Woojin Lim

The Future of Smart Machines: Intelligence, Morality, and Rights

The Future of Smart Machines: Intelligence, Morality, and Rights

The 21st century has bred the unprecedented growth of cutting-edge science and groundbreaking inventions, especially in the fields of high technology and artificial intelligence (AI). To this day, researchers have built increasingly intelligent software and human-emulating programs. Evermore pressing are the questions of whether or not AI *can* and *should* (1) have human-like intelligence, and of what kind, (2) be programmed with a set of moral duties and obligations, and (3) be considered moral agents with legal rights.

A Ghost in the Shell: Can robots have human-like intelligence?

It is far-fetched to argue that robots are able to "think" in the manner of *cogito, ergo sum*. The human mind, in its disposition to "be aware of the world and their experiences, to think, and to feel; the faculty of consciousness and thought," is unique from that crafted from steel and silicon, a mechanistic physical realm governed not by the laws of *free* spirit (Oxford University Press, 2018b). Though robots inherently have not the higher level characteristics of self-awareness, sentience, and consciousness, it is indeed possible for artificial minds to seamlessly emulate the robust and multifaceted characteristics of the human mind.

Take a look at AlphaGo, the Baduk-playing DeepMind who defeated Lee Sedol in 2016, or OpenAI, a bot that learned the game *Dota 2* from scratch by self-play, enough to beat professional eSports gamers in 2017. Without a doubt, AlphaGo and OpenAI possess *intelligence*; computers can "acquire and apply knowledge and skills" oftentimes significantly better and more effectively than humans can (Oxford University Press, 2018a). The fact that computers are programmed, consigned to act with a Turing algorithm, does not preclude innovation and creativity. Algorithms trained on romance literature, like that of Jack Hopkins, have been used to write poems, and the annual Robot Art Competition have drawn teams from

across the world to submit a range of robot-created paintings with sophisticated brushworks. The art of creativity—learning from past examples, imitating and experimenting with new arrangements and variations, gathering feedback from the audience, and accordingly problem-solving—is indeed programmable.

Creating Morality: Is there a right answer to the Trolley Problem?

Popular media often envisions of a dystopian society wherein oppressive, intelligent machines take complete totalitarian control over the world of humans. *The Matrix*, for instance, tells of a future where highly-creative machines subdue the human population to no more than an energy source. Truth be told, scientists argue that such Sci-Fi tales could actualize. Elon Musk humanizes AI as an "immortal digital dictator," a threat that poses "a fundamental risk to the existence of human civilization" (D'Angelo, 2017).

Despite these diabolical Orwellian prophecies, philosophers and engineers have called for an ethical framework to imbue a sense of rational order and morality to AI systems. In order to do so, humans first need to define and translate its imagined orders (ie. morality) in a way that computers can process. This is a difficult ordeal. Even amongst humans, the conception of morality has been a field of perennial dispute. Moral relativists claim that ethical propositions do not reflect objective or universal moral truths, but rather make propositions relative to sociocultural and personal circumstances; moral egoists believe that moral agents should act by their own self-fulfilling interests; moral nihilists insist that nothing is intrinsically moral or immoral. Even amid moral objectivists—those who believe that universal principles for morality exist beyond its subjective interpretations—no consensus has been reached in outlining what exactly those principles consist of.

There is no correct answer to the Trolley Problem: to push the lever and kill one person, or to leave be and kill five. Teleologists will vouch for the pushing of the lever to maximize human utility; deontologists will reason against the idea, alleging that when an individual pushes the lever, she harbours greater moral culpability; all the while, others will have different nuanced interpretations. This is especially problematic since robots need objective metrics that can be clearly measured and optimized: how can a machine be taught to algorithmically maximize abstract human ideals like goodness, fairness, or beauty? A machine cannot be taught morality unless the humans creating its program devises a precise definition of what components morality consists of, a matter of constant contestation as different people have different views of the world around themselves.

Isaac Asimov's *Three Laws of Robotics* provides a glimpse of hope. It establishes universal decision rules to any potential ethical dilemmas the program might encounter:

> (1) A robot may not…allow a human being to come to harm;
>
> (2) A robot must obey orders given it by human beings [without conflict with law one]...;
>
> (3) A robot must protect its own existence [without conflicting with either law]…(Vest, 2001).

Though vague in wording, the premise protecting human life above all else ensures a safeguard against dystopian robotic tyranny. Cross-cultural and agreeable general principles (ie. death and suffering are bad) can be used as groundworks to maximize human welfare. Extended from Asimov's underlying ideas, Germany's BMVI has set out initial guidelines that regulate automated driving systems. Some of its key elements touch on the issues of safety, human dignity, freedom of choice, and data sovereignty. In the event of unavoidable accident, the car itself should only be concerned to minimize damage as much as possible and should not offset victims against one another (Federal Ministry of Transport and Digital Infrastructure, 2017).

