

For the love of robots: three principles for social governance of the future of artificial intelligence and the biotech industry

By Michael Lee

Today, a man's best friend is a machine, to paraphrase the familiar English idiom. Our dependency on information and intelligence technologies is increasing at a phenomenal rate, pushing up the scale and speed of change. We have invented, through science and logic, a new techno-smart form of living empowered by efficient, intelligent technologies which continuously improve living conditions and systems.

But what is the end-game of this smart world we've engineered? In particular, what might the wisest approach be towards the growing interdependence of human and artificial intelligence?

One visionary thinker who has helped the world to envisage and comprehend the underlying dynamics giving rise to this Age of Intelligence is inventor, futurist, computer scientist and author Ray Kurzweil. He believes the law of accelerating returns (LOAR¹) will ensure this progression to an ever more intelligent technological civilisation will continue virtually unabated. He assures us we'll see continued exponential growth in the power and products of information technology. And, as he rightly says; "If all the AI systems decided to go on strike tomorrow, our civilization would be crippled."²

Kurzweil believes the next big leap for humanity will be reverse-engineering the human brain with a view to applying, as far is possible, the secrets of its incredible methods and processes once they have been decoded (like the human genome) and then simulated in even more powerful information technologies. If that project ever succeeded, it would greatly accelerate the capacity and applications of artificial intelligence, inaugurating an intelligence super-revolution.

So far, so good. But every silver lining has a cloud wrapped around it (excuse the weak pun).

¹ "My core thesis, which I call the law of accelerating returns (LOAR), is that fundamental measures of information technology follow predictable and exponential trajectories, belying the conventional wisdom that 'you can't predict the future'... The quintessential example of the law of accelerating returns is the perfectly smooth, doubly exponential growth of the price/performance of computation, which has held steady for 110 years..." Kurzweil, R. 2012. *How to Create a Mind*, p.249-50.

² *ibid*, p.158.

Here's a caveat worth considering as we contemplate the future of artificial intelligence (which we can no longer live without): humans should never submit to any forces of dehumanization. That's why talk of blurring the lines between these two kinds of intelligence, or of artificial intelligence becoming more dominant than human intelligence, needs to be treated with the utmost caution. There are critical decisions to be made over the next two decades regarding the right level of convergence, or interrelationship, between human and artificial intelligence. What balance between the two kinds of intelligence will enable us to maintain control of science and the progression of civilization? This is an issue which will shape society in profound ways – as well as our ultimate fate.

Whilst I greatly respect the insights and predictions we find liberally sprinkled throughout Kurzweil's prodigious work, I do think it's rather mischievous to conflate human and artificial intelligence as if they're only different degrees of the same phenomenon, whereas they're different phenomena altogether. Please allow me to explain.

Human intelligence emerges in real-time in Nature's four dimensional space-time continuum. Machine intelligence, by contrast, resides in systems and programs of code which, being non-organic, are stored in what I would call a-historical space. They are not part of Nature's systems. To suggest that these a-historical, artificial systems, whether robots or computer networks, will one day become conscious, and eligible for political rights as living entities in society, which is interconnected with Nature, is, I believe, poppycock. Why do I say that?

To be conscious, in the way that humans are conscious, you need self-consciousness. And to have self-consciousness, you first need a self. To have a self, in turn, you need autonomy and will-power exercised in real-time. The human brain is indivisibly connected to a self, housed within a complex neurological organism which is interconnected, in turn, to Nature. And to have that kind of complex autonomy, you need to exist in continuous time, or history, evolving in step with Nature's four dimensional space-time continuum. In short, to be a real, living, existential, evolving self, you need be part of history and evolving Nature. This can only happen in continuous space-time, the realm of change, not in the a-historical space of code programs. Ontologically speaking, the evolution in, and through, time, required for true autonomy as a being, *cannot* happen in a-historical space. Robots, that is, are not alive in the space-time continuum, just as they are not interconnected with Nature.

This ontological argument posits that a-historical entities like robots can never attain to the full autonomy required for self-conscious beings. Even learning systems, which are programmed to self-correct and accumulate new knowledge, do not possess a self, with its own autonomy. I would argue that human and artificial intelligence are not different degrees of the same phenomenon: they are different phenomena.

