Guest Editorial

ADAM A. FORD

Humanity at large is on the brink of understanding that our future will be wildly different from the past. In the coming decades we may witness the human condition transform in fundamental ways. We can see the effects of accelerating technology on our desks, in our pockets and all around us, providing transformative solutions to problems that have plagued us at least since the dawn of recorded history.

Will we foresee the potentials and the perils of advanced artificial intelligence before it materialises in full maturity? Or will we be blind-sided by technological flux? Will we react with a knee jerk when what is needed is a kit of well-considered tools to carefully navigate change?

This edition of *Issues* canvasses opinions on the future of artificial intelligence (AI) and humanity. The essays within it paint pictures of the future that are coloured with opportunity, peril, risk, reward, and many shades in between.

Ben Goertzel (p.4) introduces us to the singularity, "a hypothetical future point in time when the rate of progress in science and technology becomes so fast that the human mind can't possibly keep pace". This phenomenon is unprecedented. To help understand this, Goertzel points to a conceptual difference between normal, narrow AI and artificial general intelligence (AGI). He refers to AGI as "a system that has a lot of focus on generalisation and the ability to extend intelligence beyond one particular domain", and notes that humans do not have infinitely general intelligence.

Goertzel believes that the time is ripe to create AGI due to the convergence of a number of critical factors, such as advances in computer hardware performance, cognitive science and robotics. He predicts that "once a certain level of development is reached by any AGI research team, the rest of the world is going to wake up, [and] take notice," prompting governments to fund large AGI projects.

Kevin Korb and Ann Nicholson (p.9) begin by highlighting the idea that AI might be unbound by the limits of intelligence that humans have, and an "intelligence explosion" is possible once a certain threshold is reached; as British mathematician I.J. Good has expressed, "the intelligence of man would be left far behind".

Korb and Nicholson outline why AI may arise further into the future than what some optimists predict. There is much disagreement about when AI will achieve a high enough level of intelligence to achieve a singularity. Despite this Korb and Nicholson agree that there are potential dangers in creating intelligences similar to or greater than our own: "[The singularity's] arrival could be hugely detrimental to humanity if the first AIs built are not ethical," they write.

Hugo de Garis (p.13) poses a poignant question: "Should massively intelligent machines replace human beings as the dominant species in the next few decades?" He states that, in the near future, the artificial brain industry will be huge.

Naturally, questions will arise from having increasingly intelligent robots around: "Can the machines become smarter than humans? Is that a good thing?" From this concern de Garis predicts that a species dominance debate will arise with different ideologies clashing, resulting in an unavoidable war.

Wendel Wallach (p.20) stipulates that we don't currently understand enough about intelligence to know whether a singularity is possible. He thinks that "too much speculation, wishful thinking, fearful thinking and blind faith is posing as science. The most exciting innovations will be those whose impact we cannot anticipate."

Steve Omohundro (p.24) explains why creating smart AI could pose a threat based on the drives an AI may develop to achieve its goals. He questions the common assumption that problematic robots could be unplugged: a smart AI will try to block any attempts to unplug it, "and if you persist in trying to stop it, it will develop a subgoal of trying to stop you permanently". Furthermore, "if the robot can gain

access to its source code, it will want to improve its own algorithms," which will result in unpredictability.

Omohundro believes that the potential benefits of having a powerful AI that computes with value and meaning are *enormous*. So, to counteract unpredictability, we must build AIs with additional values beyond the goals they are designed for; that compute with meaning and take actions through rational deliberation. He suggests initially creating highly constrained AIs that act within very limited predetermined parameters. With the benefit of the intelligence of this constrained system, we can design the next generation of less constrained systems.

James Newton-Thomas (p.27) suggests that technological change will help us move further beyond our biological limitations. He says it increases our tool-making capacity and will enable us to further offload our intelligence into computing devices, which will help solve very difficult classes of problems.

Greg Adamson (p.31) discusses five barriers to socially beneficial technology: prohibition, intolerance, secrecy, greed and confusion. He suggests that these barriers are not top-of-mind for engineers, and he applauds those "advocates of the use of technology for social benefit rather than a source of non-productive profit".

Natasha Vita-More (p.35) explains the transhuman, which "marks the beginning of our evolution from human as we merge with machines". She says: "The transhuman is at a transitional stage of merging with technologies, resulting in a shedding a biological exclusivity". This intermediary stage precedes a technological singularity, or at least a time at which we can achieve independence from the material body. Vita-More explains the history of transhuman thought, and its different strands.

Meredith Doig (p.37) says that even though transhumanism may seem far-fetched, it should be approached with an open mind: "As a rationalist, I am an advocate of the human capacity to reason, particularly in the service of improving the human condition. Transhumanism does this." She observes that pursuing some of the goals of transhumanism may be like pursuing Utopia, which may irrevocably lead us away from what makes us human.

Randal Koene (p.41) introduces us to the idea that everything we experience and act upon – the very essence of who we are – can be reduced to

patterns and processes in our minds: "So, when we say that we want to extend or expand life, what we really mean is that we want to extend or expand that processing in your mind". Ultimately, the best survival mechanism, which also has the most headroom for diversity and richness of experience, is to extend and ultimately port the mind processes to more reliable and rich substrates.

In a second article, Vita-More (p.46) explores the singularity with a focus on life enhancement, allowing us to design our own experience and extend our life span significantly. "Unless human life can one day continue past its maximum biochemical process, humans can only augment, enhance, adopt and hybridise it," she says.

What excites you about the future? What frightens you? We are at a unique stage in history as scientific and technological progress accelerates dramatically. I programmed on an Apple IIc when I was a kid in the 1980s, being the only one on my street with a computer. I did not believe that computers would one day be a part of most homes, and I did not think they would migrate to our pockets. I have seen astounding technological growth, and I now find it much easier to swallow the idea that technology will continue to accelerate. Strap yourselves in for a wild ride.

An international intellectual and cultural movement is growing to support the use of science and technology to further the progress of AI and the ethics surrounding its use. Without giving the future the full attention it deserves, how can you know what sort of future you want?

If you think through the possibilities of AI, you can better appreciate the consequences. The Internet is an enormous reference to ideologies, organisations and notable individuals, as well as special references to the singularity in art and fiction. *Issues* and other publications represent an exciting lexicon of knowledge about this very important set of topics.

Do you want to help build a better future? Don't wait for permission; your energy is needed. I encourage you to continue to read and explore, form your own opinions and discover your personal potential to help in transforming the world of tomorrow.

Adam Ford is the Singularity Summit Australia and Humanity+ Summit Australia Coordinator. Find out more about the 2012 Singularity Summit at http://summit.singinst.org.au and the Humanity+ Summit at http://2012.humanityplus.org.au