Agathon: A Journal of Ethics and Value in the Modern World

AI and IA: Utopia or Extinction?



Edited by

Ted Peters

AI and IA:
Utopia or Extinction?

Edited by Ted Peters

AI and IA:
Utopia or Extinction?

Edited by Ted Peters

ATF PRESS
Adelaide
2018

Text copyright © 2018 remains with the authors and for the collection with ATF Press. All rights reserved. Except for any fair dealing permitted under the Copyright Act, no part of the publication may be reproduced by any means without prior permission. Inquiries should be made in the first instance with the publisher.

Agathon: A Journal of Ethics and Value in the Modern World
Volume 5, 2018

Agathon refers to the Greek word used by Plato in the *Republic* to refer to 'the good beyond being', most notably deployed in recent times by Iris Murdoch and Emmanuel Levinas, both of whom use this touchstone to situate ethics at the heart of all of philosophy

We live in an evolving and increasingly complex world but our ethical concepts have frequently struggled to keep pace with the change. As a result, much of what passes for public debate at present remains in the grip of either deterministic or consequentialist thinking, both built on outdated assumptions and both representing attempts to address major issues in the absence of ethical concepts. We suffer, as Iris Murdoch lamented, a 'loss of concepts, the loss of a moral and political vocabulary'.

The interdisciplinary journal, *agathon*, seeks to bring together scholars from across the humanities, social sciences and sciences, including disciplines such as philosophy, theology, law and medicine, to engage with the ethical questions that now beset the modern world..

The journal is a home for considering questions such as how we deal with competing values in ethical discourse, how ethical theory finds expression in practice, what constitutes ethical character and how it is cultivated, and what excellence and wisdom look like for the ethical person or society in the twenty-first century.

Agathon is an international and interdisciplinary refereed journal published annually by ATF Press.

Chief Editor
Dr Paul Babie, University of Adelaide Law School Professor of the Theory and Law of Property

Editorial Board
- Professor Terence Lovat, Editor in Chief Bonhoeffer Legacy, The University of Newcastle, Australia & Hon Fellow University of Oxford, UK.
- Professor Robert Crotty, former Director, Ethics Centre of South Australia, Emeritus Professor of Religion and Education, University of South Australia, Adelaide.
- Managing Editor and Publisher Mr Hilary Regan, Publisher, ATF Press Publishing Group, PO Box 234 Brompton, SA 5007, Australia. Email: hdregan@atf.org.au

Subscription Rates
Local: Individual Aus $55 Institutions Aus $65 Overseas: Individuals US $60 Institutions US $65

Agathon is published by ATF Press an imprint of the ATF Press Publishing Group which is owned by ATF (Australia) Ltd (ABN 90 116 359 963) and is published once a year. ISSN 2201-3563

ISBN: 978-1-925679-20-5 soft
978-1-925679-21-2 hard
978-1-925679-22-9 epub
978-1-925679-23-6 pdf

Published by:

An imprint of the ATF Press Publishing
Group owned by ATF (Australia) Ltd.
PO Box 504
Hindmarsh, SA 5007
ABN 90 116 359 963
www.atfpress.com
Making a lasting impact

Table of Contents

Ted Peters, *Editorial* vii

Noreen Herzfeld, *The Enchantment of AI* 1

Ted Peters, *Artificial Intelligence, Transhumanism, and Frankenfear* 15

Neville Grant Rochow, *Somnambulating Towards AI Dystopia? The Future of Rights and Freedoms* 45

Alan Weissenbacher, *Artificial Intelligence and Intelligence Amplification: Salvation, Extinction, Faulty Assumptions, and Original Sin* 69

Martinez Hewlett, *AI in Recent Literature* 89

Dan Peterson, On the Faith of Droids 107

Rajesh Sampath, *From Heidegger on Technology to an Inclusive Puralistic Theology* 117

Mark Graves, *AI Reading Theology: Promises and Perils* 133

Levi Checketts, *Idle Hands and the Omega Point: Labor Automation and Catholic Social Teaching* 153

Human Utopia or Human Extinction?

