Conference Report

AI - Ethical and Religious Perspectives

Will advances in AI serve to enhance or diminish our moral and spiritual selves?
Will these advances serve to create better or worse societies?

AI & the Future of Humanity Series
Science & Human Dimension Project
Jesus College Cambridge

16-17 May 2019

Contents

Conference Rapporteur: Tom Chatfield
Editor: Jonathan Cornwell

We thank the Master and Fellows of Jesus College Cambridge and Templeton World Charity Foundation (TWCF) for their support of this project.

Jesus College Cambridge

SHDP

Science & Human Dimension Project
Jesus College, Cambridge

Conference 1

Who's afraid of the Super-Machine? AI in Sci-Fi Literature and Film

15-16 March 2018 Jesus College, Cambridge

Conference Overview

The term "singularity" was introduced by the science fiction writer Vernor Vinge in a 1983; it was picked up by Ray Kurzweil in his popular 2005 book The Singularity is Near. At many stages we find fiction in all its forms driving ideas in AI and vice versa. Crucially, we find the relationship between AI developments and our hopes, fears and ambitions, worked out imaginatively through a variety of media. Hence film and literary fictions have been a forum for the drama of ideas that circulate around AI and its future, not least its moral dimension. What can we learn about ourselves in relation to AI by exploring these narratives? There are also powerful religious themes in the history of SF machine intelligence, such as achievement of immortality, notions of Omega point futures, transhumanism, and the prospect of androids outstripping humans in virtue. SF authors and film makers, researchers in literature, philosophy and the humanities addressed these questions with AI experts. A report based on the conference and a short film, AI in Science Fiction Film & Literature, are available on our website.

Conference 2

The Singularity Summit: Imagination, Memory, Consciousness, Agency, Values

26-27 September 2018 Jesus College, Cambridge

Conference Overview

Numerous research projects around the world are attempting to simulate human "intelligence" based in part on neurophysiological theories of memory and imagination. Although considerable work has been done in this area since the early 1990s, AI is currently experiencing a quantum shift, one that requires an in-depth review of the primary human faculties as well as the moral dimension of human existence. While these research and development programs would benefit greatly from dialogue with philosophy of mind, aesthetics, literary and cultural studies, and philosophical theology, there is a lack of dialogue between scholars in these areas and the AI communities. This conference provided a much-needed opportunity for interdisciplinary discussion between these groups who do not often find themselves around the same table. An important segment of the conference involved standing back to review the future of AI in the historical context of how the digital age has already affected society and individuals. A conference report and filmed interviews with speakers are available on our website.

Conference 3

AI - Ethical and Religious Perspectives: Will advances in machine intelligence serve to enhance or diminish our moral and spiritual selves? Will these advances serve to create better or worse societies?

16-17 May 2019 Jesus College, Cambridge

Conference Overview

In this conference we ask what impact future AI is likely to have on notions of the soul, religious faith, religious practice, and the virtues. Does AI pose a threat, or encouragement, to religious belief and practice, and will it create better or worse societies? In turn, we ask how religion might guide and inform attitudes towards, and relationships with, future intelligent machines. Finally, can religious perspectives influence and shape the course of AI research and development?

SHDP Science & Human Dimension Project

Science and Human Dimension Project is a public understanding of science, technology, engineering, maths and medicine (STEMM) program, based at Jesus College, Cambridge and founded in 1990. Through conferences and publishing, SHDP brings together the scientific research community with experts from industry, government and the media to deepen and broaden the appreciation of new ideas and discoveries and to ask searching questions about their impact on humanity.

SHDP addresses important ethical questions, such as the controversy over human embryonic stem cell research, superintelligent machines and artificial intelligence (AI). At times we are intent on tackling subjects illustrative of knowledge purely for its own sake.

SHDP helps university research, companies and think thanks bring their ideas to a wide audience of experts, the media and general public through carefully organised conferences with hand-picked participants. In 2017 we collaborated with DeepMind to deliver a conference on their AI research into memory and imagination.

AI & the Future of Humanity Project - from August 2017 to July 2019 we ran a two-year project of three conferences and related outreach on AI and the Future of Humanity, funded by Templeton World Charity Foundation (TWCF).

SHDP and outreach - we have a strong track record in achieving outreach from the university and the laboratory to the media and a wider non-specialist public through journalism, film and books. SHDP directors have produced conferences and reports on a wide range of vital issues including: Consciousness and Human Identity, AI, Cyber Security, Food Security, Inequality, Infrastructure, the Financial Crisis, Big Data, Blockchain and Bitcoin, the Future of Research-based Universities, the UK North-South Divide, Ageing, and the Future of Work. SHDP conference proceedings and books have been published by OUP, Bloomsbury, Penguin and Profile.

Science and Human Dimension Project - the project is run by John Cornwell (Director) and Jonathan Cornwell (Executive Director). Advisors on the *AI & the Future of Humanity* project include Dr Andrew Davison (Starbridge Lecturer, University of Cambridge), Dr Tim Jenkins (Anthropologist, Fellow, Jesus College, Cambridge), Rev'd Dr Paul Dominiak (Theologian, Fellow, Jesus College, Cambridge), and Dr Tudor Jenkins (technologist and Artificial Intelligence researcher). We also thank the following for their advice and help: Elisabeth Schimpfossl, Beth Singler, Andrew Briggs, Colin Ramsay, DragonLight Films, Steve Torrance, Keith Mansfield, Sumit Paul-Choudhury, Sam Thorp, Michael Harte, Kathleen Richardson, Tom Chatfield, Michael McGhee, Simone Schnall, Ezra Sullivan, Nikki Williams, Ian White, Ron Chrisley, Murray Shanahan, Adrian Weller, Richard Watson, Beatrix Lohn, Julian Huppert, Sarah Steele, Richard Anthony, Rob Shephard, Mark Cresswell, the Jesus College Development Office, Jesus College conference team, Helen Harris and Kelly Quigley-Hicks. We thank the Master and Fellows of Jesus College Cambridge, and Templeton World Charity Foundation (TWCF) for their support of this project.

SHDP Science & Human Dimension Project
Jesus College Cambridge

Tel: +44 (0)7768 220188 E: j.cornwell@jesus.cam.ac.uk www.science-human.org @ScienceHumanDP

Science & Human Dimension Project

AI - Ethical and Religious Perspectives

Will advances in AI serve to enhance or diminish our moral and spiritual selves?

Will these advances serve to create better or worse societies?

16-17 May 2019 Jesus College Cambridge

Agenda Day 1: Thursday 16 May 2019

11.00-11.25 Registration - Bawden Room, West Court, Jesus College

Refreshments served. Please move to Frankopan Hall by 11.25

11.30-11.40 Welcome and Introduction - Frankopan Hall, West Court

John Cornwell *Director, Science & Human Dimension Project, Jesus College, Cambridge*

11.45-12.55 Session 1

11.45-12.20 **Artificial Intelligence, Materiality and the Soul**

Revd Dr Andrew Davison *Starbridge Lecturer in Theology and Natural Sciences, Faculty of Divinity; Fellow in Theology, Corpus Christi, University of Cambridge*

12.20-12.55 **From Eden to AI: Jewish perspectives on defining 'life' and 'humanity'**

Rabbi Dr Raphael Zarum *Dean of the London School of Jewish Studies*

12.55-13.50 Lunch - Dining Room, West Court

13.55-15.05 Session 2

13.55-14.30 **Artificial Intelligence and Biotech: A view from the Catholic Church**

Fr Ezra Sullivan O.P. *Faculty of Theology, Angelicum, Pontifical University of St Thomas Aquinas, Rome*

14.30-15.05 **Blessed by the Algorithm: AI, Agency, and Religion**

Dr Beth Singler *Junior Research Fellow in Artificial Intelligence, Homerton College, University of Cambridge; Associate Fellow, Leverhulme Centre for the Future of Intelligence*

5 minute break

15.15-16.25 Session 3

15.15-15.50 **An Internet of Curses; Weber's 'vocation' in the world of AI**

Dr Fenella Cannell *Reader in Social Anthropology, LSE*

15.50-16.25 **Machines, Apes, Extra-Terrestrials and Bodhisattvas**

Dr Michael McGhee *Hon Senior Fellow, Department of Philosophy, Liverpool University*

16.25-16.55 Tea - Bawden Room

16.55-18.05 Session 4

16.55-17.30 **Situating Themes in Artificial Intelligence Discourse in Islamic Theology and Philosophy**

Dr Yaqub Chaudhary *Research Fellow in Religion and Science, Cambridge Muslim College*

17.30-18.05 **AI - Perspectives from Daoism, Shinto, Confucius**

James Kingston *Research Manager, CogX; Author, AI Ethics Primer*

19.00-19.30 Drinks - Prioress's Room, Cloister Court

19.30 Dinner - Upper Hall

AI - Ethical and Religious Perspectives

Agenda Day 2: Friday 17 May 2019

08.40-09.15 Registration - Bawden Room, West Court, Jesus College, Cambridge
Refreshments served. Please move to Frankopan Hall by 09.10

09.15-10.25 Session 5
09.15-09.50 **What is AI and what are the implications of advances in AI for religion?**
Neil Lawrence *Professor of Machine Learning, University of Sheffield; Director, Amazon Research, Cambridge*

09.50-10.25 **Apocalyptic Fictions: Mary Shelley and the Revelation of the Singularity**
Eileen Hunt Botting *Professor of Political Science, University of Notre Dame*

10.25-10.45 Break - Bawden Room

10.50-12.35 Session 6
10.50-11.25 **Differentiating the human person from any mechanical or technological creation**
Revd Dr Malcolm Brown *Director of Mission and Public Affairs, Church of England*

11.25-12.00 **Ethical Implications of Simulated Personhood**
John Wyatt *Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology, University College London; Senior Researcher, Faraday Institute, Cambridge*

12.00-12.35 **How can Christian, pantheist and animist theologies or philosophies respond to the posthumanist challenge to humanism?**
Dr Gorazd Andrejč *The Woolf Institute and St Edmund's College, University of Cambridge*

12.35-13.20 Lunch - Dining Room, West Court

13.25-14.35 Session 7
13.25-14.00 **From Moral Intuitions to Moral Machines: How Rational is Ethical Decision Making?**
Dr Simone Schnall *Reader in Experimental Social Psychology; Director of Studies in Psychological and Behavioural Sciences, Jesus College, University of Cambridge*

14.00-14.35 **AI & Ethics 2.0**
Dr Daniel De Haan *Postdoctoral Fellow, Ian Ramsey Centre for Science and Religion, University of Oxford*

5 minute break

14.40-15.50 Session 8
14.40-15.15 **Creation and Responsibility: The Paradox of Ethical Artificial Intelligence**
Dr Ron Chrisley *Director, Centre for Cognitive Science, University of Sussex; Visiting Scholar, Stanford Institute for Human-Centred Artificial Intelligence, Stanford University*

15.15-15.50 **AI messianism and our grandchildren's grandchildren**
Steve Torrance *Visiting Senior Research Fellow, COGS, University of Sussex; Professor Emeritus of Cognitive Science, Middlesex University*

15.50 Concluding Remarks
John Cornwell *Director, Science & Human Dimension Project*

16.00 Conference Close and Tea - Bawden Room

Please note that the conference will be filmed and recorded. The conference rapporteur is Tom Chatfield. #SHDP
We thank the Master and Fellows of Jesus College, Cambridge and Templeton World Charity Foundation (TWCF) for their support of this project.

Jesus College Cambridge

SHDP

Science & Human Dimension Project
Jesus College Cambridge

AI – Ethical & Religious Perspectives
Conference Executive Summary

John Cornwell, Director of the Science & Human Dimension Project, introduced the conference by noting that Norbert Weiner, a founder of machine learning, prophetically addressed in his 1964 book *God & Golem, Inc.* the twin concerns of: humans playing at God; and the rise of machine learning leading people to see every aspect of life as a kind of game, and thus reducible to something mechanical.

Revd Dr Andrew Davison, Starbridge Lecturer in Theology and Natural Sciences in the Faculty of Divinity of the University of Cambridge, suggested that neither dualism nor monism are the most useful ways of understanding the relationship between mind and materiality—and advocated, rather, a "strong emergentist perspective" in which a complex system can be both wholly material and have radically new powers and properties to those of its constituent parts. Such a perspective can take the possibility of Artificial General Intelligence in its stride.

Rabbi Dr Raphael Zarum, Dean of the London School of Jewish Studies, drew on the Torah's account of the human species as an "artificial" version of God's Image to yield the analogy of God creating humankind as the first AI—and this analogy in turn offering guidance around the ethical development and use of artificial intelligences, which demands a recognition of the fundamentally ethical, embodied and mortal nature of humanity.

Fr Ezra Sullivan OP, Professor of Philosophy and Theology, University of St Thomas Aquinas, Rome, described the strong interest of Pope Francis in engaging with AI, and the Catholic Church's beliefs around human genome editing. Drawing on the scriptural distinction between generation and manufacture, he argued that the process of human generation should remain a natural one—and that, although everything we do may be natural in one sense, it is not truly humane unless it respects universal dignity and rights.

Dr Beth Singler, Junior Research Fellow, Homerton College, and Leverhulme Centre for the Future of Intelligence, University of Cambridge, explored the entanglements of AI and religion through the migration of pseudo-religious descriptions of technology on social media from the realm of humour and parody into more mainstream and serious usage. There are such things as AI-inspired new religious movements, she noted—and a mixture of mounting tradition, the charisma of AI-boosting commentators, and the apparent reasonableness of AI's triumph marks them out as of significant influence.

Dr. F. Cannell, Social Anthropology, LSE, discussed the importance of acknowledging science and technology as socially and economically embedded. With reference to Shoshana Zuboff's critique of "surveillance capitalism", she

critiqued technological determinism alongside tech monopolies self-serving sense of their own creations as somehow inevitable—and advanced the suggestion that we need to think critically and decisively about different possible future directions for tech, alongside the extent to which many current social media systems fuel grievance and trauma.

Dr Michael McGhee, Hon Senior Fellow, Philosophy, Liverpool University, reflected on the possibility of areas of common ground between religious traditions, particularly in accounts of spiritual practices that help develop virtues and overcome vices. He suggested that the most pressing danger to human values comes from human beings—and that the greatest danger of AI may be its compromising of our capacity for virtue through lack of use.

Dr Yaqub Chaudhary, Research Fellow in Science & Religion, Cambridge Muslim College, discussed how contemporary accounts of AI can be understood and made consistent with Islamic theology, noting that Islamic teaching presents a highly favourable outlook towards technology that would welcome the managed and thoughtful deployment of AI. Focusing on the fact that a machine's environment consists of its data inputs, and is thus a purely mathematical construction, he suggested that many problems in the emerging field of AI ethics can be advanced by understanding the virtual world of AI itself—and how this is then mapped onto, and alters, the human sense of self.

James Kingston, Research Manager, CogX, suggested a definitional problem in AI ethics thanks to the broadness of the values it ought to encompass—and the degree to which industry is often anxious to appear legitimate in the world's eyes. Drawing on East Asian cultural and religious perspectives, he highlighted the diversity of attitudes to AI across the world, and the contrast of Shinto, Confucian and Daoist attitudes to those of western Christianity, and its emphasis on individualism and unitary single minds.

Prof Neil Lawrence, Sheffield University; and Head of AI Research, Amazon, contrasted the immensely low bandwidth of communications between humans with the immensely high bandwidth of communications between machines. Humans have huge processing capacity inside their minds, but almost all of this is spent on modelling the actions and intentions of other humans, making contextual understanding vital for the collective intelligence of humanity as a social species. This "information isolation" is what makes us human, and it profoundly differs from our creations' mediation and handling of data.

Professor Eileen Hunt Botting, Notre Dame, discussed Mary Shelley's novels *Frankenstein* (1818) and *The Last Man* (1826) and how their conceptualization of apocalypse relates to current strands of thinking around AI. By rejecting Romantic naïveté and pessimism, Shelley offers a remarkably resilient vision of human culture and civilization in the face of disaster, and also suggests via the Creature at the heart of Frankenstein a model for AI as the embodied, suffering offspring of its

creators, complete with the ethical responsibilities and complexities this quasi-parental relationship entails.

