cause harm.

References

- Adams, T. (2016), ‘Artificial intelligence: “We’re like children playing with a bomb”’, *The Guardian*, 12 June, www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine (last accessed 1 May 2017).
- Best, S. (2017), ‘Has humanity already lost control of artificial intelligence?’, *Mail Online*, 11 April, www.dailymail.co.uk/sciencetech/article-4401836/Has-humanity-lost-control-artificial-intelligence.html (last accessed 1 May 2017).
- Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.
- Cambria, E., and White, B. (2014), ‘Jumping NLP curves: a review of natural language processing research’, *IEEE Computational Intelligence Magazine*, 9.2: 48–57.
- Cellan-Jones, R. (2014), ‘Stephen Hawking warns artificial intelligence could end mankind’, *BBC News*, 2 December, www.bbc.co.uk/news/technology-30290540 (last accessed 14 July 2016).

- Chalmers, D. (2010), 'The singularity: a philosophical analysis', *Journal of Consciousness Studies*, 17.9–1: 7–65.
- CNN (2016), 'Orangutan granted controlled freedom by Argentine court', <http://edition.cnn.com/2014/12/23/world/americas/feat-orangutan-rights-ruling> (last accessed 17 July 2016).
- Cuddy, C. (2002), 'Expert systems: the technology of knowledge management and decision making for the 21st century', *Library Journal*, 127.16: 82.
- Cycorp (2016), 'A knowledge modeling and machine reasoning environment capable of addressing the most challenging problems in industry, government, and academia', *Cycorp: Home of Smarter Solutions*, www.cyc.com (last accessed 14 July 2016).
- Dymova, L., Sevastjanov, P., and Kaczmarek, K. (2016), 'A Forex trading expert system based on a new approach to rule-based evidential reasoning', *Expert Systems with Applications*, 1.51: 1–3.
- European Parliament Committee on Legal Affairs (2016), *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, Brussels.
- Financial Times* (2016), 'Why robots are coming for US service jobs', www.ft.com/cms/s/0/cb4c93c4-0566-11e6-a70d-4e39ac32c284.html#axzz4DNsK7QYF (last accessed 14 July 2016).
- Forgy, C. (1982), 'Rete: a fast algorithm for the many pattern/many object pattern match problem', *Artificial Intelligence*, 19.1: 17–37.
- Gomez-Uribe, C. A., and Hunt, N. (2016), 'The Netflix recommender system: algorithms, business value, and innovation', *ACM Transactions on Management Information Systems (TMIS)*, 6.4: 13.
- Harari, Y. N. (2016), *Homo Deus: A Brief History of Tomorrow*, New York: Random House.
- Harris, J. (2010), *Enhancing Evolution: The Ethical Case for Making Better People*, Princeton, NJ: Princeton University Press.
- Harris, J., and Lawrence, D. (2017), 'New technologies, old attitudes, and legislative rigidity', in R. Brownsword, E. Scotford and K. Yeung (eds), *The Oxford Handbook on the Law and Regulation of Technology*, Oxford: Oxford University Press.
- Hernandez, D. (2013), 'The man behind the Google Brain: Andrew Ng and the quest for the new AI', *WIRED*, www.wired.com/2013/05/neuro-artificial-intelligence (last accessed 14 July 2016).
- Honda (2016a), 'ASIMO – the Honda Worldwide ASIMO site', <http://world.honda.com/ASIMO> (last accessed 14 July 2016).
- Honda (2016b), 'Honda develops intelligence technologies enabling multiple ASIMO robots to work together in coordination', http://asimo.honda.com/news/honda-develops-intelligence-technologies-enabling-multiple-asimo-robots-to-work-together-in-coordination/newsarticle_0073 (last accessed 16 February 2017).
- House of Commons Science and Technology Committee (2016), *Report on Robotics and Artificial Intelligence*, HC145, London: Stationery Office.
- Huston, D. (2016), 'Why you should fear artificial intelligence', *Techcrunch*, 22 March, <https://techcrunch.com/2016/03/22/why-you-should-fear-artificial-intelligence> (last accessed 1 May 2017).
- Kurzweil, R. (2005), *The Singularity Is Near*, New York: Viking Press.
- Langley, P. (2011), 'The changing science of machine learning', *Machine Learning*, 82.3: 275–9.