Ted Peters

'As AI becomes invisibly ubiquitous, new ethical challenges emerge. The protection of human self-determination is one of the most relevant concerns and must be addressed urgently.'
– Mariarosaria Taddeo and Luciano Floridi[1]

The fast-moving frontier of Artificial Intelligence (AI) in league with Intelligence Amplification (IA) prompt anxiety in some quarters. When it comes to smart phones or autonomous gadgets for vacuuming our living room carpets, a brief moment of 'wow' is quickly superseded by 'oh, hum' Robotic intelligence, though initially fascinating, seems to prompt no worries about the short-range future. But, what about the long-range future? When we forecast the evolution of post-human superintelligence that will either bring a technological utopia or render *Homo sapiens* extinct, anxiety causes us to ask existential questions. In this issue of *Ethics,* we ask those questions.

Could the techie whizzes among us invent enough gadgets to relieve human beings from the drudgery of work? Can we look forward to a near future where robots and sex-bots and nanny-bots and chef-bots and mow-the-lawn-bots provide everything we previously had to strive for? Could we create our own utopia? Or, might there be a risk that the AI we create will, like the Sorcerer's Apprentice, get out of control? Might the superintelligent children of our technological procreation discard their parents as outdated junk? Will AI and IA bring us to a fork in the road where one route leads to utopia and the other to extinction?

---

1. Mariarosaria Taddeo and Luciano Floridi, 'How AI Can be a Force for Good', in *Science*, 361/6404 (24 August 2018): 751–752, at 751.

We pause here to ask: what is intelligence anyway? Could *Homo sapiens* with moderate intelligence give birth to a post-human species of more intelligent hybrids, cyborgs, or robots? If we humans are able to create cybernetic geniuses as our progeny, will we the creators attain the status of technological gods? Could *we* humans advance to the status of *Homo Deus?* Or, because that race of superintelligent and super-powerful creatures will survive us, should we designate *them* as 'gods'? Will we creatures have created a divine race to surpass us? Will the post-human discard us mere humans as obsolete trash?

These questions raise the specter of transhumanism. 'The field of artificial intelligence is deeply rooted in transhumanist visions for the future', avers Natasha Vita-More, Executive Director of Humanity + Incorporated.[2]

Like riding a jet propelled escalator, the transhumanists (abbreviated H+) among us are rocketing evolution toward an unprecedented new stage, namely, superintelligence. Whether through deep brain implants to amplify existing human intelligence (IA) or through creating robots with artificial intelligence (AI), a new posthuman species is scheduled to emerge. The event of posthuman emergence is dubbed, the *Singularity.* After crossing the Singularity threshold, superintelligence will take the reins of evolution and continue the advance to yet higher levels of intelligence until all of reality has been imbued with the cyber mind. Among other vicissitudes to be conquered is death, which will be surpassed by radical life extension or even cybernetic immortality. The transhumanists offer us a utopian vision of cosmic consciousness, even if we atavistic *Homo sapiens* may not be there to enjoy it.

These forecasts should give theologians pause, especially when it comes to anthropology and eschatology. Are the presuppositions about human nature adopted by transhumanists accurate? Is disembodied intelligence the same intelligence we in human bodies have come to know? Is the concept of whole brain emulation leading to cybernetic immortality realistic? Is the eschatological vision of the transhumanist compatible with, or contradictory to, the biblical promise of the Kingdom of God?

Some theologians, such as D Gareth Jones, warn of a competition between saviors. Who will save us? Science or Christ? 'The excesses

2. Natasha Vita-More, *Transhumanism: What is it?* (published by author, 2018). 64.

of transhumanism with its picture of a new world order, in which medicine will be devoted to conquering mortality, overcoming ageing, vanquishing neurodegenerative diseases and enabling people to live to 600 or so years of age as healthy and fulfilled individuals, rightly repel Christians....These extreme vistas represent a rerun of the science-as-saviour mentality.'[3] Or, to put the issue in First Commandment terms, does the H+ veneration of technology tempt us to idolatry, tempt us to revere the progress of science as salvific?