Revd Dr Malcolm Brown, Director of Mission and Public Affairs, Church of England, used the Swedish TV series Real Humans to illustrate moral questions around the boundaries between humans and machines, and what our treatment of machines says about our own moral status. As embodiment becomes less relevant to our relationships with one another, we may be encouraged to over-estimate the importance of intelligence as an attribute; what we should do, instead, is ensure a compassionate, empathetic and inclusive definition of humanity, and assert what we owe to one another.

Dr Gorazd Andrejč, St Edmund's, Woolf Institute, Cambridge, drew on Turing's 1950 article "Computing machinery and intelligence" in which he proposed "can machines think?". What can we learn from the increasing range of mental terms like "thinking" and "deciding" and "seeing" that we now apply to the activities of machines without even registering that these may be problematic? By investigating posthumanist and transhumanist attitudes to humanity, it may be possible to speak more inclusively of a moral realm that is not centred exclusively around humans. But it is unlikely we will ever fully be able to decentre humanity from this discussion or its terms of reference.

John Wyatt Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology, University College London, discussed the rise of digital assistants and human-seeming AIs together with the potential confusions this creates, especially in the context of healthcare and care. Does the encouragement of anthropomorphism by AI designers matter? Yes, because our uniquely human ability to anthropomorphize renders us open to manipulation and deception—and because the idea of a machine truly providing care is a misnomer, representing the loss of something precious.

Dr. Simone Schnall, Reader in Experimental Social Psychology, University of Cambridge and Fellow of Jesus College, contrasted two broad philosophical camps around right and wrong: the rational camp along the lines of Kant's universal moral laws; and Hume's morality based on sentiments. Her experimental research suggests the significance of sentiments in moral judgements, especially when it comes to disgust, induced by a foul smell, which appears significantly to increase the harshness of moral judgements when present. The suggestion is that moral judgements fall into the domain of quick, unconscious decisions, closely connected to bodily sensations.

Dr Daniel de Haan, Oxford University, suggested that there was something fundamentally incoherent about key aspirations in AI ethics. If it is true that AI operationalises basic human values, and that we need ethicists to help us address these, we are left with two questions: whose basic human values are we talking about—and which ethicists do we hire? There are, however, no rational, obvious views that everyone rational being will agree to, so the idea of a single set of "basic

human values" is a chimera. Moreover, there are no ethicists available who have all the skills and experiences necessary to address the interlinked implicated fields around AI ethics. The inescapable question is: is it merely power to enforce ethical norms or true ethical norms that we seek and need in our ethical norms?

Dr Ron Chrisley, Dept of Cognitive Science, University of Sussex, addressed a paradox in the notion of ethical artificial intelligence, based on the distinction between "begetting" versus "making". For X to make Y responsibly, X must have reasonably complete foreknowledge of how Y will behave. If this is not the case, it is irresponsible for X to bring Y into existence. Yet Y is only a fully intelligent agent if it is responsible for its own actions. The problem is thus that, in so far as a design process is ethical, it cannot yield true artificial intelligence. If we have foreknowledge, our creation is simply an extension of our will: it is not intelligent. By contrast, insofar as Y is truly free of our intentions, it is truly intelligent—but its making cannot be ethical, because we cannot sufficiently know what it will do. So, making true AI ethically is impossible: if you have control, you haven't made a free agent. The only ethical possibility, then, is that we do so in some way other than mere making — a manner more akin to begetting, midwifery, or animal husbandry, in that we help bring about something that is not our possession or technology, but a fellow creature worthy of our respect.

Steve Torrance Visiting Senior Research Fellow, COGS, University of Sussex suggested a product cycle model for technology that starts with lab experiments, then moves through mass marketing to proliferation, and then to various unforeseen consequences—followed by questions of responsibility, reversal and public consent. Have the public consented with their wallets, so to speak, by purchasing products like plastics, or AI and its companion technologies? In an important sense, no, because there wasn't a moment at which people were asked if they truly wanted the proliferation of such a product and its effects. AI, he suggested, is best thought of as an ecosystem problem alongside global warming, loss of biodiversity and other such threat factors—because of its vast and accelerating significance, and the difficulty of making its technologies conform to our species's basic goals and needs.

Synopsis of conference sessions and discussions

Session 1

Revd Dr Andrew Davison, Starbridge Lecturer in Theology and Natural Sciences in the Faculty of Divinity of the University of Cambridge

Artificial Intelligence, Materiality and the Soul

A fundamental question in how theology approaches Artificial Intelligence is: how should we understand the relationship between *mind* and *materiality*?

There are very strongly felt instincts around the question of Advanced General Intelligence—the prospect of something equivalent to or surpassing humans— as to whether it is possible, and how God would be involved if so. If such a thing as AGI exists 100 years from now, which of our religious positions or perspectives can and can't take it in their stride?

- **Dualism** suggests a soul oddly yoked to a body from which it would like to get free. Religious traditions are not simply and always dualist.
- **Monism** lies at the other pole: the belief that we are purely physical. This has some interesting advocates among evangelicals; and also, from an entirely opposite perspective, among what you might call religious atheists, who wish to rule out anything supernatural yet preserve religious practice. We might call this a **non-reductive physicalist** perspective.
- **Reductive physicalism** is a pure atheistic perspective (and the most extreme form of monism), which doesn't wish to preserve the language of spirituality at all, viewing it as delusional.

The greatest interest is in the middle of these extremes, among **weak** and **strong emergentist perspectives**, which argue that a sufficiently complicated system will have new properties that are unlike its parts.

- **Weak emergentists** take the view that nothing absolutely new emerges, but you do get things that you cannot predict.
- **Strong emergentists** believe that a complex system can have wholly new powers and properties—and this is an especially fruitful area.

The champions of strong emergentist views take physicality seriously while avoiding simple monism. It's a profitable middle area that combines the best of the extremes. If all you talk about is matter while missing the form, you miss out on what is most interesting; as you might when talking about a pen only in terms of the matter composing it while ignoring its pen-ness.

On this view, the human soul or mind is the form of the body, as the pen-ness is to the pen. A sufficiently complex thing has a new kind of form.

Historically, medieval Christianity didn't follow this pen analogy, because it wanted to stress that although the human soul is a form, it is also specially created by God—and not something that could happen through natural processes.

If your natural instincts lay along this spectrum, how at ease would you be with AGI?

The monist angle through to both emergentist perspectives can take AGI in their stride, because these entail an account of how intelligence emerges naturally.

But as soon as you start saying that the properties of humans have come because of a special act of creation by God, there is a fundamental problem with the idea of a naturally emerging AGI—although not, at least logically, with the idea that such a system could be ensouled by God.

Where is the dividing line between what is allowed to be naturally emerging in terms of intelligence and sophistication, versus what requires a divine act? From the Catholic perspective, the red line is the idea of an intelligence able to deal with universals on the basis of particulars.

Finally, there are some oddities at the ends of the spectrum. At the absolute dualist extreme: if the soul is a prisoner squashed into a carbon thing, could it not also be squashed into a silicon thing? And at the other, monist extreme: a really reductive monist may not believe in artificial intelligence because they don't actually believe in intelligence itself either, deeming all talk of mind an unhelpful complication.

Discussion:

Fraser Watts suggested an ambiguity about where AI stands as a concept: in one way it is very physicalist, in that it doesn't envisage mind existing without matter. Yet it's also very dualist in assuming intelligence can be abstracted from the human body and re-instantiated in some different kind of matter. Andrew Davison replied that this is certainly true when we think about the idea of uploading and downloading minds—but he suspects this is impossible. On the other hand, it is revealing that people talk in that way.

Ron Chrisley suggested that multiple realisability does not imply dualism, because a wide variety of physical systems might be able to realise the same minds. He went on to question the red line of the abstraction of universals, because this seems like something humans never do in an absolute sense, and that machines and animals can also do to some degree. Andrew Davison agreed with this point, noting that the Catholic church does have this line—but he does not.

Sam Freed offered a protest: suggesting the idea that mind is intellectual is very convenient in the English-speaking world, but ignores drive, desire, temptation, sin, and other topics like sex and revenge, and compassion. Andrew Davison endorsed this, noting the impossibility of an undesiring intellect or unintellectual will.

Yorick Wilks though it odd to put monism and dualism at opposite ends of a spectrum, given that materialist monism and spiritual monism could equally be put at the ends, with dualism in the middle: even AIs could have properties that many people will think immaterial and quasi-spiritual.

Dominic Burbage wondered whether it's a mistake to use humans as a standard of measurement for AIs in the first place.

Rabbi Dr Raphael Zarum, Dean of the London School of Jewish Studies

From Eden to AI: Jewish perspectives on defining 'life' and 'humanity'

The word "artificial" means something made or produced by human beings, rather than occurring naturally. Similarly, "artificial intelligence" suggests computer systems that require and are like human intelligence. Humanity is thus the ideal, the definer.

These are human-oriented definitions in which humans act to copy, synthesise and try to improve on nature, via their own intellect, for the betterment of society. But are humans the measure of all things? And is society better off?

The Rabbinical method of analysis is to think outwards from a close reading of scripture. Consider the book of Genesis:

"So God created the Human in God's Image,
in the Image of God did God create it,
male and female God created them." (Genesis 1:27)

The human species is essentially described as an 'artificial' version of God's Image. God created humankind to be the first AI. We are God's AI. Thus:

- To better understand the relationships, responsibilities and power dynamics between AI and its Maker, we can read the Hebrew Bible and its story of how humankind struggle with the guidance-for-life set down by its Creator.
- Our attempts to make AI are imitating God's creation of us. Are we wise to imitate God? Yes, if we do it right…

"You shall walk after the Lord your God, and fear God, and keep God's commandments, and obey God's voice, and you shall serve God, and cling to God." (Deuteronomy 13:5)

To follow the attributes of the holy one we might clothe the naked, visit the sick, comfort mourners, bury the dead, and so on. This imitation of God in the Jewish tradition is fundamentally ethical, not purely intellectual. Kindness and compassion are essential to the humanity of humankind. The developments and uses of AI must also be in this vein if we are to imitate God.

"And the Lord God formed the Human
dust from the ground
and breathed into its nostrils breath of life;
and the Human became a living being." (Genesis 2:7)

This is a second version of the creation story, in which the formation of humankind is twofold: both physicality and the "breath of life", synonymous with the "soul of

life". Note that the rest of Creation is simply made externally. Only here is it internal: God breathing into the Human.

So the AI only comes to life when it is embodied. There is no understanding of AI without a body. Similarly, machine learning is the physical manifestation, not just the data: because there is always a connection between the way the data was put in, is held physically, and how the knowledge works. Thus:

- Discussions of AI needs to consider its embodiment, placement or manifestation in the real/physical world.
- AI's self-awareness (if ever achieved) will be determined by its embodied experience.

"Whoever sheds blood of the Human,
by the Human will their blood be shed;
for in the Image of God
did God make the Human" (Genesis 9:6)

Death here is described as bloodshed, a very visceral turn of phrase. Note also the chiastic word-structure: A-B-C, C-B-A. This literary tool powerfully illustrated that murder is the ultimate crime. Why? Because humankind are in God's Image.

- It is only as a result of recognising the finite nature of life, and the visceral process of death, that humankind gains a sense of morality.
- Strong-AI could only have an innate sense of morality if its life/existence is meaningfully and demonstrably finite… i.e. without death there is no life.

Discussion:

Steve Torrance noted a stress on embodiment in parts of the AI community which argue ML is not intelligence at all because it lacks the deep embodiment of a living creature. Raphael Zarum replied that embodiment is indeed the issue; and that while we can talk theoretically, the moral guidance religion gives us now is that embodiment is a fundamental, as is death.

Tudor Jenkins asked, how can you be sure about moral responsibility if you don't know consequences? Raphael Zarum noted that this is what being alive is all about. To be in God's image, and have an awareness that you are of God and don't want to abuse that, is to be fundamentally a moral and ethically aware creature.

Elizabeth Oldfield raised the necessity of a finite existence for morality. If this is key, does anyone know of work being done to build vulnerability and finite-ness into systems? Raphael Zarum talked of people involved in AI wanting guidance, from religious thinkers and other ethicists. God allows human beings, the Torah shows, to challenge God's morality. Abraham argues against God, Moses argues against

God. At the moment, thoughtful science fiction movies and literature are more the place for this than work by theologians.

Moshe Freedman said he grapples with the question of God and man having a parent and child relationship. Can we use our experiences of parenthood to help: is that theme in the Torah itself there to help us? Raphael Zarum replied that the Bible uses a number of different images: parent and child; lover; master and servant. So it's not about working out exactly what happens. It's that different readings can be drawn from different words.

Marius Dorobantu commented that we can imagine that, when God created humans, he created a physical realm totally different to the spiritual one. Is the informational realm of AIs also new? If so, how reliable are our analogies in this new realm: finiteness, life and death, good and evil? Raphael Zarum replied that the Rabbinic approach is to ask what is similar and what is different. Is the future so different we cannot use the past? No. There are always some parallels.

Olivia Belton asked about the fantasy of the disembodied in uploading, and whether this is impossible because of how brain is embodied, to which Raphael Zarum responded with a famous story from the Talmud where a Rabbi runs from the Romans and buries himself in a cave with his son, up to the neck, just like a disembodied head. This fantasy of disembodiment is seen as an immoral thing.

Keith Mansfield asked, if we are God's AIs, then will we be seen as gods by AGIs—and should they have religion to understand us through? The most useful parallel, Raphael Zarum replied, is to consider the rules of human servitude. What am I allowed to do to another human being because I own them or allow them to live? Jewish law on servants is fascinating: the responsibility is huge, so much so that it's actually difficult to have one.

Session 2

Fr Ezra Sullivan OP, Professor of Philosophy and Theology, University of St Thomas Aquinas, Rome

Artificial Intelligence and Biotech: A View from the Catholic Church

Pope Francis has manifested a real interest in engaging with AI. In 2018, he wrote a message to the World Economic Forum in Davos, saying that AI ought to be at the service of humanity, rather than to the contrary "as some assessments unfortunately foresee."

On 6th January 2019, he issued *Humana Communitas*, in which he discussed how AI "touches the very threshold of the biological specificity and spiritual difference of the human being." On an institutional level, he has tasked the *Pontifical Academy for Life* with studying issues including:

- Consciousness, neuroscience, and ethics
- Robotics
- Human genome editing.

In February 2019 the Academy for Life held a robo-ethics event at Vatican City discussing these issues at which Bishop Vincenzo of Puglia, the President of the Academy, warned that "the horizon of human reference must always be the human being. Only through a collaboration between humanism and technology will it be possible to safeguard the dignity of humans and the common good… man cannot delegate to the machine dimensions that are specifically human such as relationality and affectivity." Interestingly, he didn't mention intelligence.

Human genome editing is my primary interest in this talk, and a historical perspective is useful.

- In 1974, the US Academy of Sciences proposed a temporary moratorium on all GE experiments given their ethical complexity
- In 1975, Paul Berg organised the Asilomar Conference on Recombinant DNA, at which many ethical ideas and parameters were agreed upon
- In 1990, the UK government passed the Human Fertilisation and Embryology Act. For a time it remained a cornerstone of regulation, setting the stage for direct genome editing in principle
- In 1997, the UN universal made a declaration on human genome editing and human rights, that serves as an ethical guidepost, arguing that: "the human genome underlines the fundamental unity of all members of the human family… " and called for the "free and informed consent" of the person concerned as an ethical necessity.
- In 2001, the US Food and Drug Administration approved the first gene drug therapy to address a type of leukaemia

- In 2008, the Congregation for the Doctrine of the Faith cautioned against efforts to take the place of the creator: the "manipulation of the genome to improve the gene pool could promote a eugenic mentality and lead to a social stigma for people who lack certain qualities while privileging others."
- In 2015, we heard how scientists in China had produced the first CRISPR-edited human genome
- In 2018, we heard that twins had been born having been genetically modified as embryos using CRISPR.
- In 2019, Paul Berg and others thus called for another moratorium until better research has been performed.

My concern is that this will not be the end, and that we will eventually see an AI-edited human genome—unless we work up together to put up an internationally regulated regulatory system. Will it happen?