- Lawrence, D. R. (2016), 'The edge of human: the problem with the posthuman as beyond', *Bioethics*, doi:10.1111/bioe.12318
- Lawrence, D. R., Palacios-González, C., and Harris, J. (2016), 'Artificial intelligence: the Shylock syndrome', *Cambridge Quarterly of Healthcare Ethics*, 25.2: 250–61.
- Lawrence, D. R. (2017), 'More human than human', *Cambridge Quarterly of Healthcare Ethics*, 26.3: 476–90.
- McKinley, J. (2015), 'Judge orders Stony Brook University to defend its custody of 2 chimps', *Nytimes.com*, www.nytimes.com/2015/04/22/nyregion/judge-orders-hearing-for-2-chimps-said-to-be-unlawfully-detained.html (last accessed 17 July 2016).
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, and M., Sanchez, C. A. et al. (2015), 'Reconstruction and simulation of neocortical microcircuitry', *Cell*, 163.2: 456–92.
- Matter of Nonhuman Rights Project, Inc. v. Stanley* (2015), NY Slip Op 31419, State of New York Supreme Court, <http://law.justia.com/cases/new-york/other-courts/2015/2015-ny-slip-op-25257.html> (last accessed 17 July 2016).
- Mautner, M. N. (2010), 'Seeding the universe with life: securing our cosmological future', *Journal of Cosmology*, 5.26: 982–4.
- Meister, J. (2017), 'The future of work: the intersection of artificial intelligence and human resources', *Forbes.com*, 1 March, www.forbes.com/sites/jeannemeister/2017/03/01/the-future-of-work-the-intersection-of-artificial-intelligence-and-human-resources/#38621ce56ad2 (last accessed 1 May 2017).
- Meltzer, T. (2014), 'Robot doctors, online lawyers and automated architects: the future of the professions?', *The Guardian*, 15 June, www.theguardian.com/technology/2014/jun/15/robot-doctors-online-lawyers-automated-architects-future-professions-jobs-technology (last accessed 14 July 2016).
- Muehlhauser, L., and Helm, L. (2012), 'Singularity and machine ethics', in A. Eden, J. Søraker, J. H. Moor and E. Steinhart (eds), *Singularity Hypotheses: A Scientific and Philosophical Assessment*, Berlin: Springer, 101–26
- Müller, V. C., and Bostrom, N. (2016), 'Future progress in artificial intelligence: a survey of expert opinion', in V. C. Müller (ed.), *Fundamental Issues of Artificial Intelligence*, Cham: Springer, 553–71.
- Nagata, K. (2014), 'Softbank unveils "historic" robot', *Japan Times*, 5 June, www.japantimes.co.jp/news/2014/06/05/business/corporate-business/softbank-unveils-pepper-worlds-first-robot-reads-emotions/#.U5hbl_m1ZbU (last accessed 1 May 2017).
- National Science and Technology Council Committee on Technology (2016), *Preparing for the Future of Artificial Intelligence*, Washington DC: Executive Office of the President.
- New Scientist* (2005), 'Humanoid robot gets job as receptionist', www.newscientist.com/article/dn8456-humanoid-robot-gets-job-as-receptionist (last accessed 14 July 2016).
- New Scientist* (2006), 'The word: common sense', www.newscientist.com/article/mg19025471.700-the-word-common-sense (last accessed 14 July 2016).
- Newell, A., and Simon, H. A. (1976), 'Computer science as empirical inquiry: symbols and search', *Communications of the ACM*, 19.3: 113–26.
- Nonhuman Rights Project, Inc., on Behalf of Tommy, v. Patrick C. Lavery* (2014), 518336, State of New York Supreme Court, <http://decisions.courts.state.ny.us/ad3/Decisions/2014/518336.pdf> (last accessed 14 July 2016).

- Osborne, M., and Frey, C. B. (2015), *Technology at Work: The Future of Innovation and Employment*, Oxford Martin Programme on Technology and Employment, Oxford: Oxford Martin School and Citi GPS, www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work.pdf (last accessed 2 May 2017).
- Popular Science* (2010), 'A ping-pong-playing terminator', www.popsci.com/technology/article/2010-02/ping-pong-playing-terminator (last accessed 14 July 2016).
- Raibert, M., Blankespoor, K., Nelson, G., and Playter, R. (2008), 'BigDog, the rough-terrain quadruped robot', *Boston Dynamics*, www.bostondynamics.com/img/BigDog_IFAC_Apr-8-2008.pdf (last accessed 14 July 2016).
- Rawlinson, K. (2015), 'Microsoft's Bill Gates insists AI is a threat', *BBC News*, 29 January, www.bbc.co.uk/news/31047780 (last accessed 1 May 2017).
- Rodgers, P. (2014), 'Elon Musk warns of terminator tech', *Forbes*, 5 August, www.forbes.com/sites/paulrogers/2014/08/05/elon-musk-warns-ais-could-exterminate-humanity (last accessed 14 July 2016).
- RT International (2017), 'Artificial intelligence to take over half of all jobs in next decade', 28 April, www.rt.com/business/386452-ae-replace-half-jobs-technologist (last accessed 1 May 2017).
- Russell, S., and Norvig, P. (2003), *Artificial Intelligence: A Modern Approach*, 2nd edn, Upper Saddle River, NJ: Prentice Hall.
- Savulescu, J., and Kahane, G. (2009), 'The moral obligation to create children with the best chance of the best life', *Bioethics*, 23.5: 274–90.
- Searle, J. R. (1980), 'Minds, brains, and programs', *Behavioral and Brain Sciences*, 3.3: 417–57.
- Studer, G., and Lipson, H. (2006), 'Spontaneous emergence of self-replicating structures in molecule automata', in L. M. Rocha (ed.), *Proceedings of the 10th Int. Conference on Artificial Life (ALIFE X)*, Cambridge, MA: MIT Press, 227–33.
- Syntouch (2016), 'Biotac', syntouchllc.com, www.syntouchllc.com/Products/BioTac (last accessed 14 July 2016).
- Taylor, C. (1985), *The Concept of a Person. Philosophical Papers*, Volume 1, Cambridge: Cambridge University Press.
- Thompson, C. (2015), 'The real problem with artificial intelligence', *Business Insider*, 10 September, <http://uk.businessinsider.com/autonomous-artificial-intelligence-is-the-real-threat-2015-9?r=US&IR=T> (last accessed 1 May 2017).
- WIRED (2012), 'Google's artificial brain learns to find cat videos', www.wired.com/2012/06/google-x-neural-network (last accessed 14 July 2016).
- Yudkowsky, E. (2011), 'Artificial intelligence as a positive and negative factor in global risk', in N. Bostrom and M. Cirkovic (eds), *Global Catastrophic Risks*, Oxford: Oxford University Press.
- Zhu, W., Miao, J., Hu, J., and Qing, L. (2014), 'Vehicle detection in driving simulation using extreme learning machine', *Neurocomputing*, 128: 160–5.