Agathon Articles

In the articles to follow here in this edition of *Ethics*, we will ask whether the disjunction between utopia and extinction provides the most realistic set of alternatives. According to Alan Weissenbacher, book review editor for *Theology and Science,* there is no reason to think that development of AI or IA will represent anything dramatic at all, especially once one assesses critically several of the assumptions that lead some futurists to overenthusiastic optimism. In his contribution 'Artificial Intelligence and Intelligence Amplification: Salvation, Extinction, Faulty Assumptions, and Original Sin', Weissenbacher analyses three questionable assumptions made by techno-optimists and transhumanists. First, the definition of intelligence we find in AI and IA technology is vague and confused. Wet brain human intelligence and dry machine intelligence are so different, they cannot rightly be compared. They are two different things, even if they functionally overlap at times.

Second, according to Weissenbacher, there appears to be an implicit assumption especially among transhumanists that higher intelligence equates to greater morality. This assumption, every theologian knows, is clearly false. Some of the brightest minds in history have devised and disguised the most horrendous evils. There is no relationship between enhanced intelligence and enhanced morality.

Third, closely related to this is the assumption that greater intelligence will bring with it the will to address societal problems. This is by no means the case. One could easily forecast that the generation of human beings who initiates the era of superintelligence could use

3. D Gareth Jones, 'A Christian Perspective on Human Enhancement', in *Science and Christian Belief,* 22/2 (2010): 14–16, at 14.

this newfound power for greedy or tyrannical purposes. Technological advance does not entail moral advance.

Unless, of course, you're Mark Graves. Graves, currently pursuing research at the University of Notre Dame in the US, is author of 'AI Reading Theology: Promises and Perils'. Graves is both a computer scientist and a theologian who raises the question: should we ask AI to read and interpret theology, especially the theology of St. Thomas Aquinas. Given the current computational ability to analyze and generate theological text, reasonably near future AI may plausibly read theology comparable to a first-year seminary student and have the capacity to learn more. Should theologians contribute to AI learning theology? Yes. Benefits include technologically enhanced theological interpretations and, note, improved AI morality. AI reasoning necessarily entails moral reasoning, he says. Really? I find this doubtful, frankly, because AI does not include selfhood; and it takes selfhood for a sense of moral responsibility and culpability. Be that as it may, Graves invitation for AI to enter the very heart of theological and ethical deliberation is both unique and exciting.

Asking a computer to reason morally would be to ask too much, according to Noreen Herzfeld, a hybrid theologian and computer science professor. Herzfeld fears that too much hype is leading us to unrealistic expectations. Despite nearly three quarters of a century of research, to date no electronic device yet invented exhibits intelligence. Computers compute, to be sure; but they cannot pass the Turing Test. They are not intelligent. Even though it's too early to predict what will happen in the future, Herzfeld bets that authentic machine intelligence remains science fiction. In her article, "The Enchantment of Artificial Intelligence," Herzfeld suggests that the H+ vision is extravagant, unrealistic. Even if Weissenbacher and Herzfeld are correct about long range forecasts, however, AI developments will have implications in the near future for human work and human dignity.

Despite our enchantment with AI, IA, and the H+ vision, all this might be considered dystopian rather than utopian. According to Australian barrister Neville Rochow, such dystopias are familiar to us. But, rather than focus on the technology associated with Amazon, Apple, Facebook, and Google plus the horizon of new AI advances, Rochow worries about the business side of technology. In his contribution, 'Somnambulating Towards AI Dystopia? The Future of Rights and Freedoms', Rochow says history teaches that supranational cor-

porations cannot be trusted with unchecked power. Corporations do not respect human rights. No one can predict the effect AI will have on human rights. We should take our rights seriously before the biotechnological lottery begins. Prompt international action is required to regulate AI developers and corporations.