I would further suggest that we need to add a Cartesian test of autonomous self-awareness to the Turing Test³ for machine intelligence, which is that a computer would deserve to be called intelligent if it could deceive a human into believing that it was human. Descartes's "I think therefore I am" is a much better test. It indicates a being (Descartes) who knows what he knows, because he's part of history (including his personal history and education, as well as the history of philosophy he understands).

"I think, therefore I am" indicates a being who knows he has spontaneously and intuitively created an original idea and a new principle of philosophical truth.

While logical processes, in isolation, can certainly be replicated in artificial programs, there are three aspects of this great milestone in philosophical thought which cannot be replicated in code:

- Selfhood itself, with full autonomy and the exercise of will in, and through, time – the "I";
- Spontaneous, free-flowing thinking producing original out-of-the-box ideas – the "think"; and,
- Self-knowledge – the "I am".

Machine intelligence should not just fool humans into thinking they are speaking with another person, as suggested by the Turing Test. It should be required to pass the Cartesian Test and produce thinking about thinking to attain true self-knowledge. Until a machine can say "I think therefore I am" and know that it knows what it's saying, its intelligence will remain qualitatively inferior to human intelligence.

³ "The Turing test is a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. In the original illustrative example, a human judge engages in natural language conversations with a human and a machine designed to generate performance indistinguishable from that of a human being. All participants are separated from one another. If the judge cannot reliably tell the machine from the human, the machine is said to have passed the test." http://en.wikipedia.org/wiki/Turing_test

Given the qualitative differences between machine and human intelligence, here are three suggested core guiding principles, which could be used outside of the noble medical profession, for the humanitarian and social governance of the future of artificial intelligence, and new, related biotechnologies:

- Humans should not become technology. This would lead inevitably to objectification of our *being*.
- Humanity should not be abandoned to predominant control by artificial intelligence systems. This would lead to the depersonalization of human life, while constituting a high-risk abdication of social and moral responsibility.
- Humans should not attempt to alter the intrinsic nature of reality revealed to us by the laws and knowledge of science, nor contravene Nature's evolutionary processes, by creating in the laboratory cloned humans or any new kinds of manufactured creature, including through the augmentation of animal intelligence with artificial intelligence.

These three broad governance principles are about reverence for humanity, evolutionary Nature and science, in that order. They set reasonable boundaries for acceptable use of science to shape the future of society and Nature through a judicious use of empowering technologies. These principles are in keeping with our classic treasure chest of collective philosophical, religious and scientific knowledge gleaned throughout the history of civilization.

Underpinning these principles is a core assumption that there are ontological boundaries between humans, animals and robots which should be respected *ad infinitum*.

For example, each has a different kind of intelligence, a different form as well as different roles or purposes for existence. I'm convinced that most humanists would agree with me: we don't want to blur these boundaries and manufacture hybrid entities, whether human-animal beings, human-robot beings or animal-robot beings. I'm not talking about exceptions like use of medical technology to prolong the life, or improve the health, of a person with a faulty organ, medical condition or a disability. I am talking about trying to manufacture in the laboratory, for purposes not related to medicine, new hybrid beings intrinsically combining characteristics of humans, machines or animals. This would represent a new level of hubris, a moral failing of premeditated ontological blurring.

How we view humans, both now and in the future, who wish to merge permanently with technology and become cyborgs, with cybernetic implants, is going to be a major 21st century political question which will require a definite, rational and humane set of policies. It's an issue that will need to be continuously evaluated and wisely governed. For cyborgs, becoming a human-machine hybrid is an issue of the "right to bodily autonomy",⁴ whilst for humanists this quickly becomes a question of natural law and social equality. There has been much debate and discursive literature about this intriguing and important topic and it's not within the scope of this piece to cover all the complex theoretical ground which has already been argued. What I'm saying, however, is that an inflexion point will soon be reached when artificial intelligence becomes so powerful, and therefore influential, that social and political policies will need to be reviewed fairly urgently in order to determine the role we wish artificial intelligence to play in human life and society.