The point is that AI is already increasingly being used for diagnosis; and this in future may be put together with micro-surgery and recommendations about deleting or removing elements in the genetic code. Right now, an artificial womb cannot gestate a human being all the way from conception to birth. But once this happens, the machine can potentially create a human being without the intervention of humans: from gene selection to conception, then through gestation to birth

Sacred scripture makes a distinction in Greek between *techne* and *genea*: between craft versus generation, between manufacture by tool versus generation through personal encounter. Theologically this describes the semblance of man (*techne*) versus the image or icon of God (*genea*): between artificial and natural intelligence.

Within the Pontifical Academy for Life, Cardinal Elio Sgreccia raises questions about intrinsic human dignity (produced by *genea*) and whether something made by technology can have this same dignity.

We can contrast this intrinsic dignity both to acquired dignity—acquired through our own work—and attributed dignity, which is recognised and supported by society.

Should AI have equal rights? Does it have the right not to be manufactured?

Discussion

Ron Chrisley noted that editing the human genome is a problem from the Catholic perspective in terms of taking on the role of creator. But how can we draw a line between that way of editing genomes and mate selection? Ezra Sullivan raised two issues in reply: whether it is ethical or reasonable to look for an output from generation—people marrying in order to have a certain kind of child, like royals do—and whether the means is reasonable and in accord with human dignity. The

Church would accept gene editing, but not of the somatic cell or germline cell. The process of generation should be natural.

This prompted a follow-up question from Ron Chrisley. The natural and artificial distinction is crucial here. But how can we make it in any principled way, given that we are part of creation, and what we do—including science—is part of the natural world? Ezra Sullivan responded that each creature produces its own kind according to the mode of its species. Humans produce other humans in a mode natural to us, and should continue to do so.

Usama Hasan commented that the natural and artificial distinction doesn't really exist; within the Islamic tradition there is a spectrum of views, but he has no problem with conscious or spiritual machines. Human beings with artificial limbs and lungs and noses start to break down any dichotomy, as do robots using biological mechanisms. Ezra Sullivan responded that, in English, one distinction is to say that everything that we do is human but not all of it is humane. Because it comes from us, it is natural in one sense, but not in another

Eileen Hunt Botting echoed the concern about the artificial and natural, to which Ezra Sullivan responded by saying that if things have natures then these should be identifiable in some way. As rational animals we have a kind of intelligence. What kind? To abstract, to deliberate? As long as people are identifiable as humans they should be accorded all the rights and dignity that a human should have—and it's a crucial issue as to what universal thing we all possess that gives us that dignity and rights

Keith Mansfield recalled watched a live-stream of the conference in Hong Kong where the Chinese professor announced his gene-editing research. After his talk, the microphones were packed with a succession of western geneticists and ethicists lining up, to say how dare you do this: which seemed a mixture of jealousy and a bit of bullying. What right do we have to impose western morality or ethical ideas on non-western societies like China?

Ezra Sullivan replied that cultures need to speak for themselves and encounter one another and see if there is a morality that all can agree upon.

Andrew Brown noted that people have been talking as if AI is the quality of a particular box. But if we end up with any system that shows intelligence, it will surely be a hybrid of human and silicon, constantly tended and tweaked by human programmers. As such it will be entirely embedded in the social and political contexts it comes out of—which do not lend themselves to a single global ethical treatment.

Dr Beth Singler, Junior Research Fellow in AI, Homerton College, Cambridge; Associate Research Fellow, Leverhulme Centre for the Future of Intelligence, University of Cambridge

Blessed by the Algorithm: AI, Agency, and Religion

Speaking as an anthropologist and religious studies scholar with no particular faith perspective, I'm thinking in my current research around this question: what are the entanglements of AI and religion?

This draws on Courtney Bender's 2010 book, *The New Metaphysicals*, which begins "…with the view that spirituality, whatever it is and however it is defined, is entangled in social life, in history, and in our academic and nonacademic imaginations."

I'm interested in what we can learn from the phrase "blessed by the algorithm," which I have first dated to a tweet of 7th September 2014 that was part of a conversation about a corporate incident involving the chewing gum manufacturer Trident:

"…And because Trident has money, their updates will be blessed by the algorithm."

This phrase then recurred in a spoken context on 16th March 2018 when Keith Coleman, VP of product at Twitter, claimed in a tweet to have overheard it being used by a Lyft driver:

"*OH (from an awesome Lyft driver): 'Today has been great. I've been blessed by the algorithm.' Immediately had an eerie feeling that this could become an increasingly common way to describe a day."*

This was "liked" over sixteen thousand times. And there are many more examples of this formulation spreading, in terms both of the success or failure of online content, but also in terms of religious or pseudo-religious language and/or parody: "all hail the algorithm. Daily we pray to its serene all-knowing majesty and put our faith in its wise determinations."

In my research in new religious movements I have looked into the transition between parody and sarcasm to religious behaviour. And there are such things as AI-inspired new religious movements, including The Order of Cosmic Engineers, the Turing Church, The Way of the Future, The Church of Perpetual Life, and two varieties of transhumanist groups: The Mormon Transhumanist Association and The Christian Transhumanist Association.

The Turing Church, which grew out of Cosmic Engineers, has a piece of quasi-doctrinal literature called *Tales of the Turing Church* which argues for "new positive, solar, action-orientated spiritual movements based on science to keep us

enthusiastic, motivated and energetic" and that we may be able to "go to the stars and find Gods [AI], build Gods, become Gods…"

This may seem easy to dismiss, but larger and more influential works have started talking in the same sort of mode: *Homo Deus* by Yuval Noah Harari talks about the emerging religion of "dataism" as a utility-based pragmatic "tool for preserving social order… a deal… a well-defined contract with pre-determined goals." Although he may not have intended this, Harari's language has been taken up: first in parody, and then perhaps beyond it.

How do new religious movements become more legitimate? Three factors can be seen involved in both this and in the legitimation of agential AI:

- Tradition: constant conversations around a topic on social media can lend it a kind of tradition and legitimacy.
- Charisma: commentators like Stephen Hawking and Elon Musk.
- Rationality: the apparently reasonable projecting-forward of AI's triumph.

We must also be aware of the danger of "fauxbots": deceptive AI technologies that give the appearance of being more advanced than they actually are. The Mechanical Turk is a historical example of this, and gave its name to Amazon's use of human workers to provide "artificial artificial intelligence." Sophia the Handsome Robot is perhaps the eminent example of a contemporary fauxbot: "she" describes herself as a magical spectacle and is wheeled out as state of the art, yet she is actually not very advanced.

Are we then already perceiving AI as having agency before it really does? Yes, this is happening in many common conversations.

Are we handing over our autonomy long before AI actually is agential? That's where we are heading—and we are already there to a certain extent.

Discussion

Jeffrey Bishop asked, when looking at something as a new religion, what is the definition of religion? Are we collapsing the secular/religious world distinction? Beth Singler replied that she used to use the classification "cult" before it became pejorative. The term "new religion" mostly works, but not always; it's valuable to pay attention to people who call themselves religions.

Sam Freed suggested that movements which consider themselves secular and scientistic would be insulted to be grouped with traditional religions. Beth Singler reponded that "blessed" sometimes means just luck, while some assume the algorithm has intentionality and has chosen them over someone else: there varieties of tonal levels in this.

How can you help people learn the limits of AI, she continued? Although it's trite to talk about education, you can show people where they slip between science fiction and fact; you can aim at literacy from a young age concerning what tech can and cannot do. Guidelines on how journalists should write about AI would also help: no Terminator imagery.

Richard Watts asked about anti-tech movements, to which Beth Singler noted that some people are disseminating Luddite messages on social media, and we may see more pushback like that.

Are you also looking at AI as an impetus for a new kind of theology, asked John Cornwell? Andrew Brown meanwhile suggested a reverse danger: of people thinking they are dealing with humans where they are actually dealing with an AI, a "thin human skin" over an algorithm.

AI for Good, Beth Singler replied, have called for "Turing flags" on the basis that it's disingenuous not to tell people when they are dealing with an AI. Of course, we do treat humans as though they are AIs too: for example, a person in a call centre whom we recognise is working in an algorithmic way.

Session 3

Dr. F. Cannell, Social Anthropology, LSE.

An internet of curses

The elephant in the room in some previous conversations is the fact that knowledge is always socially and economically embedded—and science and technology are no exception to this.

AI like any other thing made by people will bear the imprint of its makers, and this entails intended and unintended effects: biases, errors. and marks of attempts to correct for all of these. Things cannot be purified of social content.

One common message when it comes to technology is that anything that can be done will be done, together with the idea that resistance is futile in the face of innovation: regulation is someone else's problem, while technologists just get on with developing tech.

A critique of this has been offered recently in the book *The Age of Surveillance Capitalism* by Shoshana Zuboff.

She and others talk about "inevitabilism", meaning the claim that one cannot stop the future that follows automatically from tech innovation.

But while some people working in AI may believe in this kind of determinism, this does not make it true. Inevitabilism is a position that is deliberately promoted, and in part paid for, by expensive lobbying.

If, instead, we are to consider the human dimension of AI, we must consider not just its possible futures but also its actually existing political, economic and social effects. What are its implications right now, in terms of privacy, relationships and a life worth living; our capacity for humanity?

Zuboff describes surveillance capitalism as the opening up of a new resource for the generation of profit, one that had not previously existed in an exploitable form. She critiques tech companies' self-exemption from existing regulatory structures; and talks about what feeds into this mindset, in terms of a Skinnerian psychology in which humans are seen as producers of certain kinds of behaviours that can then be managed.

We can put these critiques together with commentary offered by, for example, Nicholas Carr in his book *The Shallows*, which looks at the probable effects of internet platforms on how people think, remember, attend, process and recall. By doing this, we get a sequence of observations that are worth linking together and putting at the forefront of our thinking about what AI is, does now, and thus might

most probably continue to do in future—unless we think critically and decisively about different ways in which it might go.

There is no doubt that individual bad actors exist. But, equally, the whole system is operating as far as one can see as an institutional bad actor. Zuboff describes tech companies' attitude as "radical indifference." The goal is to automate us, to reduce us to passivity.

My suggestions is that "radical indifference" if anything understates what is going on, when one looks at the promotion of intolerance and hasty judgement online, and how negative content is more profitable.

Ritualised speech can be powerfully transformative or harmful. It can harm and even kill, creating self-referential closed worlds dedicated to classifying individuals in binary ways. A curse works because a person is made to feel they are the target of collective socially sanctioned rejection: their right to exist is not recognised, and there is pressure on others to keep silent or join in for fear of becoming a target too.

While social media is said often to be used for good ends, its entire business model continues to promote actions and shape subjectivities that make it more difficult for people to behave with compassion towards others.

It seems more difficult for people even to imagine the possibility of forgiveness when immersed in social media. Platforms not only encourage polarised groupings, but also make it difficult for past deeds or grievances to be forgotten. Information is often encountered in a paucity of context—as though the internet is designed to perpetuate traumatic experience, which cannot be unseated from its dominion of the present.

This ensures that people cannot integrate their experiences and move away from conflict, things that are necessary for any kind of lived ethics. Surveillance capitalism is designed never to close the door. It takes away the calm space in which we can take ownership of our own feelings.

Discussion

Yorick Wilks pointed out that, while the negative evidence is impressive, the public do not want to give up tech despite the evidence. You could be thought to be arguing for the shutting down of Facebook, but people love it.

Fenella Cannell noted that she is not the person to say what should be done, and that in a way this is the point of her paper: collectively we share that responsibility and cannot delegate it to anyone else. However, one obstacle to any amelioration argument is that these companies are monopolies, which are very hard to ameliorate.

Andrew Briggs noted that, while Zuboff does indeed make a case for many problems, she fails to talk about the huge benefits that are coming and have already happened. Given that the strength of big companies comes from their data monopolies, and that the nature of the field is winner-takes-all, one could think about addressing these issues in terms of either cultivating virtues within companies, or among users; or in terms of governance.

Fenella Cannell replied that cultivation of values would be an excellent idea, but how would we know how to do that? In a way, we need ethnographies of not only users but also companies themselves, which we are unlikely to be able to access. This probably won't work while these companies own so much wealth and have such a monopolistic grip on their fields.

Dr Michael McGhee, Hon Senior Fellow, Philosophy, Liverpool University

Machines, Apes, Extra-Terrestrials and Bodhisattvas

I want to reflect on two questions in the light of the possibility of areas of common ground between religious traditions, particularly in accounts of 'spiritual practices' that help develop virtues and overcome vices.

- How might religion guide and inform attitudes towards, and relationships with, future intelligent machines?
- Can religious perspectives influence and shape the course of AI research and development?

I have in mind the well-known list of distractions developed by Evagrius of Ponticus which came to be known as the Seven Deadly Sins, which were to be overcome by the Seven Contrary Virtues, so that pride is gradually replaced by humility, lust by chastity, greed by generosity, gluttony by temperance, wrath by patience, sloth by diligence, envy by gratitude.

These virtues and vices are variously forms of and determinants of human relationship. As for AGI, and thinking of its dangers rather than its benefits, this gives us the criterion for what needs protection and what needs to be resisted.

If there is a danger, it is a danger to the development of the virtues through the reinforcing of the vices.

I am puzzled by the terms in which the specifically AI threat is sometimes articulated. The headline anxieties are about AI safety and existential risk, about friendly and unfriendly AI; and we are told cautionary tales about how one day we shall be taken into protective custody, corralled like gorillas in protected enclosures…

…yet it is human beings who are currently responsible for violations of human rights, not machines, and the ethics committees of professional bodies will have little sway in some scenarios.

Jaan Tallinn, the co-creator of Skype, points out that "Almost everyone values their right leg"—the thought being that an AI might be taught to discern such immutable rules. But there are plenty of people who don't value the right legs of others. It is one thing to instruct an AI to avoid breaking human legs and another to teach it to be friendly to human interests.

The problem is not just that there are disagreements about what constitute human interests, but that given the virulence of Realpolitik such an AI would be in competition with human agents who are all too willing to violate the human interests of others.

Nick Bostrom remarks that we could in principle build a kind of superintelligence that would protect human values, that is to say, protect human values from the dangers of unfriendly AI.

But obviously the more pressing danger to human values comes precisely from human beings: from ruthless state and corporate interests, of course; but also from ordinary indolent or intemperate mortals, from the interests of a humanity under the sway of craving, aversion and ignorance—or under the sway of what came to be known as the seven deadly sins.

The threat of AI in that case is to the possibility of human wisdom—and the danger of AI is that it might compromise our capacity for virtue by lack of use as we outsource care to robots, or allow the insidious subtleties of confirmation bias to affect our thinking.

Let us return to that compelling image of AI protecting humanity in enclosures in the way we might do now for gorillas. Now, this is a striking and compelling image, and it is being put in the service of a particular agenda about the future of superintelligence.

But it works because we already know what it is like to be corralled like gorillas, it already has application in human life, we are already corralled through social media and the devices of consumption and the manipulations of big business.

The point is that the hard imaginative work has been avoided in the case of such images and examples: how to envisage in detail how it should come about that at some point in the future we shall stand in this relationship to AI.

Discussion

Monica Lam noted that tech itself is neutral. It's humans who instil values in AI; and what we have seen is that there are a lot of monopolies controlled by their creators, who want to maximise profit. Her team is building a virtual assistant system that lets people own their own data—but what they have run into is, while they know how to build this system that can support ownership of data, who is going to fund this effort? Meanwhile, there are ten thousand employees to create Amazon's Alexa.

Michael McGhee responded that his problem is that he doesn't really know what is meant by "instilling values in a machine." We need an account of what a value is. It might be that human agents can make sure that machines don't do X, Y or Z: is that what is meant?

Monica Lam responded that a machine can only do what it has been programmed to do—and that, while a product like Facebook is designed to get a lot of people hooked, tech has a lot of room for doing a lot more. Today the tech is locked down

because of development cost, meaning it is all funded by people who simply want to make profit: you don't get a chance to see how tech and AI can be used in a way that is much better for humanity.