'Faith creates the self in human beings.' Following Paul Tillich, Dan Peterson recognizes the decisive role played by faith in human personhood and then applies it to Strong AI. If AI is ever to equal the human level of intelligence, it will need to acquire both personhood and faith. We can anticipate what this looks like if we watch fictional robots on the big screen. Peterson offers a delightful exegesis of R2-D2 in *Star Wars: A New Hope* and K2-SO in *Rogue One: A Star Wars Story*. Peterson shows how faith as a capacity would be indispensible if Strong AI is ever to attain *imago Hominis*. In this essay, Peterson provides an excellent example of how a theologian can contribute to the creative mutual interaction between science and theology.

Shakespeare exclaims rhetorically: 'What a piece of work is man! How noble in reason! how infinite in faculties! in form and moving, how express and admirable! in action how like an angel! in apprehension, how like a god!' Now, are we talking about the humanity we know? Or, are we talking about the post-human that artificial intelligence and intelligence amplification could bring into existence? Will this amount to what is truly human, or what is post-human?

Martinez Hewlett, evolutionary biologist and fiction author, poses these questions in his article, 'What a Piece of Work is Man: On Being and Becoming Human in Science Fiction'. Hewlett interrogates science fiction, because there the question of what it means to be a human person is brought into sharp focus. Hewlett selects three authors whose work has influenced the genre and has been the inspiration for film adaptations: Isaac Asimov (*I, Robot*), Philip K Dick (*Blade Runner*), and Richard K. Morgan (*Altered Carbon*). All three challenge us by pushing the limits of robotics, androids, and mind uploading. They scenario visions of what the future might be in the wake of technological advances in artificial intelligence. Who shall we become?

In his article, 'Idle Hands and the Omega Point: Labor Automation and Roman Catholic Social Teaching', theologian and ethicist Levi Checketts assesses the impact of AI on work. Will AI replace work? He examines three polar theologies of work present in Roman

Catholic thought: labor as drudgery, labor as dignified calling, and labor as an obstacle. These three theologies inform an eschatological vision of labor—Christians are called to work to build the kingdom of God, not merely to acquire material possessions. What might be the implications for the distribution of resources and the problem of idleness in a post-labor world. All of this suggests Catholics must promote a future of dignified labor dedicated to love of neighbor and God.

In my own article, 'Artificial Intelligence, Transhumanism, and Frankenfear', I argue that even if AI (beyond rapid computation and cute gadgets) never comes to pass, theologians must speculate with transhumanist visionaries about the prospect of superintelligence, about the prophesied extinction of *Homo sapiens,* and about the survival of a post-human species. While advances in AI technology that benefit humanity should be celebrated, extravagant utopian promises should be met with a healthy dose of Frankenfear, that is, caution. AI and IA advances in the near future may very well make life on Earth for human beings and even the ecosphere better; yet the more extravagant utopian promises for the long-range future could risk large scale destruction.

Intelligence versus Love

Utopian promises beyond the Singularity threshold are unrealistic on two counts. The first is anthropological. To surmise that this generation of human beings could create a superior intelligence—where the creature is more intelligent than the creator—defies the logic of plausibility. Yes, we human beings have created power tools that are stronger than we who designed and manufactured them. This is a positive precedent. Yet, intelligence seems to require more than mere design. Because intelligence as we experience it in ourselves and certain animals requires both the capacity for insight and for decision-making, it's difficult to imagine that superintelligence could be the product of a design. It's difficult to imagine that what is designed would be more intelligent than the designer.