Transhumanism is open and honest about working towards a post-human future, or condition, daring openly to alter human nature and transcend its biological limits, through biological and neurological augmentation, while, at the same time, advocating the use of chemicals to enhance subjective states of happiness. The 1999 Transhumanist Declaration, for example, states: "Transhumanists advocate the moral right for those who so wish to use technology to extend their mental and physical capacities and to improve their control over their own lives. We seek personal growth beyond our current biological limitations".⁵

In addition, the Declaration makes clear that a key principle of this movement is the well-being of all sentience (whether in artificial intellects, humans, non-human animals or possible extra-terrestrial species). This point effectively replaces the individual human person (the "I think, therefore I am"), which has been the cornerstone of modern Western philosophy and culture, with a radical new yardstick of Transhuman sentience. Such a subtle shift of emphasis is actually one of the most profound paradigm shifts in the history of thought. And it completely ignores and bypasses the ontological boundaries between human, animal and robot: "I firmly believe that uplifted chimps will force us to admit that intelligent personhood, not humaneness, is the ticket to citizenship."⁶

⁴ Hughes, J. 2004. *Citizen Cyborg*. p.130.

⁵ Hughes, J. 2004. *Citizen Cyborg*, p.177

⁶ Hughes, J. 2004. *Citizen Cyborg*, p.94.

Once this triple distinction between the being, identity, intelligence and purpose of humans, animals and robots is thus eliminated in Transhumanism, we'll find ourselves on a slippery slope of ontological blurring that leads to some radical, and unexpected, envisaged outcomes, such as:

- The creation, in real laboratories, of mythological and fantasy creatures: “When I was reading Harry Potter to my kids they asked if there really were unicorns, elves and centaurs. I told them no....not yet. But in the coming decades and centuries we will create all the creatures that populate our mythologies.” (Hughes,J. 2004. *Citizen Cyborg*, p.92)
- The equipping of animals with human level intelligence : “We will give non-human animals human-level intelligence, and then have to decide if they should have rights in our society or not...Soon we will be able to genetically enhance primates to have human intellectual capabilities, a project dubbed ‘uplifting’ by David Brin.” (Hughes,J. 2004. *Citizen Cyborg*, p.92-4)
- Genetic modification of the human body for aesthetic or cosmetic reasons as well as genetic mixing of human and animal DNA;
- The granting of political rights to gorillas and chimps: “...bioethicist Peter Singer and an international group of activists have organized around the Great Ape Project (GAP). They propose that we should extend the boundaries of rights first and most extensively to great apes – gorillas, bonobos and chimpanzees – since we have strong evidence that they share our capacities for self-awareness.”⁷
- The merging and blurring of artificial and human intelligence: “We will incorporate computers into our bodies and brains and simulate human brains in computers.”⁸ “Humans already constitute spiritual machines. Moreover, we will merge with the tools we are creating so closely that the distinction between human and machine will blur until the difference disappears. That process is already well under way, even if most of the machines that extend us are not yet inside our bodies and brains.” (Kurzweil, R. 2012. *How to Create a Mind*, p.223-4)
“...we will merge with the intelligent technology we are creating.” (Kurzweil, R. 2012. *How to Create a Mind*,p.279)

⁷ Hughes,J. 2004. *Citizen Cyborg*,p. 93.

⁸ Hughes,J. 2004. *Citizen Cyborg*, p.77.

- Filling human bodies and brains with intelligent nanobots: “Intelligent nanobots in our bloodstream will keep our biological bodies healthy at the cellular and molecular levels. They will go into our brains noninvasively through the capillaries and interact with our biological neurons, directly extending our intelligence...”(Kurzweil, R. 2012. *How to Create a Mind*, p.279);
- The attainment of personhood by machines around 2030: “My objective prediction is that machines in the future will appear to be conscious...We will come to accept that they are conscious persons.” (Kurzweil, R. 2012. *How to Create a Mind*, p.207)
- The universal use of drugs to artificially manage human happiness and other states of mind;
- Sex with robots (sexbots) “... a robot will be able to offer the lightest of touches and the most gentle of caresses to its lover. And in return, the receptors in its artificial skin will provide the robots with sensuous feedback.” (Levy, D. 2006. *Robots Unlimited*, p.352)
- The colonization of the universe by artificial intelligence: “...waking up the universe, and then intelligently deciding its fate by infusing it with our human intelligence in its non-biological form, is our destiny.” (Kurzweil, R. 2012. *How to Create a Mind*, p.282)