Though ethical norms cannot be always clearly standardized, policymakers should make a conscious effort to universalize widely-accepted truths about the human condition, and work with provided data in light of cooperation and transparency. The public must have access to the algorithms engineers have used to quantify ethical values, and to the outputs the AI has produced. All automated decisions should be kept and reviewed to ensure ethical accountability. In interactions with humans, robots should be programmed not to impose its moral agenda, but be open-minded. On the whole, the platform of AI morality should remain flexible and open to public scrutiny, discourse, and reform.

The Charter of Robots and AI-doms: Can and should robots have moral rights?

In October 2016, the social humanoid Sophia became the very first robot to obtain citizenship. Some people contest the sheer absurdity of a robot being granted more rights than women, migrant workers, or displaced refugees who suffer from a lack of agency in that very country, and view Sophia's citizenship as degrading the concepts of rights for actual living, breathing humans. Others, in contrast, uphold the decision as a historic milestone for technological innovation, arguing that robots now have human-equivalent intelligence to ground rights. Whereas Sophia *can* (and did) become a citizen of Saudi Arabia (an ontological 'is' question), the more so relevant question is whether or not she *should* be granted citizenship (an axiological 'ought' question). In a larger context, the question becomes: can highly intelligent robots satisfy the normative conditions that would entitle them to moral and/or legal rights? If so, should they be granted exercise over those rights? Amid the various modalities to this equation, the most plausible answer, personally, is the view that robots cannot have moral rights exactly like that of humans due to underlying functional limitations, but a *sui generis* legal conception of rights and

obligations, a necessary corollary of strengthened human-robot relationships, should be concocted and respected.

At least as of right now, robots do not (and perhaps cannot) have the necessary properties that ground rights, to be fully considered legal and moral agents. Intelligence does not equate other capacities fundamental to morality such as self-awareness, which by its nature, necessitates the provision of minimal Enlightenment virtues, including the right not to be owned as property and the autonomy to life and liberty (Dvorsky, 2017; Ward, 2017). Though robots may not experience any suffering or feel pain from human violence, the act of striking a robot should be sanctioned by law, "not to avoid material damage in itself but rather to safeguard human feelings and uphold the interests of our society" (Cahen, 2016). There is something qualitatively and phenomenologically different about the way in which humans perceive and interact with AI; as robots increasingly grow and resemble humans in both complexity, intelligence and appearance, there is a notable difference in how humans engage with socially-interactive robots, compared to interactions with less-complex machines like toasters (Gunkel, 2017). As humans anthropomorphise robots, certain cognitive capabilities and emotions are projected onto them, which foments ethical obligations and especial treatment (McNeal, 2015).

In the same way corporations are recognized as legal bearers of rights and obligations despite not having traits like self-awareness, robots should also be provided a list of legal rights and obligations unique to its kind—though, not to the extent of human rights, as limited functional agents—as they continue to deepen in their relations with other human beings. Though it is unclear what exactly the future of robotics and AI holds, until then, humans should work collaboratively to develop a framework that firmly conceptualizes the limits and obligations of what exactly robots should or should not be able to accomplish, ethically and

legally, and establish a consensus-based grounding of *summum bonum* principles that humans believe all robots should aspire to uphold.

Bibliography

Cahen, M. (2016). The Rights of Robots. *Avocats*. murielle-cahen.com/publications/robot.asp

Cellan-Jones, R. (2014, December 2). Stephen Hawking warns artificial intelligence could end mankind. *BBC News: Technology*. Retrieved from bbc.com/news/technology-30290540

D'Angelo, M. (2017, July 17). Elon Musk: 'AI is a fundamental risk to the existence of human civilization'. *Tesla News*. Retrieved from teslarati.com/elon-musk-ai-fundamental-risk-existence-human-civilization/

Dvorsky, G. (2017, June 2). When Will Robots Deserve Human Rights? *The Gizmodo Review*. Retrieved from gizmodo.com/when-will-robots-deserve-human-rights-1794599063

Federal Ministry of Transport and Digital Infrastructure (BMVI). (2017, June). Ethics Commission: Automated and Connected Driving. *The Federal Republic of Germany*. Retrieved from bmvi.de/report-ethicscommission

Gunkel, D. J. (2017, October 17). Ethics and Information Technology. *Springer Netherlands*. Retrieved from doi.org/10.1007/s10676-017-9442-4

McNeal, G. S. (2015, April 10). MIT Researchers Discover Whether We Feel Empathy For Robots. *Forbes*. Retrieved from forbes.com/sites/gregorymcneal/2015/04/10/want-people-to-like-your-robot-name-it-frank-give-it-a-story/#7176954c48f9

Oxford University Press. (2018a, May 25). Definition: Intelligence. *English Oxford Living Dictionaries.* Retrieved from en.oxforddictionaries.com/definition/intelligence

Oxford University Press. (2018b, May 25). Definition: Mind. *English Oxford Living Dictionaries.* Retrieved from en.oxforddictionaries.com/definition/mind

Vest, F. (2001). Isaac Asimov's "Three Laws of Robotics". *Auburn University Division of University Computing*. Retrieved from auburn.edu/~vestmon/robotics.html

Ward, T. (2017, June 5). Should We Give AIs the Same Rights as Humans? *Futurism*. Retrieved from futurism.com/should-we-give-ais-the-same-rights-as-humans/