Intelligent robots must uphold human rights

The common fear is that intelligent machines will turn against humans. But who will save the robots from each other, and from us, asks **Hutan Ashrafian**.

There is a strong possibility that in the not-too-distant future, artificial intelligences (AIs), perhaps in the form of robots, will become capable of sentient thought. Whatever form it takes, this dawning of machine consciousness is likely to have a substantial impact on human society.

Microsoft co-founder Bill Gates and physicist Stephen Hawking have in recent months warned of the dangers of intelligent robots becoming too powerful for humans to control. The ethical conundrum of intelligent machines and how they relate to humans has long been a theme of science fiction, and has been vividly portrayed in films such as 1982's *Blade Runner* and this year's *Ex Machina*.

Academic and fictional analyses of AIs tend to focus on human–robot interactions, asking questions such as: would robots make our lives easier? Would they be dangerous? And could they ever pose a threat to humankind?

These questions ignore one crucial point. We must consider interactions between intelligent robots themselves and the effect that these exchanges may have on their human creators. For example, if we were to allow sentient machines to commit injustices on one another — even if these 'crimes' did not have a direct impact on human welfare — this might reflect poorly on our own humanity. Such philosophical deliberations have paved the way for the concept of 'machine rights'.

Most discussions on robot development draw on the Three Laws of Robotics devised by science-fiction writer Isaac Asimov: robots may not injure humans (or through inaction allow them to come to harm); robots must obey human orders; and robots must protect their own existence. But these rules say nothing about how robots should treat each other. It would be unreasonable for a robot to uphold human rights and yet ignore the rights of another sentient thinking machine.

Animals that exhibit thinking behaviour are already afforded rights and protection, and civilized society shows contempt for animal fights that are set up for human entertainment. It follows that sentient machines that are potentially much more intelligent than animals should not be made to fight for entertainment.

Of course, military robots are already being deployed in conflicts. But outside legitimate warfare, forcing AIs and robots into conflict, or mistreating them, would be detrimental to humankind's moral, ethical and psychological well-being.

Intelligent robots remain science fiction, but it is not too early to take these issues seriously. In the United Kingdom, for example, the Engineering and Physical Sciences Research Council and the Arts and Humanities Research

> ## INTELLIGENT ROBOTS REMAIN SCIENCE FICTION, BUT IT IS NOT TOO EARLY TO TAKE THESE ISSUES SERIOUSLY.

Council have already introduced a set of principles for robot designers. These reinforce the position that robots are manufactured products, so that "humans, not robots, are responsible agents".

Scientists, philosophers, funders and policy-makers should go a stage further and consider robot–robot and AI–AI interactions (AIonAI). Together, they should develop a proposal for an international charter for AIs, equivalent to that of the United Nations' Universal Declaration of Human Rights. This could help to steer research and development into morally considerate robotic and AI engineering.

National and international technological policies should introduce AIonAI concepts into current programmes aimed at developing safe AIs. We must engage with educational activities and research, and continue to raise philosophical awareness. There could even be an annual AIonAI prize for the 'most altruistically designed AI'.

Social scientists and philosophers should be linked to cutting-edge robotics and computer research. Technological funders could support ethical studies on AIonAI concepts in addition to funding AI development. Medical funders such as the Wellcome Trust follow this model already: supporting research on both cutting-edge healthcare and medical ethics and history.

Current and future AI and robotic research communities need to have sustained exposure to the ideas of AIonAI. Conferences focused on AIonAI issues could be a hub of research, guidelines and policy statements. The next generation of robotic engineers and AI researchers can also be galvanized to adopt AIonAI principles through hybrid degree courses. For example, many people who hope to get into UK politics take a course in PPE (politics, philosophy and economics) — an equivalent course for students with ambitions in robotics and AI could be CEP (computer science, engineering and philosophy).

We should extend Asimov's Three Laws of Robotics to support work on AIonAI interaction. I suggest a fourth law: all robots endowed with comparable human reason and conscience should act towards one another in a spirit of brotherhood and sisterhood.

Do not underestimate the likelihood of artificial thinking machines. Humankind is arriving at the horizon of the birth of a new intelligent race. Whether or not this intelligence is 'artificial' does not detract from the issue that the new digital populace will deserve moral dignity and rights, and a new law to protect them. ∎

Hutan Ashrafian *is a lecturer and surgeon at Imperial College London, UK.*
e-mail: h.ashrafian@imperial.ac.uk