From the perspective of the Christian theologian, intelligence is a curious category. Why does it rise to the top of the values list for transhumanists? Why is it the trump card which defines the desirable future? For the follower of Jesus, love belongs on top of the values list,

not intelligence. To the children--not to the technological whizzes--belongs the kingdom of God, said Jesus. 'Let the little children come to me; do not stop them; for it is to such as these that the kingdom of God belongs' (Mk 10:14). With the logic of intelligence presupposed by H+, could H+ apply their technology to the capacity to love? Could we procreate a future generation of creatures more dedicated than we to love, compassion, healing, and peace? As laudable as increased intelligence might be, it should not sit atop our list of values.

The second unrealistic assumption made by H+ has to do with eschatology. Transhumanists assume that evolution is progressive. Plus they assume that increased intelligence leads to increased moral resolve. A more intelligent superspecies, H+ assumes, will have the moral capacity to heal all psychological, sociological, and ecological ills so that all consciousness can enjoy a higher level tranquility and even harmony. Yet, I ask: can a leopard change its spots? Can a sinful human race beget progeny healed of their sin without an act of divine grace? To assume that an increase in intelligence leads automatically to an increase in moral will is fallacious. It defies history, experience, and theological insight.

The ultimate transformation Christians look forward to is eschatological. It will be a gift of divine grace. It would behoove the present generation to view AI advance as penultimate, not ultimate. A healthy dose of caution is warranted.

Conclusion

In treating our theme, 'AI and IA: Human Utopia or Human Extinction?' this edition of *Ethics* engages in public theology. *Public theology*, I believe, should be *conceived in the church, reflected on critically in the academy, and addressed to the world for the sake of the world.*[4] Public theology has five tasks: pastoral, apologetic, scientific, political, and prophetic tasks. The writers of theme articles in this volume are taking up the task of engaging science and technology from a theological perspective.

4. Ted Peters, 'Public Theology: Its Pastoral, Apologetic, Scientific, Political, and Prophetic Tasks', in *International Journal of Public Theology,* 12/2 (2018): 153–177; https://brill.com/abstract/journals/ijpt/12/1/ijpt.12.issue-1.xml.

Readers should take away two significant points. First, God created the human species to be creative; and the future of science and technology represents the *imago Dei* at work within history. Second, because scientific and technological advance occur within history, we should avoid asking for more than it can deliver. No scientific insight or technological innovation can accomplish what God's promise of new creation can deliver. The wisdom of Reinhold Niebuhr abides today as well as it did three quarters of a century ago. 'The Christian hope of consummation of life and history is less absurd than alternate doctrines which seek to comprehend and to effect the completion of life by some power or capacity inherent in man and his history.'[5] When comparing the transhumanist promise with the biblical promise, the latter is less preposterous.

5. Reinhold Niebuhr, *The Nature and Destiny of Man,* Gifford Lectures, 2 Volumes (New York, Scripbners, 1941) 2:298.

The Enchantment of Artificial Intelligence

Noreen Herzfeld

'Against utopianism the Christian faith insists that the final consummation of history lies beyond the conditions of the temporal process. Against other-worldliness it asserts that the consummation fulfills rather than negates the historical process.'
– Reinhold Niebuhr[1]

Abstract. Does the advance of AI technology promise utopia? Or does it pose an existential risk to human life on Earth as we know it? Fears of a super-intelligent robot apocalypse may be premature, yet we dare not hope for more benefits than AI technology can actually deliver. AI will disrupt (1) the economy; (2) politics; and (3) our private lives. Mindless programs will overtake many of our currently mindless jobs. Data mining will give us new correlations on which we will base an increasing number of decisions, some of them leading to false premises and unjust actions. Social media and robotic sex partners will take the place of authentic face to face relationships. But AI will not tell us how to respond to these changes. Only we humans can do that.

Key Terms. Artificial Intelligence, robot, Singularity, Reinhold Niebuhr, Ray Kurzweil, Elon Musk

Bio. Noreen L Herzfeld is Nicholas and Bernice Reuter Professor of Science and Religion at the College of St. Benedict and St. John's University in Collegeville, Minnesota. She is author of *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis: Fortress Press, 2002).