The canned heaven of Transhumanism with its medicated humans and electronically fortified brains, and Dr. James Hughes's fantastical future Narnia society populated by a proliferation of transhuman creatures, complete with unicorns and elves from the laboratory, would violate evolution, natural law and, for the religious majority of human beings, divine law. Are we really so clever and so empowered that we want to disconnect our history from Nature, the sacred source of life, and put ourselves under the control of exponential forces of artificial intelligence and in the hands of genetic engineers? This Brave New World of hybrid creatures, clones, human-machine entities, sexbots, the domination of human intelligence by artificial intelligence systems and chemical manipulation of the human mind, will undermine human character, will and autonomy, turn the world into a gigantic laboratory and destroy our ecological relationship with Nature.

Those who may wish to play God by going beyond the ancient boundaries of Nature and human nature should please first prove to the world that they're both benevolent *and* omniscient.

Dr Noel Sharkey, Emeritus Professor of Artificial Intelligence and Robotics at the University of Sheffield, and Dr Amanda Sharkey, Senior Lecturer at the Department of Computer Science, wrote in the October/November edition of *The World Today*: “It is important to gain benefits from the advances in robotics while avoiding the dystopian consequence of a misplaced trust in robot capabilities. We as a society need to discuss the issues now rather than sleepwalk into an inhumane nightmare.”⁹

US writer, futurist, philosopher and transhumanist, Zoltan Istvan, gave a speech at the latest Artificial Intelligence and Singularity Conference in Oakland, California, in which he said: “The key, of course, is not to let artificial intelligence run wild and out of sight, but to already be cyborgs and part machines ourselves, so that we can plug right into it wherever it leads. Then no matter what happens, we are along for the ride. After all, we don't want to miss the Singularity.”

This is not moral guidance, it's abandonment of the species to forces not even the experts and enthusiasts fully understand. Istvan then went on to say: “In my philosophical novel *The Transhumanist Wager*, I put forth the idea that all humans desire to reach a state of perfect personal power— to be omnipotent in the universe. I call this a Will to Evolution.” This, my friends, smacks of megalomania – the desire to be omnipotent. Even Istvan, though, warns that artificial intelligence systems need to come with built-in, fool-proof “on/off switches”.

It would be foolish and counter-productive to oppose the progression of artificial intelligence because there are so many incredible benefits accruing to humanity through automated systems, robotics and digitization.

I actually believe that the greatest contribution of artificial intelligence to humanity will turn out to be: increasing the rationality of human beings forced to compete with AI in the workplace over the coming decades. That increase in rational behaviour would be very welcome, Mr.Robot!

However, any widespread Transhumanist abandonment of the ontological barriers between humans, animals and robots would almost certainly undermine, probably terminally, the essence of the planetary eco-system of Nature and Civilization which has evolved over many millennia.

⁹ “In the care of Nurse and Ann Droid” by Noel Sharkey and Amanda Sharkey, *The World Today*, Vol 70, Number 5, October/November 2014, p.22.

This ecosystem of humanity and Nature is still the only known form of successful social life in our whole wondrous cosmos. I vote to preserve it the way it is.

Acknowledgments

Hughes, J. 2004. *Citizen Cyborg*. Cambridge, MA: Westview Press.

Kurzweil, R. 2012. *How to Create a Mind*. New York: Penquin Books.

Levy, D. 2006. *Robots Unlimited*. Wellesley, Massachusetts: AK Peters, Ltd.

Sharkey, N and Sharkey, A. "In the care of Nurse and Ann Droid". The World Today, Vol 70, Number 5, October/November 2014.

http://en.wikipedia.org/wiki/Turing_test