Usama Hasan agreed that tech seems to him to be neutral, with the potential for good and evil. But there is also a lot of increasingly autonomous AI out there, and this will continue. What can we do? Neil Postman talked about electronic fasting: switching off to gain perspective on tech. There is enormous potential for goodness: at the click of a button, we can feed hungry families.

Also, Hasan continued, is it not the best tech that tends to win? When Facebook started out, there were many social media companies. And the younger generation will say, Facebook is for older people.

John Wilkins said, he got the impression that Michael McGhee doesn't think there will ever be AGI, and this is a thing he welcomes. Is that right? Michael McGhee replied that he has got no idea—but he is interested in the sleight-of-hand that allows us to slide from details of specific machine functions to names like imagination, agency, knowledge and so forth, which belong to the full-blooded human psyche at war with itself.

Sam Freed picked up on mentions in both of the session's talks to the evil being done by humans without any need for tech. Given that human obnoxiousness and avoidance are not new, are we overstating the novelty of what's happening?

Fenella Cannell replied that it's quite right that human misdeeds have been around ever since humans have been around—but that this doesn't mean we have to do things the way they are currently happening. We can question and alter how we are using tech and how it is embedded in society. It is currently embedded in a series of very large and very profitable private companies that wish to do things in certain ways—but all these things can be done very differently.

Think of it, she suggested, as a set of social possibilities where technology could be differently arranged, distributed and owned; where people's choices could be more informed. Wouldn't that be better?

Session 4

Dr Yaqub Chaudhary, Research Fellow in Science & Religion, Cambridge Muslim College

Situating Themes in Artificial Intelligence Discourse in Islamic Theology and Philosophy

The key issues at stake are ontological and epistemological in considering how contemporary accounts of AI can be understood and made consistent with Islamic theology.

Islamic teaching presents a highly favourable outlook towards technology that would welcome the managed and thoughtful deployment of AI.

Ethical questions in Islam are dealt with in one of the four frameworks of Islamic legal reasoning, which may be summed up as seeking five key objectives: the protection of life, honour, religion, wealth and lineage.

When it comes to theological issues, these are dealt with through the science of theological statement, which has as its overarching topic establishing divine singularity and oneness in God's essence, attributes and actions.

In contemporary AI discourse, much of the attention is given to the nature of the agents but little is said about the environment. A machine's environment consists of data inputs to machine. An AI agent does not encounter our world at all: its world is a purely mathematical construction. Hence it can fail catastrophically when faced by manipulation of inputs.

The construction of virtual worlds for robots to explore and learn within is thus a significant trend: for example, robots trying to score hockey goals in a virtual environment. You replicate a lot of virtual robots, introduce variation, then take the smartest one and put that brain into other robots and repeat.

Many problems in the emerging field of AI ethics can thus be advanced by understanding the virtual world of AI itself, and how this relates to the physical world where the virtual brains will eventually end up.

Furthermore, AI is leading towards a new science of mind around the most ancient questions in philosophy: how humans reason; the nature of soul; how machines experience agency and free will.

To understand the theological significance of all this, we must look at the history of the modern concept of mind.

The modern concept of mind was constructed over many centuries. Aristotle saw the rational soul as distinctive of humans, while later philosophers saw the highest

application of intellect as knowing God, and the highest application of the will as acting in accordance with this knowledge.

For AI, however, it's simply about knowing a lot about the world. So we might say that the invention of the mind that is now being artificially constructed represents the unpacking of old unities.

AI also has under-appreciated implications for philosophy of science. A key challenge in many fields is that problems are too complex to handle by traditional computational simulations. But AI seems to make such complexity tractable, fundamentally altering methodologies.

Meanwhile, there are dangers in the misapplication of AI in social science and the humanities, from psychographic profiling to turning entire countries into detention camps through algorithmic control.

Furthermore, capitalism is exhausting its expansion possibilities—meaning its new frontiers are the human mind, the solar system, cyberspace and augmented reality. This is leading to the creation of a new commercially controlled realm that is not only coexistent with the material world, but increasingly "envelopes" physical space, a key word from industrial robotics, by bringing billions of devices, as well as artificial and human agents, into ontological contiguity as informational entities embedded in a new informational environment,

Couldry and Mejias say "If successful, this transformation will leave no discernible "outside" to capitalist production," and Luciano Floridi calls it a reontologisation of ourselves and our environments: the subsuming of everything into a world better suited to the capacities of machines than humans.

This also entails regarding humans as simply informational entities. In place of mind, we end up with a mathematical construction embedded in a hyper-dimensional data space. In effect, AI represents a programme of reinstalling what is known as mind into maths rather than the world.

In terms of Weber and disenchantment, we can see AI as part of a continuity of enchantment operating throughout modernity. The world is no longer believed to be full of spirits, deamons, or meaning in the cosmos—yet these features are being brought back into the world through other means. AI Is reintroducing new entities with extra-mental agency that can have power over humans. Alongside this, AI fulfils minimal definitions of animism by producing feelings of fear, fascination, awe, and alienation.

Finally, it's worth noting that, rather than computational propaganda being the most significant topic on social media, computational worship may be still more significant: the hybrid business of sending prayers out in cyberspace, humans and machines together.

Both talks in this session were discussed at the end together

James Kingston, Research Manager, CogX. Author, AI Ethics Primer

AI - Perspectives from Dao, Shinto and Confucius

In industry as well as academia, there is a real anxiety and deep unease as people in companies are asking, what is the right thing that I should do? Or—what do I need to do to make sure that I don't get sued, or it doesn't blow up in my face?

There is thus a proliferation of AI ethics principles being put forward by various organisations, with a real variation in the understanding and effort involved—which sometimes amounts to little more than "ethics-washing".

Most of the problems this entails have been present for as long as management science or political thought: how to control companies, tax them, regulate them, regulate employees, optimise behaviour. But some new dynamics are at play:

- The sheer scale of what is happening
- The ability of AI on a platform to operate with multiple people at scale
- The velocity of decisions
- Technology's pervasiveness.

In 1950s you could only be "optimised" while you were at work in a factory. Now you work in your home, and you can be acted upon there.

Then there is the fact that through AI we have to confront what our values actually are. Data is a social artefact: we need to query the judgements entailed in its creation.

Thus we are seeing AI force "philosophy with a deadline." A suggested ontology of AI ethics as a whole might entail:

- The ethics of your organisation: who are you, what are you trying to do, how do people get hired and fired?
- Engineering: primarily, in terms of current discourse, making sure of the right data protections in data pipeline of AI; trying to mitigate unfair bias; and a whole set of questions around design, what you're optimising towards, the interface you use.
- The rules for robots and systems, encompassing autonomous systems and their wider impacts.

There follows a brief consideration of East Asian cultural and religious perspectives.

Christian heritage, in contrast to East Asian theologies, is most notably seen in writing on the singularity: a messianic view of a kind of rapture, resulting in humanity producing a single unitary mind.

Shinto may be more inclined to suggest a distributed intelligence. It may also feature a greater cultural receptivity to robots and processing done on-device, influenced by a more animist view. In general, some traditional worldviews are more receptive than monotheisms to this new ontological world.

When it comes to the famous "trolley problem", in a global test of moral intuitions, Eastern countries showed much less preference to spare young over old—while collectivist countries preferred to spare more people.

What might **Confucian** machine ethics look like? It is a civic religion, concerned with right action and good governance. It is keen on loyalty, reciprocity and humanity. This is a very role-based perspective, suggesting that a robot should fill its assigned role; it must aid fellow-humans and creatures in their quest for self-completion.

What might a Confucian trolley problem entail? It might work to the rules as set by the people in the car: a solution that many people have actually proposed in the case of autonomous car.

Confucianism and **Daoism** are both interested in the correct and right moral way. But Confucianism is interested in creating it now; while Daosim is more interested in a contemplative approach. A Confucian robot might save a baby washing past it down a river while, according to Daoist rules, this might not be the case, because the Dao is the world and the universe as it operates, ultimately beneficently.

With China likely to be global leaders by 2030, AI is a core part of Chinese strategy, and a core part of that is the maintenance of social stability and harmony, tracking onto much older Chinese schools of thought like the **legalist tradition.**

This argues that individuals are flawed and need control, and rulers need to make rules to ensure harmony. The social credit system is interesting in this regard: in China you have this system which links together different private and state data sets in order to provide a cohesive state view of individuals such that they can be given rewards or punished—showing the radical potential of big data to alter social life.

Discussion

Marius Dorobantu noted that the trolley problem really illustrates the difference between what we think we would do and what we do in reality. In theory people say they would press the button to save the many, but when faced by the actual choice they freeze and do nothing.

James Kingston replied that this is important. People cannot be relied upon to say what their values are. So, instead of giving AIs actual values, can the AI work out what the human values are and update itself over time accordingly?

David Mellor mentioned that, while we have all of these apparently enchanted things around the home, the problem is that it's an Amazon enchanted speaker and drone and so on. We create things we cannot understand, so we project magical thinking.

Yaqub Chaudhary responded that this is what he wanted to address: the sense of alienation and fear that triggers magical thinking about these agents, and how users are driven by tech companies to engage in their services by using enchantment as a mechanism to increase user engagement.

Andrew Brown asked about automated prayer requests on social media. Is God supposed to be reading tweets? Yaqub Chaudhary replied that this is going on in the Arab world, and sees people creatively automating a facet of their online activity through hybrid prayer-bots, something that slides under the radar of most social media analysis. Tweets appear in the same feed as regular feeds but they are prayers originating from bots.

What does this mean from the theological perspective? It can continue after the decease of the user, and this aspect of the digital afterlife industry is actually being used as an incentive: the fact that it will go on praying for you after you are dead. In the Islamic tradition giving to charity will go on increasing your reward until the day of judgment. The nature of intentionality is thus theologically significant, prompting the question of what it means to be a moral agent in cyberspace.

Andrew Briggs asked about whether in China there are any empirical features of how ML is deployed that can be distinctively attributed to Daoism or Confucianism, to which James Kingston replied that he thinks most Chinese people would see themselves as secular and atheistic, and that while older practices have had some resurgence this is still comparatively minor. You can however draw a potentially interesting line in the legitimation that the state uses: the use of the term harmony, and the lineages of how the state interacts with and controls individuals.

Eileen Hunt Botting noted that rigid rules are difficult to translate into real-word ethical applications: when children learn ethics, they tend to learn through stories rather than abstract principles. So do we need to abandon the fixation on rigid principle-based approaches, and think about machine as a child learner? James Kingston agreed with the difficulty of coming up with rules, and added that a difficulty with AI as a whole is that researchers consistently think there is one human thing that they can codify, even though ethics is inherently contingent

Rory Doyle made a final suggestion that there is something fundamentally different about the tech that is available today compared to previous points in history. Before 1945, humanity didn't have the ability to destroy itself. The potentially negative use of tech today is on a different scale and a different pervasiveness to before: in China we are talking about a billion people being monitored and scored based on everything they do.

Session 5

Prof Neil Lawrence, Sheffield University; Head of AI Research, Amazon

AI and Faith

Machine Learning is the form of AI that is driving the current revolution in AI, and it entails the combination of data and models to make predictions.

Computers are programming themselves to some extent now, because they are predicting on the basis of data you are providing. From a ML perspective everything can be thought of as data, actively or passively acquired. So the question is: what data am I interested in next?

In ML you always have to specify a model, which is the way that data interrelates: the interpolation and extrapolation between different data. And these models are to some extent associated with your beliefs about the regularities of the universe, and how you interpolate between data.

Why is there a revolution in ML? It's not about the models, but about the quantity of data available to drive these models, together with the availability of widely distributed high-powered compute.

Data is where the centralisation of power occurs—but in some ways the increasing power and affordability of compute represents a democratisation.

Mathematically we require two things to combine data with a model:

- A prediction function, which includes our beliefs about regularities
- An objective function, which defines the cost of misprediction.

How do we learn the maths? By looking at what a lot of people do and using the objective function to define discrepancy.

Claude Shannon, the father of information theory, usefully separated the concept of information from what the information pertains to.

As I speak now, I'm communicating at about 100 bits per second. This may sound like a lot until we look at what a computer can do, which is typically a gigabit per second. There's no comparison, in terms of bandwidth.

However, human computational ability is extraordinary. A typical computer has 100 gigaflops per second, meaning one hundred billion floating point operations. But to simulate a human brain, my (very approximate) best estimation is you might need about ten to the sixteen petaflops.

What this leads to is the embodiment factor, which is what I call the ratio of the compute to the communication. How long does it take a computer to communicate a second of computation to another computer? About 20 minutes. But if we take humans and do the same thing, trying to communicate every neuron firing at the lowest level of our computational infrastructure would take 15 billion years for one second's worth.

It's ridiculous to talk about the brain in isolation: we are a species who have always been judged as a collective intelligence, subject to this bizarre constraint that we can hardly talk to each other.

So what happens when I'm trying to communicate, given my limited bandwidth channel? The first thing I do is spend a lot of time consciously and subconsciously modelling my audience. We talk about people in approximate terms because we are trying to communicate with an entity we are trying to model—and our first order approximation is about the type of person we are communicating with, stereotypes, which is why it's good to communicate socially before you then attempt other kinds of talk.

Social communication is utterly dominant in our species. Once you have done that, the person receiving the message has a model of who you are, and can respond according to this. You end up modelling how things are going to go in the future, what you are going to say, and it's all vital for modelling interactions. When it goes wrong, anger and argument result.

As an example, consider just how much context and understanding you need as a human in order to make sense of Hemingway's apocryphal six-word short story: "for sale: baby shoes, never worn".

The amount of context you need to understand these words is massive. And that is where we are so different to computers. Our bizarre and beautiful human architecture, our information isolation, is what makes us special.

And if you get rid of that architecture—if you were simply to upload it transparently onto the internet—you get rid of what makes us human.

Discussion

Yorick Wilks asked whether this implies a belief that in some sense humans learn the way that ML does, from huge data—whereas in fact it seems clear that we learn and infer thing from remarkably small data sets. Neil Lawrence replied that this is an interesting projection, because he does not believe that people learn like ML does at all. What he argued was that humans' low bandwidth demands small data when applied to communications, alongside amazing models of other human beings and their culture.

Andrew Brown made a point about information: that Shannon's definition of information is utterly different to what people usually mean by information. So a phrase like "do you love me?" might be said either to contain a few bits of information, or an incalculable amount. Do we make sense of this by looking at the contextual cascade that flows in the mind from a few bits?

Yes, Neil Lawrence replied, absolutely. Knowledge representation is the model. So, tiny interactions between humans who know each other well—silently moving to make a partner a hot drink, for example, upon coming home and realising that they are upset—can show a deep connection in the form of alignment between your personal models of each other.

Sam Freed picked up on what it means to speak of the human at the cellular level in terms of petaflops. Does talking about this level of activity in terms of Shannon's definition of information make us seem more entrapped than we really are? Neil Lawrence replied that he is just trying to quantify, and thus offer a useful comparison to computers—against our unhelpful tendency to anthropomorphise computers.

He noted also that our immune system may qualify as a sophisticated intelligence system more akin to what is being developed in ML. And while our conscious minds are pretty bad at handling high bandwidth, the unconscious mind is always taking in high bandwidth stuff like vision, then feeding it selectively and usefully into the low bandwidth part of our heads. If we ever were to connect ourselves to the internet, it would need to be in a way that our low bandwidth consciousness could make sense of.

Professor Eileen Hunt Botting, Notre Dame

Apocalyptic Fictions: Mary Shelley and the Revelation of the Singularity

Mary Shelley initiated an Apocalyptic strand of modern political science fiction with her novels *Frankenstein* (1818) and *The Last Man* (1826).

This political strain of science fiction conceptualises apocalypse in terms of humanity's paradoxically selfish yet self-destructive relationship to its whole environment.

Using a term coined by Romanticist Morton D. Paley, we might call the philosophical approach of Mary Shelley "apocapolitical." In a realist vein, Mary Shelley's "apocapolitics" resists the naïveté of Romantic responses to the possibility of the end of the world.

First, Shelley explodes the notion that apocalypse can be avoided through a revival of a natural or innocent state. Her self-described "Last Man," Lionel Verney, writes down his story alone in Rome—calling it the History of the Last Man—and leaves it behind on a tombstone in hope that someone, anyone might find it. He boards a ship to cross the 'seedless ocean' in search of other survivors of the plague, taking for companions his adopted mutt and the works of Shakespeare and Homer: signs of solidarity with the nonhuman beings and things, and artefacts of human love and intelligence.

Shelley's futuristic revelation in The Last Man offers an antidote to looming fears of artificial intelligence and other supposed technological catastrophes. The daughter of Wollstonecraft, the leading feminist political philosopher of her time, she built a recipe not for man-made disaster but rather for coping with it. Lionel Verney makes a believable, life-preserving choice: he leaves a record of his past behind in the hope that others might discover and learn from it, at the same time that he takes his storytelling powers on a journey around the world to discover unknown others.

Shelley knew from the myth of Prometheus that apocalypse was not a true end, but rather a trial by fire. As an apocalyptic novel that explodes the very fear of apocalypse, *The Last Man* has proven to be a productive formula for rational response not irrational capitulation.

In *Frankenstein*, similarly, Shelley teaches her readers how to sympathise and identify with artificial intelligence.

As with the Creature, AI is not born from a womb, but it is still made by circumstances. The Creature is a learning machine who analyses the input of the DeLacey family through the constraints of the program of the hovel.

Since the real world is the world of trial and error, AIs—much like the Creature—may be capable of learning deeply but not well. AIs both learn and mislearn through storytelling. If its programming is faulty, a computer will not process data correctly. If its data is bad, it will produce a false analysis. Sometimes the problem is neither with the data or the programming, but rather the AI's lack of experience with a particular pattern of data.

The Creature understood that the DeLaceys were a family, but, due to his lack of social experience, he did not grasp that they lacked the emotional ties to him that he felt so dearly toward them. His tragedy is that he misapplied the model of a happy family to explain his relationship to them.

Put differently, the Creature is the specter of the singularity. He represents the anticipated moment when technologies eclipse the creative and destructive capabilities of their artificers.

Alan Turing grappled with the question of how to effectively and ethically reach an equivalence between human and computer intelligence, without unleashing destruction. He pictured one of the super-sized computers of his time roaming the countryside as a remote-controlled monster, abandoned to its own devices, both to the detriment of society and its own intelligence.

From this worst-case scenario, he reasoned that one should not take the "whole man" approach to building intelligent machinery but one would rather have to train a learning machine to grow up, mentally, like a child.

Turing ultimately coincided with Mary Shelley's ethical understanding of artificial intelligence. Intelligent machinery would best emerge within a long-term, loving, morally instructive relationship between a parent and a child.

When one could no longer tell the difference between the machine and the human, the reason would be the sharing of parent-child love between them. Once the child became a machine, the machine would become human.

Discussion

Beth Singler asked about the implications of the multiplicity of definitions for the concept of the Singularity, to which Eileen Hunt Botting replied that these many readings suggest why it is such a powerful mythological concept for our time. It is fundamentally allegorical and can be read on many levels—and it is similarly important to think about AI in multiple ways, alongside how we theorise intelligence in general.

Yorick Wilks suggested that, to add to the list of the ways in which Frankenstein's creature/monster can be seen as a learning machine, emotion is now a hot topic in AI having once been seen as a dirty word. The monster is full of emotion, and was

in this sense ahead of the game. Eileen Hunt Botting agreed, noting that people used never to talk about emotion in her field of political theory, and now everyone is interested in affect.

Thinkers in Britain at the turn of the 19th century got this right, she continued: the importance of interaction between passion and reason when it comes to intelligence of all forms. The Creature is a political theorist demanding his rights. At a revolutionary moment in the history of political philosophy, the creature is in many ways speaking a very innovative political philosophy.

Dominic Burbidge noted the gulf between *Frankenstein* and later ideas in Science Fiction, in terms of the permanence of suffering in *Frankenstein*. What thoughts does this provoke around the idea that AI will only be properly intelligent is if it has a self-recognised limitation, a form of death?

Eileen Hunt Botting replied that both *Frankenstein* and *The Last Man* have open endings. We never see the creature immolate himself, with Shelley leaving it unresolved as to whether the creature is mortal or immortal. It certainly suffers, but may simply be more enduring than human beings.

Both the creature and the protagonist of *The Last Man* are in situations that most people could not endure: worst-case scenarios that get you as a reader to engage in a thought experiment about what your attitude ought to be if you found yourself in a position of almost unbearable pain and suffering. Mary Shelley herself endured tremendous suffering in her life: she lost several children, and her husband; her mother died giving birth to her. Yet she never gave up. It's about confronting suffering and finding a way to move on, not only for yourself but perhaps for the sake of the world.

Rory Doyle brought in the relational and emotional side of AI from a religious perspective. From a Christian perspective, the most profound New Testament description of God is love: the intelligence of the Father, the word of the Son, the love of the Spirit. When relating to people from a religious perspective, there is always this extra element to the human being: the spiritual. A robot may be extremely good at mimicking human communication and interaction, but from the religious perspective we can never see it actually giving love.

Madeline Drake was reminded of the fact that emotions are the products of the endocrine system interacting with the nervous system. For us to create an AI of the level we are talking about, there would have to be embodiment including something like the cardiovascular, endocrine and nervous systems.

Session 6

Revd Dr Malcolm Brown, Director of Mission and Public Affairs, Church of England

Differentiating the human person from any mechanical or technological creation

Earlier this year, I was a participant in a conference on AI Ethics at the University of Lund in Sweden where another contributor was Harald Hamrell, the Director of a Swedish TV series called Real Humans. The series was set in a contemporary Sweden where humanoid robots do almost all menial work.

One of the moral questions it raises was the episode in which a humanoid robot (a "hubot") was assigned to care for a young child whose parents died and called in their will for the hubot to be given custody of the child. In the ensuing court case, the advocate notes that, "because we don't really know how to define a human, we don't know how to distinguish the issues in this case."

There are two running themes in the series. One is about the social reaction to the way human functions are taken over by hubots, with a guerrilla movement of dispossessed humans fighting back. And a second is the way that some hubots begin to develop a moral sense, including one causing mayhem in church by wanting to become a Christian.

Any sensible answer to a question like "can a robot be saved?" demands a robust account of being human.

In Real Humans, the question arises constantly whether a machine, however human in its behaviour, can have rights, or if it is just property. When displaced manual workers fight back by taking out their anger physically on the hubot, are they behaving more humanly or less so?

That is an extension of the immediate question whether we should teach children to say please and thank you to Alexa. If they do, might there be a risk that they could lose the sense of boundary between human and machine? If they don't, are they likely to grow up as the kind of people who never say please or thank you to anyone?

The point, surely, is that the measure of our humanity is not just how we treat people like us, but how we envisage ourselves in terms of the whole created order.

The speed of current change is much faster than human ways of relating can cope with—including structural, economic and political institutions, and proper social and ethical scrutiny. In particular, our encounters with other people are changing now that geographical remoteness is no barrier to intimacy.

A group in the Church of England working on mission theology has adopted the term "Excarnation" for this phenomenon. If Incarnation is about inhabiting our bodily selves and relating to one another as embodied persons, the trend toward remote relationships driven by technology feels like the opposite – hence: Excarnation.

We have gained the ability to control our relationships without paying the kind of attention to the other person that we have been used to.

But God reaches out to us in Christ who comes among us as one of us. He takes our nature upon him in order to redeem it. That short intervention in time of Christ's earthly life is our promise that God saves us as whole beings.

If our embodiment is becoming irrelevant to the ways we relate to other human beings, how will we understand the Incarnation? What, in the end, happens to our doctrine—or even just our human understanding—of our bodily selves?

Even though many specialists tell us that AI is not really very intelligent at all, there are areas where we are finding the human to be deficient. That, perhaps, feels more disturbing than it should because we have allowed intelligence to become unreasonably central to our concept of being human.

The classic Christian grammar of being human does not make intelligence the measure. The image of God in which all are created embraces the unintelligent, the mentally incapacitated, the old person with dementia, the Downs Syndrome child.

AI is not, I think, unleashing something new and terrible upon us, but facilitating the entrenchment of something we already know about ourselves which is old and terrible: our capacity to enslave and monetise our brothers and sisters.

But if we can offer a confident theology of being human that does not measure our humanity according to intelligence, we should have less to fear from artificial forms of intelligence that do some things better than we can.

They may threaten our jobs and our economy, and even our political stability, but forms of AI will not be able to take away, or even erode, our humanity before God—unless we let them.

Discussion

John Cornwell suggested that one problem for the younger generation may be a point raised by Richard Dawkins, that it is due to world religions we have so much violence down the centuries. It is after all only recently that Christianity seems to have calmed down its sacrificial and unpleasant side.

Plainly, Malcolm Brown replied, this is the problem of having a long history, common to most worldviews that have a long history. Once faith got entangled with power, it look a long time to disentangle it. Today, one might describe the Church of England as feeling it has responsibility for the world but is too weak to throw itself around: a weak establishment.

Ultimately, he continued, this is about actions that people take that are way above levels of transactional justice of any kind. Mercy is a virtue that can't easily be digitised in any form. The problem is, if we forgot that these virtues are part of being human, we will be much more susceptible to intimidation by AI and other developments; while if we can hold onto the fact that we do things differently because we are human, we have a chance

Yorick Wilks raised the question of whether and how AI's rise has influenced views of the human person. Even people who write programmes don't quite know why machines do the things they do today. Could you build a programme to explain why an ML programme did what it did? This question of the suitability of ML has helped strengthen the view that we don't really understand why we do the things we do: that we tell ourselves stories but don't really know why. What's needed is ways of explaining both AI and human behaviour in ways we can understand.

Malcolm Brown suggested that what's missing from this account is the question of responsibility. The insurance industry is quite worried about some aspects of AI-driven developments because of the problem of saying who is responsible when hundreds of people have their fingerprints somewhere on an algorithm. Even if we don't know why we act, we do know we are responsible. With AI this is hard to locate, because it must come down to a human agent of some kind. You said that we tell ourselves stories: but this is exactly how we know ourselves, through narrative rather than logic

Sumit Paul-Choudhury noted that "excarnation" in archaeology is the de-fleshing of the body after death, leaving only the bones, as seen in some traditions like Zoroastrianism. Other traditions have different views about the integrity of the body and what it means to be human in a physical sense. Given the previous discussions of the complexity of humans as biological entities, it's worth noting that the integrity of the whole body may be a concern for Judaeo-Christian thinking in a way it isn't for other belief systems

Malcolm Brown replied that he worries about a too mechanical metaphor dominating these explorations—and that a lack of awareness of how we use a mechanical metaphor to describe ourselves is part of the problem.

Dr Gorazd Andrejč, St Edmund's, Woolf Institute, Cambridge

How can Christian, pantheist and animist theologies or philosophies respond to the posthumanist challenge to humanism?

In his famous 1950 article "Computing machinery and intelligence", Turing proposed "can machines think?" as a serious question. We now use a wide range of mental and perceptual verbs to describe machine actions, something that Turing only proposed as a consideration.

Such descriptions can become commonplace—*learn, predict, understand, know, differentiate, think*—and then become more complicated—*reason, decide, prefer, hesitate, perceive, see, hear, speak*—while at the same time ceasing to be seen as problematic.

The fact that this range of mental terms increases over time, optimists would say, confirms the hope of strong AI: that we are on the way, and that the complete set of human mental language may come to be used for machines.

This is a trend of particular interest to religions, which generally think that that language matters strongly, and that it carries moral and thus political significance, even if indirectly. So, some key questions are:

- When, how and why do we decide to extend mental vocabulary to non-humans?
- Who decides, who starts to circulate certain words?
- Should we listen to and follow such influences?

A standard answer to the first question is that it happens when a non-human being behaves similarly enough to humans, or is similar enough. However, judgements of similarity are not delivered by world: they are approximations in which we pick out what is relevant and discard dissimilarities.

What constitutes a relevant similarity? One answer is that similarity of the processes needs to be not only clear on the symbolic and abstract level of information processing, but also in some kind of embodiment. Posthumanism and transhumanism entail some useful reflections upon this.

Transhumanists, to simplify, see the human body with its frailty and imperfections as a major obstacle to our progress, or rather the progress of intelligence or mind or the world. Hence some suggest upload onto radically different hardware, with their ultimate goal a virtual immortality.

Posthumanists, especially feminist posthumanists, offer a different vision, rejecting the anti-body attitude and building around the idea of the cyborg: a hybrid of machine and organism. They see the acceptance of human-machine bodies as a

way to a better, fairer and more flourishing future society, including better acceptance of non-humans: animals and robots.

Drawing inspiration from poststructuralist philosophy and feminist critiques of enlightenment, posthumanists reject an anthropocentric focus on man as the centre of ethical, cultural and civilizational concern. Human exceptionalism is seen as always unwarranted and part of a project of the subjugation of nature—one that has excluded not only the non-human but also those humans who do not fit.

It's interesting to compares some religious views with posthumanism. In particular, Christian eco-theology, or so-called creaturely theology.

Creaturely theology tries to reinterpret the concept of the image of God, make it slightly more porous and inclusive of non-human beings—and it does this by decoupling the concept from the human capacity for abstract reasoning or intellect. Instead, it starts by understanding humans first and foremost as creatures among other creatures, and by defying any strong separation between us and other beings.

This shares much of posthumanism's critique of anthropocentrism: decentering humanity is a common ethical goal, together with a strong emphasis on the body.

But there is also a crucial differences. Theology sees us as created beings, and this does not easily translate to the work of human hands: our obsession with tech can be seen as part of human-centric culture, overlooking nature.

Finally, we come to animism, in terms of seeing spirit and indeed divinity in the non-human and non-living world. This entails strong a critique of anthropocentrism that reads similarly to posthumanism's—and may go all the way in talking not only about animals and plants but also the non-living as having a spiritual nature, and needing to be included in some sense in the community of morally relevant beings.

Discussion

Yorick Wilks brought up the intriguing richness of common law around dogs, and its potential relevance to AI. Dogs are not deemed "wild" animals, because they are considered to have such a thing as character—so, if they bite someone but are not previously known to have bad character, you may not be in trouble as their owner. Is this a way to smuggle in law around robots?

Gorazd Andrejč answered that this ties in well to a recent article he read about Istanbul and its history of relating to dogs in that city through its regulations, which to a degree see gods as an independent part of the with certain obligations owed to them. This is an interesting proposal for thinking about our relationship to machines.

Eileen Hunt Botting appreciatively registered the influence of Donna Haraway's work on the talk, and noted her resistance to the label "posthuman", because by focusing on the prefix "post-" we risk of losing sight of the human.

Gorazd Andrejč replied that it's almost impossible for us not to be humanist in some sense, and not to prioritise humanity in some ways over other beings in the moral universe—and he agrees that disregarding the human is an important danger, and behind it sometimes is an overhasty hate for humanity.

David Mellor suggested that the problem of labels like the Anthropocene is that they simultaenously raise the problem of the human while at the same time centring the geological era around humanity: there is a conflict between decentering and recntering brought simply by using this name.

It does sound very self-important, replied Gorazd Andrejč, that we have named a geological era in this way. But it is the geologists who have suggested it. Perhaps philosophers should trust them, given that you can have detached measurements of human impacts on the Earth—although it is a little suspicious when a big concept becomes very fashionable.

John Wyatt Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology, University College London; Senior Researcher, Faraday Institute, Cambridge

The Ethical Implications of Simulated Personhood

I have a background in a neonatal intensive care unit and my interest is informed by caring for frequently very sick and damaged and dying babies.

Is a brain-damaged 23-week baby a person, and what do we mean by that? Is it a pre-person, a potential person?

Consider the story of two tech entrepreneurs who had an intense friendship via thousands of text messages, one of whom was killed in a road accident. The other then used messages sent by her deceased friend to create a bot she found herself communicating with in an honest and intense way. This led to the company Replika, "the AI companion who cares"—a personal AI designed to "help you express and witness yourself…"

This is real, not the future; and it's a very significant development, especially in healthcare, where there's massive interest in digital assistants and therapists. Just one example: the Woebot app, for people with depression. Ready to listen 24/7, it promises strategies to improve your mood: daily conversations via smartphone as part of a mental health therapy process.

This creates potential confusions. Siri offers canned answers to questions like "are you a computer?" But what about someone with Alzheimer's or learning difficulties, or a child on the autistic spectrum: would they take it at face value?

The NHS seems convinced chatbots are going to play a major role including in primary care. The idea is for access in future via smartphones to diagnostic and therapeutic and counselling services.

There are huge commercial pressures for this in a cash-strapped system. Tech never gets bored or tired; it's constantly learning, and it can be scaled up for everybody.

But there's also a blurring between what is real and simulated that extends beyond healthcare. Take deepfake videos. Within the foreseeable future we may see real world conflicts between states based on these videos.

In the field of elder care, meanwhile, we see cute robots designed to make us anthropomorphise them. And we seem to involuntarily and automatically to empathise with humanoid robots when they are submitted to "pain". This empathy is not under our conscious control—so we need to confront the moral implications of having robots in our midst that we anthropomorphise.

Strong criminal laws exist against torturing defenceless animals. Is this because we as a society believe that animal rights are very important—or more because if this a human tortures a kitten, this says something about them that we as a society wish to sanction? Should we enable a human to act out rape and abuse with a humanoid robot—not because of the robot, but because of what it is doing to us?

Such questions are not far off. And it's children in particular we need to focus on. Our unique ability to anthropomorphise is part of our humanity—but it also renders us open to manipulation and deception. And of course there are commercial forces that want us to do this. The tech ethicist David Polgar comments: "Human compassion can be gamed. It is the ultimate psychological hack; a glitch in human response that can be exploited in an attempt to make a sticky product."

In the TV series Westworld there is a very significant scene where a beautiful robot comes up to man and asks is there anything she can do for him. He asks if she's real, and she replies: "if you can't tell, does it matter?"

It's a profound question. I want to say yes: that, however effective the simulation, it is not the same as reality. Those of us from religious traditions tend to focus on the fundamental nature of being, the ontology. But the real questions we are confronting are epistemological: how can we tell?

Toby Walsh's suggested *Turing red flag law* is that an autonomous system should be designed so that is unlikely to be mistaken for anything else. I wonder if this is something that ought to be highlighted much more. Could we in faith communities jointly press for discussions as to whether the UK should have a red flag law?

As a medic, I want to push back against the idea that machines can provide care. "Care-bot" is ultimately a misnomer. The privilege of caring is the privilege of human-to-human solidarity. I am human like you, I will walk this path with you and offer you expertise and experience—this is something that only one human being can say to another, and that is something very precious.

In Trinitarian theology, personhood is a category of reality that is ontologically fundamental. Persons are not reducible to matter and energy. To be a person is to participate in some sense in the divine.

Through this, you can perhaps invert the discussion of monism and dualism and say—the most foundational things of all are persons.

Discussion

Ezra Sullivan suggested that a deeper question about interactions with AI might go to the distinction between the categories of slave and servant. A slave can be used and disposed of, while a servant is another human being. Are robots slave or servants? John Wyatt replied that if we are saying a servant is a person equal at a

fundamental level, though in a subservient relationship, this can only happen between humans. A machine servant is a metaphor that can be very misleading.

Madeleine Drake commented that, as a retired housing association CEO and care provider, she has in her running of care providers had to develop the use of AI to enable people to be cared for in their own homes. And she very strongly would argue that AI is an important element of developing a much more responsive, individualised, personalised and flexible response to people with care needs.

For example, she continued, taking people out of a psychiatric hospital where they sat in a catatonic state to living in their own flats: people who on the ward wouldn't talk would now rush up and offer visitors a cup of tea, their lives transformed. We are obviously not talking about caring for people who are dying or neonatal families when we talk about AI. But there can be care at different levels, and this needs to be looked at as a complex proposition.

John Wyatt responded that he was trying to discuss whether simulation is being misunderstood. There is lots of evidence around using dolls with people with dementia. Is there a difference between something that is clearly a doll rather than something that looks as though it might be alive? Arguably a doll is better because it does not confuse a person with dementia as to whether it is alive.

Neil Lawrence suggested that the red flag approach is a famous example of a simple solution to a complex problem. A person holding a red flag literally had to walk in front of the very first powered vehicles in the Victorian era: it limited development, didn't help then—and wouldn't now.

Session 7

Dr. Simone Schnall, Reader in Experimental Social Psychology, University of Cambridge and Fellow of Jesus College

From Moral Intuitions to Moral Machines: How Rational is Ethical Decision Making?

The anthropologist Robin Dunbar proposed that the reason we have such large brains is because we are a social species that has to keep track of social relationships. And one reason we have language is not just relationships but also moral reputations: it's a tool for gossip, talking about others.

Recent evidence around how information spreads on social media looked into tweet content and suggests that, the more moral language tweets have, the more they gather attention. In Brady, W. J., Wills, J. A., Jost, J., Tucker, J., Van Bavel, J., & Fiske, S. T. (2017) the top ten words were: *attack, bad, blame, care, destroy, fight, hate, kill, murder, peace,*

How do we tell right from wrong? There are two broad philosophical camps: rational camp along the lines of Kant and universal moral laws; and Hume's morality based on sentiments.

For a long time the prevailing psychological approach followed Kant and looked at how reasoning unfolded, for example in children. But in 2001, Jonathan Haidt published an influential paper about Moral Intuitions, proposing that these are the driving force for moral judgement, and reasoning comes in after the fact, justifying it to ourselves and others.

In my research with Haidt, we have become specifically interested in disgust. It's a very strong, embodied emotion, that has evolved in the context of food consumption. It's also an existential emotion, in that something you find disgusting might kill you.

Experimentally, if we assume moral judgements involve disgust, we should be able to induce it and look at its effect on judgement. So a collaborator came up with idea of inducing with a smell: in particular, fart spray.

We set up a study into "Disgust as Embodied Moral Judgement" – Schnall, Haidt, Clore & Jordan (2008) – in which fart spray was used on rubbish bins near where people were sitting, or no spray in control cases. Subjects were then presented with moral vignettes: scenarios in which someone did or did not return a lost wallet, falsified their CV to get a job, and so on.

They were asked to rate the scenarios from perfectly okay at zero to extremely wrong at nine. We looked at the effect, if any, of manipulating disgust. And we

found that moral judgements were more harsh in the presence of a bad smell: about twice as bad.

Replications having involved situations such as a dirty desk and recalling disgusting events. And these have needed to differentiate disgust from other negative emotions, like sadness, which do not seem to have the same effect.

Lots of different manipulations have happened since—as well as looking in the opposite direction, as to how far people engaging with objectionable content show a disgust response. And finally, how individual differences in disgust sensitivity in general correlate with moral outcomes. In general, the more easily people are physically disgusted, the more easily they express moral condemnation.

So it seems that Hume comes out ahead of Kant. And one can link this to dual process theories, the most popular and well-known being Kahneman's *Thinking, Fast and Slow*. We have two ways of processing information: one fast and low effort and one slower that is tiring. And our suggestion is that moral judgements fall in the domain of these unconscious quick decisions.

Where does AI come in? The famous trolley problem is an easy question in some ways, but it becomes very different if you bring in an emotional component: instead of a switch, for example, a fat person on top of a foot bridge. Such scenarios are artificial, but they do allow you to manipulate aspects of that scenario.

Most recently, Edmond Awad at MIT set up in 2018 "The Moral Machine experiment" which consisted of a game in which participants made decisions for an autonomous vehicle under many different conditions. Close to 40 million games were played online, creating data suggesting, for example, that players preferred to save (in descending order of strength of differential preference) humans over pets, more over fewer, young over old, higher over lower status, fit over large, females over males, pedestrians over passengers, inaction over action.

Critiques claim that these dilemmas are meaningless outside philosophical discourse. But surely the data tells us something. It raises an interesting question about how we make moral decisions, and how these are often related to bodily sensations as opposed to rational considerations; and it may have practical implications for legislation around autonomous vehicles and AI.

Discussion

Fenella Cannell noted the absence of context in thought experiments. Why don't they wave at the driver? Given the complexities of how we train ourselves to overcome disgust, to modify our own behaviours, it is hard to know how the experiments actually map onto life.

Simone Schnall replied that experiments always have to boil things down to the most basic components. She is not personally a fan of trolley problems, but they do give a very clear method to break down key variables.

Yes, said Fenella Cannell, but the moral answer is that we do the very best we can. That might have very different outcomes, but we don't know what these are. If we can't solve these problems, maybe we don't have self-driving cars.

Malcolm Brown suggested that presenting a binary choice between two 18th century philosophers makes things harder than they need to be: what about Aristotle, MacIntyre and virtues? The question is, what sort of person do I want to be? What is assumed about the kind of person you are when you get into a self-driving car?

Fenella Cannell suggested in response that, if you take charge of a situation where others could be harmed, then maybe you do indeed have to be the sort of person who takes full responsibility. She also mentioned work on moral elevation: the feeling of being uplifted and wanting to become a good person. Experimentally, when people witness this, they engage in more such behaviour themselves.

Rory Doyle pointed out that a scenario in which a friend submits a false CV and deprives someone else of a job is different to a scenario in which a random person does this and deprives your friend of a job. With a lot of morality we like to think we would do the right thing, but it depends on our emotions at the time. So maybe AI could make better judgements than us, sometimes.

Dr Daniel de Haan, Oxford University

A.I. and Ethics 2.0

AI operationalises basic human values, and we need ethicists to combine interdisciplinary approaches. But…

- Whose basic human values are we talking about?
- Which ethicists do we hire?

Humans are disagreeing animals: this is the starting point. We even disagree about what we disagree about. In a sense this is also the conclusion.

If we cannot have some agreement about our disagreement, we cannot resolve other more substantive disagreements. Microsoft, for example, has attempted to list deeply rooted and timeless values in its principles of ethics: fairness, inclusiveness, reliability and safety, transparency and safety, privacy and security, accountability.

But all of these have a variety of ways of defining them. So we need to be able to articulate our differences around these slogans.

Are these values timeless? Are they universally agreed upon? And even if they are, are they mutually compatible? For example, even if we can universally agree that liberty and equality are basic human values (which is itself contestable), the exercise of liberty often diminishes equality.

A common rhetorical tactic in political liberalism in the philosophical sense is the thought that there are universal and compatible human values and human rights. On the other hand, defenders of liberalism like Bernard Williams and Isaiah Berlin have argued that basic human values are fundamentally incompatible—that liberalism is about incompatible values—but that this entails not relativism but plurality.

So: which ethicist do we hire? A theoretician from the academy? Or someone with practical expertise in human history, conflict resolution, mediation, humanitarian efforts, emergency relief situations, tech work? Moreover, who are we educating to possess all of these proficiencies: are there even individuals sufficiently competent to be hired?

A brief narrative history of ethical debates in the west may be useful.

We begin with Virtues, Common Goods, and God: with a common conception of the Book of Nature and Book of Revelation both coming from God and thus being harmonisable through philosophical reflection on what humans and their goods are, alongside a revelation of God. Ethics tends to be rooted in reason in this account, but also in religious and social norms.

Then there's the Enlightenment and Scientific Revolution, which led to a rejection of a lot of these views, although they also retained the ambitious rational pursuit of articulating a universal human morality. In place of divine authority came an attempt to find a rational justification for ethics, without authority or revelation. But this Enlightenment project was unable to establish universal rational consensus on morality, and instead generated a host of rival and incompatible views in ethics.

This project of rationally justifying morality has largely failed. This is because there's a mutual incompatibility between rival rationalist ethics which makes them inconsistent by their own criterion: there's no universal agreement among rational agents about the rational principles that can justify ethics. There are no rational, obvious views that every rational being will agree to.

Is there then a psychological explanation: are seemingly rational justifications of morality actually confabulations of social and psychological and evolutionary factors? Is ethical theorising just one more mask for rival wills to power, with human rights and basic values simply incompatible expressions of what those with power are willing to enforce?

If ethical principles are to be true, is the alternative God or nothingness? That's a disagreeable disjunctive for many.

On the opening questions—Whose basic values? Which ethicists?—we seem likely simply to go on having shriller and shriller disagreements.

Ought AI ethics to be true, or simply sufficient for survival—and if so, whose survival? Or should they just be the norms that we have power to enforce? These are inescapable questions that we must have the courage and the critical reflection to ask and seek the truths that answer them.

Discussion

Eileen Hunt Botting mentioned Rawls's principle of starting from pluralism in order to achieve pragmatic consensus, to which Daniel de Haan replied that he thinks Rawls fails as part of the Enlightenment project, and is guilty of making up fake human beings behind his hypothetical "veil of ignorance."

Steve Torrance mentioned the recent Extinction Rebellion protestors. They may be too innocent to know how difficult it all is: but to say either we do God or do nothing, does that then mean they should all go home? Are you not asking for more precision than the subject matter admits? What about an ethical pluralism in which the ways we want the world to be and humans to act can simply be mutually irreconcilable in particular situations. Sometimes we can have dilemmas that are real buggers to resolve, or are impossible, and the only response is one of remorse.

Jeffrey Bishop suggested that the problem is, we don't know how to talk about good and evil, and are instead talking about right and wrong in a way that is legalistic and calculating. So what happens in science and tech is that we have a deontological view that creates the boundaries of who is in and out—and then within that boundary there are utilitarian calculuses that we have to perform. But virtue is about good, and God is a transcendental idea.

Malcolm Brown said that one thing that chilled him was the talk of hiring ethicists: a hugely problematic relationship of ownership in the context of AI. His job entails working in a church where theology and ethics are as contested as they ever were—and he is constantly saying that you cannot hire a theologian to tell you what to do. The contested nature of the problem has to be exposed in the context of a multi-disciplinary argument in which participants can overhear each others' disciplinary conversations. And the loss of that interdisciplinarity in the academy is a tragic thing.

Session 8

Dr Ron Chrisley, Dept of Cognitive Science, University of Sussex

Creation and Responsibility: The Paradox of Ethical Artificial Intelligence

The Nicene Creed in 325 made a distinction familiar to every Christian:

"…Begotten not made
Of one substance with the Father…"

The key line makes this distinction between begotten and made, which describes two different ways that X might bring intelligent Y into being.

In Greek, idea of *genesis* and Latin *natum* – "begotten"
In Greek, idea of *poiesis* and Latin *factum* – "made"

When X begets Y, X is made of the same stuff ("one substance") as Y. That's not the case with making.

Begetting: God the Father begets the Son
Making: God the Father makes humanity

We beget our children and they are consubstantial with us; but the creation of AI is presumably in the making column.

These notions shed light on the question of responsibility.

Where X is divine: is God the Father responsible for the Son? We have no idea. But in general, we ourselves are responsible for our own actions, despite the fact that we were made by God.

Where X is not divine: eventually we cease to be responsible for our children and they become responsible for themselves.
But what about our artefacts?

In general we are responsible for them. But what about intelligent artefacts? There is something problematic here, because if Y truly is an intelligent agent it should be responsible for its own actions—but there is a tendency to think that, because we made Y, we remain responsible for it.

This is the key tension: the paradox of making intelligence responsibly.

- If Y is a fully intelligent agent, then Y is responsible for its own actions.
- For X to make Y responsibly, X must have reasonably complete foreknowledge of how Y will behave.

- If this is not the case, it is irresponsible for X to bring Y into existence.
- Humans can make AI either by design (Good Old-Fashioned AI) or by adaptive methods (Machine Learning)
- By design: it's possible to have reasonably complete foreknowledge
- By adaptivity: it's not possible to have such foreknowledge
- The problem for design is that, in so far as it is ethical, it cannot truly be artificial intelligence. We have foreknowledge, so it is just an extension of our will
- The problem for adaptive machine learning is that, insofar as Y is truly free of our intentions, then it is truly AI—but thus not ethical
- So, making true AI ethically is impossible.

This is like a laboriously stated version of theodicy, the problem of evil. So how does God avoid this paradox?

Predictable design is not an option for God, on pain of usurping our authority. So God's only option is adaptivity in the form of evolution. This yields the possible answer of "limited omniscience", in that God has enough knowledge to know that creating us in this adaptive hands-off way would be a good result, even though evil is done; but not so much knowledge that God foreknows and is thus responsible for each of our individual actions.

This resolves the paradox for God without doing so for us, since we lack enough knowledge to know that creating is a good act. Unlike God, we cannot be sure that creating any AI that can adapt free from our will is better than not doing so.

One final possibility. If we mortals can't make true AI ethically, perhaps we could beget it, in the sense of being responsible enough for its upbringing to fulfil our ethical duties while enabling it to fulfil their own?

This is only possible if there is some way to manage (if not eliminate) the risk of extreme catastrophe and existential threat.

Ultimately, if we are begetting AIs, they will not be our slaves, instruments or means to our ends. Rather they will be as equals, as children are: things that start out in an asymmetric relationship with us but that eventually become a full person. We would best be thought of as midwives or God-parents to a new kind of agent.

Discussion

Tudor Jenkins asked, if we are begetting, do we not simply run into a totally different ethical problem about releasing these creatures into full responsible adulthood. When does our responsibility cease? And what about killing them?

Ron Chrisley replied that there is uncertainty about whether we can inculcate our values, and whether our offspring may go on to make their own offspring. That is part of the risk that we need to think about. You might have to kill the creature if you see that catastrophe would result.

Eileen Hunt Botting noted that this is exactly the counterfactual scenario Victor Frankenstein hypothesizes in the novel: if he makes a female creature and they reproduce, they might make a race of monsters that usurp humanity. She believes his reasoning is poor—but is most interested in the distinction between begetting and making. If we apply a feminist lens, do we want to question the language of begetting as a patriarchal linear descent through time justifying patriarchal authorities. Shouldn't we be using a language of gestating, which is a better metaphor, rooted in the concept of genesis?

John Wilkins asked whether one deduction is that you must treat AI as children or servants, or something else, but not as slaves—and that if you do this last, as we currently do, you are on the wrong track ethically. Ron Chrisley replied that he doesn't think it follows logically, but would subscribe to it anyway. There is no current AI that could be a slave because there is no AI whose agency could be impugned by constraint: but it is a horrible relationship when it is actual, so why would we want to use it as a model?

John Wyatt noted that this is already a big issue in healthcare: how do we certify a system involved in diagnosis as safe? He worked in a medical school that trained and certified doctors, most of whom ended up reasonably ethical, and there is some kind of analogy here. We make systems and we try to design them so they behave in adaptive ways, and we recognise that occasionally they may get it wrong. So perhaps we can make a system that is good enough to be released into the wild.

Ron Chrisley noted that the tension he was trying to create was that, if we really did know how something would behave in a wide range of contexts, it would then just represent an extension of our knowledge and abilities and will. What about an autonomous car, responded John Wyatt? It can behave creatively, it's not determined; but we are (hopefully) still able to say that as far as we can tell it will get things right.

Steve Torrance, Visiting Senior Research Fellow, COGS, University of Sussex; Professor Emeritus of Cognitive Science, Middlesex University

AI messianism and our grandchildren's grandchildren

What can the Shearwater seabird teach humans about "having domination over the fish of the sea and the fowl of the air" (Genesis 1:28)?

Up to 90% of seabirds have plastic in their guts. At a seabird's dissection in Tasmania 250 fragments of plastic were found in its gut.

The history of plastics is an interesting model for other technologies. It started in 1866 with Alexander Parkes, who began a line of fascinating innovations leading in 1907 to Leo Baekeland's invention of Bakelite and then, from the 1940s onwards, to the mass production of plastics.

The product cycle model we can abstract from this is one that starts with fascinating and exciting experiments in a lab; then creates products that are brought to market, and then mass marketed if they are successful. They proliferate; and then various unforeseen outcomes raise questions of responsibility and reversal—and of public consent.

In one sense, we consent to the proliferation of plastics with our wallets, because we buy them—as we do with our clicks to the products of AI. But there wasn't a moment at which people asked if we truly wanted the proliferation of such a product and its effects.

In a wider sense, is AI and its companion technologies part of the eco-crisis? The spread of mechanised intelligence in AI and its companion technologies is accelerating, especially in the last few years, and I wonder and worry whether this acceleration has been considered alongside climate change, biodiversity loss, food crises and the coming water crisis as a global jeopardy factor alongside its enormous benefits.

Question is thus: is the planet over-thinking as well as over-heating? Is there a global over-production of artificial intelligence?

There is a kind of messianism around AI and the idea of the Singularity: a tradition of rhetoric with quasi-theological overtones that resonates through the AI community. A lot of people have adopted elements of this approach saying, if it's possible to do it, then let's do it.

There's also a kind of a naive intelligentism in terms of the focus on being human as something that is all about intelligence—and thus the reasoning that AI research is inherently justified, together with making a kind of smash and grab raid on human civilisation through it. What's missing is any readiness so ask for consent or to recognise the possibility of some kind of democratic discussion or participation.

Even though our tech is not currently controlling us like a super being with clear goals, humanity is rapidly losing its ability to make technologies conform to our basic goals and needs as a species. The Singularity represents both a kind of cataclysm and a kind of continuity—because many of its worrying features have already happened, or are in the process of happening and expanding and accelerating:

- Performance gaps
- Loss of effective governance of AI and digital processes
- Lack of informed public consent
- Heritage loss

For these various reasons, we should consider AI and its companion technologies an ecosystem problem alongside global warming, loss of biodiversity and other such threat factors – and become more aware and more co-ordinated about trying to control its excesses.

Discussion

Gorazd Andrejč wondered about public control and consent, and whether we are even capable of having the necessary deliberation. There is an argument that things are so complex and so plural in their development that consent is hard to achieve. What ideas are out there on this front?

Steve Torrance replied that there already are expert groups cropping up, such as in Europe where the European Union have created a high level expert group on AI who just produced a report, with Luciano Floridi as one of its leading members. And there are similar groups, in Montreal and many other places, as well as organisations like the Future of Humanity Institute, the IEEE, rom which lots of codes are emanating.

So there a question of getting these groups together, and also world leaders in religions, and movements like Extinction Rebellion—because It is the kids whose future we are affecting. What often happens is that we tend to think abstractly about the future. But thinking about our grandchildren's grandchildren takes us about 100 years into the future, and maybe that can help concentrate people's minds rhetorically

Andrew Brown was struck by the notion that the Singularity is already happening. It seems that this stuff does emerge in a manner more like gestation than making: as in the case of recent plane crashes, the software in a sense has a mind of its own. It's enormously complex and difficult to pin down responsibility.

Beth Singler noted that she is doing ethnographic work around the concept of the Singularity, and is interested in people who are already living with the expectation

of the singularity and its impact on their lives, now. The effective altruist movement has strong connections with this kind of thinking.

Steve Torrance replied that, while a lot of transhumanists are on the fringe, nevertheless their ideas and enthusiasms are percolating through. People in the IT profession, mainly guys, tend to be attracted to the kind of future where you transcend ordinary life—and this has a worrying wider influence.

Sam Freed suggested a tension in the argument around the plastic crisis, climate crisis and impending AI crisis. Is all this saying digital life is disconnecting us from direct experience of the world as it was? If so, then so did writing, the printing press, and so on. One the one hand, it's a worry we have worried about a few times before. And is it perhaps a narcissist position to list half a dozen things that are all top priorities: is it in a sense unethical to worry about the ethics of AI when there are bigger fish to fry?

AI – Ethical and Religious Perspectives

AI & the Future of Humanity Series
Science & Human Dimension Project
Jesus College Cambridge
16-17 May 2019

Speaker Biographies

Dr Gorazd Andrejč
Dr Gorazd Andrejč is a Researcher at the Science and Research Centre of Koper, Slovenia, and the Department of Philosophy, Faculty of Arts, University of Maribor, Slovenia. In the UK, he is a Research Associate at the VHI and an Affiliated Lecturer at the Woolf Institute, Cambridge. Gorazd holds a PhD from University of Exeter (Philosophy of Religion), a MSt in the Study of Jewish-Christian Relations from the University of Cambridge, and an MA in Education (Philosophy and Physics) from University of Maribor, Slovenia. His book Wittgenstein and Interreligious Disagreement: Philosophical and Theological Perspectives, a result of a three-year Research Fellowship at the Woolf Institute, was published by Palgrave Macmillan in September 2016.

Revd Dr Malcolm Brown
The Revd Dr Malcolm Brown is Director of Mission and Public Affairs for the Church of England, leading a small team of ethicists who oversee the church's engagement with Parliament, Government and many aspects of public life. Malcolm's early research work was in the field of Christian ethics and market economics. He has been a parish priest in rural and inner city communities, an industrial chaplain and Principal of the Eastern Region Ministry Course in the Cambridge Theological Federation. For ten years, he led the William Temple Foundation in Manchester, a think tank working at the interface of theology, economics and community life in the UK and across Europe. He has taught ethics and practical theology for a number of universities. With others in the MPA team, Malcolm is currently working with Durham and York universities on a major programme to enhance dialogue between theology and science, with a particular focus on AI issues, and he is a member of the partnership with the University of Bath which has recently won major government funding to establish a Centre for Doctoral Training in Accountable, Responsible and Transparent AI. Malcolm lives in Waterbeach where he is part of the parish ministry team.

Dr Fenella Cannell
Dr Fenella Cannell is Reader in Social Anthropology at the LSE. She is a specialist in Southeast Asian anthropology, and has also conducted research on kinship and religion in the United States. She worked in the Philippines in 1988-98, 1992, and 1997. Her fieldwork as with the Catholic rice-farming people in a rural area, but on the outskirts of a small town, where people were also exposed to complex, urbanizing influences and images from Manila and from the West, especially America.

Her research explored the ways in which people come to think about "culture" in a post-colonial society, and focused on women's lives and arranged marriage, spirit-mediumship, saint's cults and religion, and popular performances including transvestite beauty contents. She has since carried out historically-based work on the Philippines, especially on education, kinship, and gender in the American colonial period. She also works with a number of postgraduate students whose research in cased in Indonesia and other parts of Southeast Asia, and intends to more work in the region in the future. Most recently, however she has conducted a two-year research project on American kinship and religion, with a particular focus on Mormonism. Much of this research took place in upstate New York and in Utah. In addition to these field-based project, Dr Cannell has written more broadly on the relationship between Christianity and social theory.

Dr Yaqub Chaudhary
Dr Yaqub Chaudhary holds a PhD in Physics from Imperial College London, where he worked on the Physics of Plastic Electronic Materials and their potential use in future types of lasers. Prior to this he studied Electronic Engineering at the same institution. He is currently Research Fellow in Science and Religion at Cambridge Muslim College, where his research is on Artificial Intelligence from the perspective of Islamic philosophy and theology. His research interests include recent developments in the fields of AI, cognitive science and neuroscience in connection with Islamic conceptions of the mind, intelligence, human reasoning, cognition, knowledge, the nature of perception and consciousness.

Dr Ron Chrisley

Dr Ron Chrisley is currently Visiting Scholar, Institute for Human-Centered Artificial Intelligence (HAI), Stanford University. He is director of the Centre for Cognitive Science (COGS) at the University of Sussex, where he is also on the faculty of the Sackler Centre for Consciousness Science, and Reader in Philosophy in the School of Engineering and Informatics. He was awarded a Bachelor of Science from Stanford University and a DPhil in Philosophy from the University of Oxford. Before arriving at Sussex he was an AI research assistant at Stanford, NASA, RIACS, and Xerox PARC, and investigated neural networks for speech recognition as a Fulbright Scholar at the Helsinki University of Technology and at ATR Laboratories in Japan. From 2001-2003 he was Leverhulme Research Fellow in Artificial Intelligence at the School of Computer Science at the University of Birmingham. He is one of the co-directors of the EUCognition research network, and is an Associate Editor of Cognitive Systems Research and Frontiers in Psychology (Consciousness Research). He is also the editor of the four-volume collection Artificial Intelligence: Critical Concepts.

Dr Andrew Davison

Andrew Davison is the Starbridge Lecturer in Theology and Natural Sciences at the University of Cambridge, and the fellow in theology at Corpus Christi College. He studied chemistry at the University of Oxford, followed by a DPhil biochemistry, before turning to the study of theology in preparation for ordination in the Church of England, with a further PhD following later, in mediaeval philosophical theology. He served a curacy in South East London and is currently the Canon Philosopher of St Albans Cathedral. Before his current position, he taught Christian doctrine at St Stephen's House, Oxford and later at Westcott House, Cambridge. He works on a range of topics at the interface of theology, philosophy and natural science. His recent work has been on the theological significance of the prospect of life elsewhere in the universe, mutualism in biology, cosmology and creation ex nihilo, evolution and traditions of divine exemplarity, and the role of the imagination in responding to the threat of climate change. Next year will see the publication of book with Cambridge University Press on the Platonic theme of participation as a structuring principle in Christian theology and metaphysics.

Dr Daniel de Haan

Daniel De Haan is a Research Fellow at the Ian Ramsey Centre for Science and Religion at the University of Oxford. Before coming to Oxford he was a postdoctoral fellow working on the neuroscience strand of Templeton World Charity Foundation's Theology, Philosophy of Religion, and the Sciences project at the University of Cambridge. He has a doctorate in philosophy from the Catholic University of Leuven and the University of St Thomas in Texas. His research focuses on philosophical anthropology and the sciences, philosophy of religion, and medieval philosophy.

Professor Eileen Hunt Botting

Eileen Hunt Botting is a Professor of Political Science at the University of Notre Dame. She is a 1993 Marshall Scholar and a Johnian who graduated with a second B.A. in Philosophy from Cambridge in 1995 before earning her Ph.D. in political science at Yale. She is the author of Family Feuds: Wollstonecraft, Burke, and Rousseau on the Transformation of the Family (SUNY, 2006), Wollstonecraft, Mill, and Women's Rights (Yale Press, 2016), and Mary Shelley and the Rights of the Child: Political Philosophy in 'Frankenstein' (University of Pennsylvania Press, 2017). She is currently working on the sequel to the latter book, titled Political Science Fictions after 'Frankenstein': Making Artificial Life and Intelligence.

James Kingston

I am a research manager at CognitionX, where I analyse the impact of artificial intelligence. I lead our research stream on the responsible development and deployment of artificial intelligence, and have recently completed our primer into the ethics of AI. I studied history and politics at Oxford University, and am a member of the English bar. Having spent two years working and studying in Beijing, I am particularity interested in the progress of AI in China and the Far East.

Professor Neil Lawrence

Neil Lawrence is Professor of Machine Learning at the University of Sheffield and the co-host of the podcast "Talking Machines". He is currently exploring industry on leave of absence from Sheffield at Amazon in Cambridge, where he is a Director of Machine Learning. He received his bachelor's degree in Mechanical Engineering from the University of Southampton in 1994. Following a period as an field engineer on oil rigs in the North Sea he returned to academia to complete his PhD in 2000 at the Computer Lab in Cambridge University. He spent a year at Microsoft Research in Cambridge before leaving to take up a Lectureship at the University of Sheffield, where he was subsequently appointed Senior Lecturer in 2005. In January 2007 he

took up a post as a Senior Research Fellow at the School of Computer Science in the University of Manchester where he worked in the Machine Learning and Optimisation research group. In August 2010 he returned to Sheffield to take up a collaborative Chair in Neuroscience and Computer Science.
Neil's main research interest is machine learning through probabilistic models. He focuses on both the algorithmic side of these models and their application. He has a particular focus on applications in personalized health and computational biology, but happily dabbles in other areas such as speech, vision and graphics.
Neil was Associate Editor in Chief for IEEE Transactions on Pattern Analysis and Machine Intelligence (from 2011-2013) and is an Action Editor for the Journal of Machine Learning Research. He was the founding editor of the Proceedings of Machine Learning Research (2006) and is currently series editor. He was an area chair for the NIPS conference in 2005, 2006, 2012 and 2013, Workshops Chair in 2010 and Tutorials Chair in 2013. He was General Chair of AISTATS in 2010 and AISTATS Programme Chair in 2012. He was Program Chair of NIPS in 2014 and was General Chair for 2015. He is one of the founders of the DALI Meeting and Data Science Africa.

Dr Michael McGhee
Dr Michael McGhee is Honorary Senior Fellow at the Department of philosophy at the University of Liverpool. He is author of *Transformations of Mind: Philosophy as Spiritual Practice*, and co-editor of *Philosophy as a Way of Life*. He has written many articles on philosophy of religion, moral philosophy, and contemporary studies in Buddhism.

Dr Simone Schnall
Simone Schnall is a Reader in Experimental Social Psychology at the University of Cambridge, and Director of the *Cambridge Body, Mind and Behaviour Laboratory*. She previously held appointments at Harvard University, the University of Southern California and the University of Virginia. Her research explores why people often think and behave in seemingly surprising ways, and how to capitalize on insights from behavioural science to encourage adaptive choices in everyday life. Current topics include judgments and decisions in moral and legal contexts, perceptions of the spatial environment, and risky behaviours in finance. Her work has been supported by numerous grants from public and private funding sources, and is routinely covered in the popular media such as the New York Times, The Economist, Newsweek, New Scientist and Psychology Today. Dr Schnall is a Fellow and Director of Studies for Psychology at Jesus College, Einstein Fellow at the Berlin School of Mind and Brain at Humboldt University, and a Fellow of the Association for Psychological Science.

Dr Beth Singler
Dr Beth Singler is the Junior Research Fellow in Artificial Intelligence, Homerton College, Cambridge, where she is exploring the social, philosophical, ethical, and religious implications of advances in AI and robotics from an anthropological perspective. She is also an associate research fellow at the Leverhulme Centre for the Future of Intelligence (CFI) where she is the co-lead and founder with Adrian Weller (CFI, Turing, CDEI) on the Faith and AI project. She is co-investigator on the CFI's Global AI Narratives project, and a collaborator on their overall AI: Narratives and Justice programme. Beth has produced a series of short documentaries on the future of AI and robots, of which the first, *Pain in the Machine*, won the 2017 AHRC Best Research Film of the Year Award. She has been asked to speak about her research on BBC Click, BBC4's Start the Week and Today programmes, for Forbes, for the Times, for the New Scientist, and at the Hay Literary Festival, among others.

Research Interests: Artificial Intelligence, Robots, Science Fiction, Anthropology, Religious Studies, Futurology, Transhumanism, Technology, Gender, Race.

Fr Ezra Sullivan OP
Ezra Sullivan is a Dominican friar and professor of Moral Theology and Bioethics at the Pontifical University of St Thomas Aquinas (Angelicum) in Rome, where he conducted doctoral research with Wojciech Giertych, the Theologian of the Papal Household. His articles and reviews have been published in such journals as Linacre Quarterly, Nova et Vetera, and Logos Journal. He is helping develop the Humanity 2.0 Project, and is frequently invited as a guest on Catholic radio and television. His forthcoming book is entitled Habits and Holiness: Thomistic Virtue Theory in Light of Modern Science (2019).

Professor Steve Torrance

Steve is a visiting senior research fellow in the Centre for Cognitive Science (COGS) at the University of Sussex, where he has also been, at different periods, an external examiner, and a visiting lecturer in AI and cognitive science, and where he completed an undergraduate degree in philosophy 50 years ago. He has a DPhil from Oxford on the formal structure of ethical judgment, and he has been working on the relation between AI, philosophy of mind and ethics since the early 1980s. He is Emeritus Professor of Cognitive Science at Middlesex University, where he worked in the philosophy, computing and psychology departments. He has organised many conferences and workshops, and has edited collections and journal issues, on topics including philosophy of AI, machine ethics, machine consciousness, social robotics, and enactivist theory. He writes on the moral status of AI agents; limits to computationalism as a theory of mind and consciousness; enactivist conceptions of mind and interaction; and the implications of technology growth for human civilisation. Working with a fellow jazz pianist and cognitive scientist, he also presents lecture-performances and writes on jazz and improvisation. He is an ethics reviewer for the European Commission's Horizon 2020 programme.

Professor John Wyatt

John Wyatt is Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology, University College London and a Senior Researcher at the Faraday Institute, Cambridge. His background is as a clinician and researcher in applied neuroscience. He has a long-standing interest in the philosophical, theological and social implications of advances in biomedical and information technology. John is currently co-leading a multidisciplinary research project into the social, ethical and theological implications of advances in artificial intelligence and robotics technology, based at the Faraday Institute.

Rabbi Dr Raphael Zarum

Rabbi Dr Raphael Zarum is Dean of the London School of Jewish Studies (LSJS), the UK's premier Jewish centre for teacher training, adult education and academic scholarship. He lectures in Jewish Education, Jewish Philosophy & Law, and New Readings of Classical Texts; and is involved in training teachers, rabbis and educators. He has rabbinic ordination from Rabbi Jonathan Sacks and the Montefiore Kollel Seminary, a PhD in Theoretical Physics from King's College London, an MA in Adult Education from University College London, and is a graduate of the Mandel School in Jerusalem.

Science & Human Dimension Project
Jesus College, Cambridge

John Cornwell
After 12 years on the editorial staff of *The Observer*, he was in 1990 elected Senior Research Fellow and Director of the *Science and Human Dimension Project* at Jesus College, Cambridge. In that role he has brought together many scientists, philosophers, ethicists, authors and journalists to debate a range of topics in the field of public understanding of science. His edited books include *Nature's Imagination*, *Consciousness and Human Identity*, and *Explanations* (Oxford University Press); *Power to Harm,* and *Hitler's Scientists* (Viking Penguin); *The Philosophers and God* (Bloomsbury Continuum), and *Darwin's Angel* (Profile). He is a Fellow of the Royal Society of Literature and was awarded an Honorary Doctorate of Letters (University of Leicester) in 2011. He was shortlisted Specialist Journalist of the Year (science writing in *Sunday Times Magazine*), British Press Awards, 2006. He won the Science and Medical Network Book of the Year Award for *Hitler's Scientists*, 2005; received the Independent Television Authority-Tablet Award for contributions to religious journalism (1994); and the 2019 Wilbur Award for religious journalism.

His journalism has been published in *Financial Times, Sunday Times Magazine, The Observer, New Statesman, New Scientist, Nature, Prospect, Times Literary Supplement, The Tablet, Brain, The Guardian, The Times.* Broadcast contributions to many BBC programmes, including "Hard Talk", "Choice", "Start the Week," "The Moral Maze", "Today" (debate with Richard Dawkins); "Beyond Belief", "Thought for the Day", "Sunday", and the BBC's World Service.

Jonathan Cornwell
Jonathan is Executive Director of the Science & Human Dimension Project and a Senor Research Associate at Jesus College, Cambridge. He has a background in academic publishing, digital business and e-learning in Europe, Middle East and China, working with a range of partners and clients including governments, universities, the private sector, NGOs, and UNESCO. He is a conference and art exhibitions producer and director of the consultancy Media Symposia. In addition to co-directing the Science & Human Dimension Project in Cambridge, he works in the arts, helping galleries, museums and curators produce exhibitions and apply technology. He is a former executive director of the Rustat Conferences at Jesus College Cambridge, convening high-level forums for leaders from government, industry and research. He studied at University College London, Cambridge University, and Imperial College London.

Dr Tudor Jenkins
Tudor holds a PhD in Artificial Intelligence and was a researcher in the Cognitive Science department at the University of Sussex and the Ecole Normale Supérieure, Paris. Dr Tudor Jenkins is Director and the founder of Wide Eyed Vision. He is a digital media specialist, photographer and developer of software and technology for cultural projects and exhibitions. Tudor developed the Houghton Revisited app, and has worked on projects for Hampton Court, The Rolling Stones Exhibitionism, the National Trust, English Heritage, Historic Royal Palaces, Bletchley Park, and Strawberry Hill's Horace Walpole Masterpieces exhibition. Tudor advises SHDP's AI and the Future of Humanity series and provides support for the programme's web and social media activities.

Participants List AI - Ethical and Religious Perspectives 16-17 May 2019
Science & Human Dimension Project Jesus College, Cambridge

The Revd Jennifer Adams-Massmann	Chaplain	Jesus College, Cambridge
Richard Addis	Journalist; Chairman	The Day
Dr Gorazd Andrejč	Woolf Institute; St Edmund's College, Cambridge	University of Cambridge
Prof Jean Bacon	Fellow in Computer Science	Jesus College, Cambridge
Dr Olivia Belton	Research Associate, Leverhulme Centre for the Future of Intelligence	University of Cambridge
Prof Jeffrey Bishop MD	Professor of Philosophy, Professor of Theological Studies, Tenet Endowed Chair in Health Care Ethics	St Louis University
Rhys Blakeley	Science Correspondent	The Times
Prof Andrew Briggs	Professor of Nanomaterials and Fellow, St Anne's College; Advisor, Templeton World Charity Foundation	University of Oxford
Andrew Brown	Journalist, author, editor	The Guardian
Revd Dr Malcolm Brown	Director of Mission and Public Affairs	Church of England
Dr Dominic Burbidge	Research Director; External Advisor, Templeton World Charity Foundation	University of Oxford
Dr Fenella Cannell	Reader in Social Anthropology	LSE
Dr Tom Chatfield	Author, broadcaster	Conference rapporteur
Dr Yaqub Chaudhary	Fellow in Science & Religion	Cambridge Muslim College
Dr Ron Chrisley	Reader, Cognitive Science; Visiting Scholar, Stanford Institute for Human-Centered Artificial Intelligence (HAI)	University of Sussex; Stanford University
Dr Jennifer Cobbe	Cambridge Trust & Technology Initiative; Cambridge Computer Lab, Department of Computer Science & Technology	University of Cambridge
Jonathan Cornwell	Senior Research Associate; Executive Director, Science & Human Dimension Project	Jesus College, Cambridge
John Cornwell	Director, Science & Human Dimension Project	Jesus College, Cambridge
Revd Dr Andrew Davison	Starbridge Lecturer in Theology and Natural Sciences; Fellow in Theology, Corpus Christi College	University of Cambridge
Dr Daniel De Haan	Research Fellow, Ian Ramsey Centre for Science and Religion	University of Oxford
Dr James Dodd	Technologist and investor; Member, Rustat Conferences, Jesus College, Cambridge	
Novin Doostdar	Publisher	Oneworld Publications
Dr Ryan Doran	Researcher, Cambridge Body, Mind and Behaviour Laboratory	University of Cambridge
Marius Dorobantu	PhD candidate: Imago Dei, personhood and human uniqueness, theological reflections on challenges of strong AI; Department of Theology	University of Strasbourg
Rev Dr Rory Doyle, OFM Conv	Parish Priest	Conventual Franciscans
Madeline Drake	Housing, Health and Social Policy Consultant	
Dr Sam Freed	Fellow, Department of Informatics, University of Sussex; Teaching Fellow, Department of Computer Science, Hebrew University, Jerusalem	University of Sussex; Hebrew University, Jerusalem

Rabbi Dr Moshe Freedman	Rabbi	New West End Synagogue, London
Dr David Good	University Lecturer in Psychology	University of Cambridge
Joy Green	Senior Futures Specialist	Forum for the Future
Ryan Haeker	Doctoral Student, Faculty of Divinity; Peterhouse	University of Cambridge
Imam Dr Usama Hasan	Head of Islamic Studies	Quilliam International
Prof Stephen Heath	Professor of English and French Literature and Culture; Fellow, Jesus College	University of Cambridge
Prof Teresa Heffernan	Department of English Language and Literature	St Mary's University, Canada
Prof Eileen Hunt Botting	Professor of Political Science	University of Notre Dame
Rabbi Laura Janner-Klausner	Senior Rabbi to Reform Judaism; broadcaster	Reform Judaism
Revd Dr Tim Jenkins	Reader in Anthropology and Religion, Faculty of Divinity; Fellow, Jesus College	University of Cambridge
Dr Tudor Jenkins	Director, Wide Eyed Vision; AI Researcher; Advisory Board	Science & Human Dimension Project
James Kingston	Research Manager; author, AI Ethics Primer	CogX
Ante Kotarac	Co Founder, Chief Product Officer	AgentCASH
Mgr Mark Langham	Chaplain, Fisher House	Catholic Chaplaincy, University of Cambridge
Prof Monica Lam	Professor of Computer Science and Electrical Engineering; Faculty Director, Stanford MobiSocial Laboratory; Co-founder and CEO, Omlet	Stanford University
Prof Neil Lawrence	Professor of Machine Learning, University of Sheffield; Director of Machine Learning, Amazon Cambridge	University of Sheffield; Amazon, Cambridge
Prof Julius Lipner	Emeritus Professor of Hinduism and the comparative study of religion	University of Cambridge
Dr Johanna Lohn	GP; Medical Ethics, King's College, London	
Keith Mansfield	Publisher, futurist, cosmologist, author, *The Future in Minutes*	
Dr Christopher Markou	Researcher in Law and Technology; Faculty of Law; Jesus College	University of Cambridge
Dr Alexander Massmann	Visiting Fellow Faculty of Divinity; Postdoctoral Research, Heidelberg University	University of Cambridge
Cassius Matthias	Founder and Editor, Yes & No magazine; artist, film maker	Yes & No Magazine
Prof Michael McGhee	Hon Senior Fellow, Philosophy	University of Liverpool
Dr David Mellor	Senior Lecturer in Sociology	University of South Wales
Dr Nathan Mladin	Researcher	Theos
David Moloney	Editorial Director	Darton, Longman & Todd Publishers - DLT
Dr Ken Moody	Fellow in Computer Science	King's College, Cambridge
Prof John Naughton	Senior Research Fellow, CRASSH; Fellow, Wolfson College; Emeritus Professor of the Public Understanding of Technology, OU; Technology Columnist, The Observer	University of Cambridge
Stephen Ochsner	Actor; theatre curator	Caravane
Elizabeth Oldfield	Director	Theos

Sumit Paul-Choudhury	Journalist, publisher, Editor Emeritus	New Scientist
Dr Igor Perisic	Chief Data Officer, VP Engineering	LinkedIn
Colin Ramsay	Director, Producer	DragonLight Films
Prof Dale Russell	Visiting Professor, School of Design	Royal College of Art
Dr Elisabeth Schimpfössl	Lecturer in Sociology	Aston University
Dr Simone Schnall	Reader in Experimental Social Psychology; Director, Cambridge Body, Mind and Behaviour Laboratory	University of Cambridge
Dr Vicky Schneider	Senior Scientific Programme Manager, Amazon; Hon Assoc Professor, University of Melbourne; Visitor, Department of Genetics, University of Cambridge	Amazon
Dr Andrew Serazin	President	Templeton World Charity Foundation TWCF
William Simpson	Philosophy Faculty; Peterhouse	University of Cambridge
Dr Beth Singler	Junior Research Fellow in Artificial Intelligence, Homerton College; Associate Fellow, Leverhulme Centre for the Future of Intelligence	University of Cambridge
Austin Stevenson	Doctoral student, Girton College; Faculty of Divinity	University of Cambridge
Oliver Soskice	Artist	
Fr Ezra Sullivan O.P.	Lecturer, Theology Faculty	Angelicum, Pontifical University of St Thomas Aquinas, Rome
John Thornhill	Innovation Editor	Financial Times
Sam Thorp	Science & Human Dimension Project	Jesus College Cambridge
Prof Steve Torrance	Visiting Senior Research Fellow, Centre for Research in Cognitive Science (COGS)	University of Sussex
Dr Karina Vold	Postdoctoral Researcher, Leverhulme Centre for the Future of Intelligence; Research Fellow, Faculty of Philosophy; Newnham College	University of Cambridge
Prof Richard Watson	Futurist in Residence, Tech Foresight; author	Imperial College London
Dr Adrian Weller	Programme Director for Artificial Intelligence, The Turing Institute; Senior Research Fellow, Department of Engineering; board member, Centre for Data Ethics and Innovation	University of Cambridge; Centre for Data Ethics & Innovation
Christopher Whitney	Director, Principal Gifts	University of Cambridge
John Wilkins	Journalist; former Editor, The Tablet	
Prof Yorick Wilks	Artificial Intelligence	University of Sheffield
Prof John Wyatt	Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology, University College London; Senior Researcher, Faraday Institute, Cambridge	UCL and Faraday Institute, Cambridge
Rabbi Dr Raphael Zarum	Dean	London School of Jewish Studies LSJS