1. Reinhold Niebuhr, *The Nature and Destiny of Man,* Gifford Lectures, 2 Volumes (New York, Scribners, 1941). 2:291.

In July 2017 at a meeting of the National Governor's Association, Tesla founder and CEO Elon Musk issued the following warning regarding the future of artificial intelligence: 'AI is a fundamental existential risk for human civilization, and I don't think people fully appreciate that.' Claiming access to cutting-edge AI technology, Musk called for proactive government regulation, noting that while such regulation is generally 'irksome, . . . by the time we are reactive in AI regulation, it's too late . . . I think people should be really concerned about it', Musk said: 'I keep sounding the alarm bell.'[2]

Musk is not alone. Several years ago physicist Stephen Hawking told the BBC: 'The development of full artificial intelligence could spell the end of the human race.'[3] According to Hawking, AI could 'take off on its own, and re-design itself at an ever-increasing rate . . . Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded.'[4]

This concern has been a staple of science fiction for decades (see *The Terminator* or *2001*). However, those with a more intimate knowledge of AI disagree. As MIT computer scientist Rodney Brooks has wryly pointed out, Musk and Hawking "don't work in AI themselves. For those who do work in AI, we know how hard it is to get anything to actually work through product level."[5] Virtual reality pioneer and Microsoft resident guru Jaron Lanier says anyone with experience of modern software should know not to worry about our future robotic overlords. 'Just as some newborn race of superintelligent robots is about to consume all humanity, our dear old species will likely be saved by a Windows crash. The poor robots will linger pathetically, *begging us to reboot them*, even though they'll know it would do no good.'[6]

2. https://www.c-span.org/video/?431119-6/elon-musk-addresses-nga
3. https://www.bbc.co.uk/news/technology-30290540
4. https://www.bbc.co.uk/news/technology-30290540
5. Connie Loizos, 'This famous roboticist doesn't think Elon Musk understands AI', https://techcrunch.com/2017/07/19/this-famous-roboticist-doesnt-think-elon-musk-understands-ai/, accessed 9/14/18.
6. Jaron Lanier, 'One-half of a Manifesto', in *Wired*, December 1, 2000. https://www.wired.com/2000/12/lanier-2/

Who is right? Does AI pose an existential risk to humankind?[7] Not for the reasons Hawking and Musk imagine. We are unlikely to have intelligent computers that think in ways we humans think, ways as versatile as the human brain or even better, for many, many years, if ever. However, that doesn't mean we are out of the woods. 'AI' programs that do one thing and do that thing very well (think Deep Blue) are progressing by leaps and bounds and stand to undermine, or at least drastically change, our economy and our politics. In fact, they are already doing so, as seen in the 2016 American election or in predictions, such as a recent one by the McKinsey Global Institute, that in ten years up to thirty per cent of our current jobs will be altered or made obsolete by AI.[8]

Such dislocation in the workplace, coupled with the potential for social dislocation from applications as disparate as social media and sexbots challenges our perception of who we are as human beings and what we are worth. Nor does AI need to be totally successful to effect this challenge. While the dire predictions of Musk and Hawking are unlikely to unfold, the simple idea of AI enchants us, obscuring its true risks and pushing many toward a kind of magical thinking that blinds us to who and what we are, and who we, as Christians, are truly called to serve.

Why the Singularity (Probably) Won't Happen

Early AI researchers believed a machine that could solve calculus problems or play a credible game of chess would be intelligent. Today we carry such machines—our smartphones or calculators—around in our pockets every day but do not consider them particularly intelligent. Computer scientists quickly learned that many problems we consider difficult turn out to be relatively easy to program while the skills of a toddler such as emotion recognition, or traversing a crowded room, are much more difficult. However, recent advances in machine learning and sensory interpretation are starting to conquer these areas as well.

7. An existential risk is a risk of an event that would either annihilate human life on Earth or permanently and drastically curtail its potential.
8. https://www.mckinsey.com/featured-insights/future-of-organizations-and-work/Jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages