Artificial Intelligence in Sci-Fi Film and Literature Imagining AI - the View from Mythology, Fantasy, and SF Conference Report

Science & Human Dimension Project Conference 15-16 March 2018 Jesus College, Cambridge

Contents

Introduction	1
Report	4
Conference Overview	12
Conference Agenda	13
Speaker Biographies	16
Participants List	24

Introduction: John Cornwell, Director, Science & Human Dimension Project, Al and the Future of Humanity series:

This meeting today is the first in a new cycle of three conferences on AI, and our focus is how artificial intelligent systems are likely to affect the ways we think about ourselves as persons, as individuals, old and young, members of families, of communities, of society: as individuals and groups with convictions, values and beliefs. Will we continue to see machines as things, as in the past, or will the differences between persons and things begin to blur? One detects a distinct feeling of unease about the coming of superintelligent machines, quite apart from the predictions of existential risks.

Our focus is summed up in that powerful phrase uttered by Kent to Oswald in King Lear about teaching differences. And a problem from the outset, it seems to us, is not only on the side of understanding machines that learn, difficult as that may be, but understanding what it means to be a person.

For example, despite rapid and broad advances in cognitive neuroscience - a not always happy combination of neurophysiology and psychology - there are significant differences in the qualitative conclusions of philosophers of mind (albeit literate in neuroscience) as they address the most crucial aspects of traditionally accepted differences between things and persons. Witness the perspective of the constituency of influential thinkers as represented by Dan Dennett and his many followers who deny the existence of the Self, deny the existence of an interior mental life, deny Free Will, deny consciousness as a reality (consciousness, as Dennett has put it, is an illusion thrown up by a mundane bunch of tricks, the same goes for the Self and Free Will): hence a short step, one might think, from human

to machine equivalence. Moreover, Professor Dennett asserts that there is something wrong with us if we disagree with him. Well, he may be right. Although it has to be said that if you follow soaps like the BBC's Archers and EastEnders, traditional notions of persons with a sense of I and Thou and with responsibility for one's actions are alive and well. And there are well-respected philosophers, equally versed in neuroscience, Mary Midgely, Antonio Damasio, David Chalmers, for example, who would vigorously defend the notion of self, free will; the hard problem of consciousness, and the reality of a mental life.

Meanwhile, something strange and interesting occurred at our last AI conference, which we held here at Jesus College, Cambridge in September 2017, which might have the power to by-pass the stories told by philosophers skilled in neuroscience. And the key is imagination - a faculty or behaviour rarely mentioned by philosophers or neuroscientists.

Dr Demis Hassabis Co-founder and CEO of DeepMind and twenty six of his colleagues joined us to discuss "human and machine memory and imagination" with a group of academics in philosophy, anthropology, literary studies and theology.

That meeting marked a striking alteration in the direction of travel that metaphors explicating machine and human differences usually take. There has been a tendency in the early to the late modern period to seek to understand the mind-brain relationship in terms of the machines that fascinate us, in other words mechanical metaphors: Leibnitz exploited the cogs, wheels and belts of the windmill, Victorian psychologists invoked the thermostat of the steam engine, the Edwardians the telephone exchange, as late as 1942 the neurologist Sir Charles Sherrington in his Man on his Nature appealed to "an enchanted loom where millions of flashing shuttles weave a dissolving pattern." Early cyberneticians found parallels in the feedback of anti-aircraft guidance systems. In the second half of the 20th century it was, and still is for many, the computer with its hardware, software, programmes, downloads, databases, and retrievals, despite the fact that neural nets made their debut as early as the middle of the Second World War.

Dr Hassabis and his colleagues were not seeking to explain the human mind with recourse to machine analogies, they were seeking to exploit known features of the human mind-brain function to assist in, and explicate, their AI design strategies.

Memory in the AlphaGo system is not the retrieval of inert bits of programmed information from a database, but memory as construction or reconstruction, creation even, from the dynamic activity of the neural nets akin to neuronal groups in the human brain. And not only does this work for the system's memory, they claim, but the capacity of the system to envisage a range of future options, and calculate the far reaching consequences of those options aided by algorithms appropriate to a value, namely winning the game, before making each move. It was the system's capacity to construct, or reconstruct, memory and future scenarios and their consequences that Dr Hassabis described as "imagination", a term that seriously scandalised many of our humanities audience at the conference. "Imagination," Dr Hassabis said, " is one of the keys to general intelligence, and also a powerful example of neuroscience-inspired ideas crossing over into Al."

Another shift, worth noting, was the comment by AlphGo technicians that they could not explain how the system, in its ultimate game with Lee Sedol in Seoul, calculated its winning moves except to invoke the word intuition in addition to imagination. What a step from the days when the cyberneticians argued that in theory a machine could replicate any function or behaviour that could be defined in a finite number of words.

Try defining Intuition, Imagination, in a finite number of words!

I suspect that powerful as the resemblance between human and AlphaGo's function may be in forward thinking, the terms intuition and imagination are more metaphorical than literal. And as that great early modern philosopher of imagination, Giovanni Battista Vico warned, it is hazardous to mistake our metaphors for reality.

And perhaps I could be allowed a footnote to illustrate my point: in his book *Consciousness Explained*, Dan Dennett comes up with the idea that the self, which, according to him, does not exist, is no more and no less than a series of multiple drafts of one's life stories, or, a narrative centre of gravity, a notion he believes not to be a metaphor, but the thing itself. But, and here's the thing, he concedes that he found the concept and the definition in a novel by David Lodge, *Nice Work*.

Nice Work is about an exchange scheme whereby the Managing Director of an engineering factory shadows Robyn, an Eng Lit department lecturer in literary theory. The Managing Director, attending one of Robyn's seminars, finds himself totally baffled as Robyn, who has been overdosing on Derrida, expounds this same multiple drafts theory of the self in a seminar.

So here's the situation: Dennett's illusory version of the self is borrowed from the fictional deconstructionist notion of a literary theorist, who has overdosed on Derrida, as ridiculed by a satirical novelist who, if you'll excuse the expression, is satirising the whole idea.

So who ideally can teach us differences in the realms of imagination, intuition, metaphor persons and machines? They are of course the artists and writers across the broad scope of mythology, speculative fiction, gothic and horror fantasy, literature of ideas, supernatural and superhero fantasy, and the rag bag of genres known as Science Fiction.

From the Prometheus myth to the Book to Genesis, to early modern narratives of the Hebrew Golem, to Mary Shelley's Frankenstein, to War of the Worlds, the Man in the Moon, Metropolis, the explosion of 20th century short stories, comic books, stage and broadcast drama, novels, movies, TV series, painting, sculpture, poetry, dance and music, imaginative artists have explored the borderlands of the human - machine divide, equivalence, competition, antagonism, from the theme of man playing God, to machines playing God, to Aliens and Monsters, to the notion of artefactual entities whose existences are unbearable to them, with every imagined possibility, and impossibility, hope, fear, unbridled hubris, catharsis and apocalypticism...and in there somewhere we might find room for UFO cults like Heaven's Gate whose devotees were prepared to sacrifice their lives for their fantasies.

Which is why we're truly excited by the possibilities of this conference on AI in Science Fiction. Who knows what dramas and psycho dramas of ideas and visions, fantasies, and emotions, will be unleashed? But I hope that by the end of tomorrow afternoon we might have made some connections, formed some valuable insights, deep or superficial, which will take our teaching and learning of differences to a new level?

But before we take off into outer space, or inner space, I'm going to ask Professor Murray Shanahan of Imperial College, and DeepMind, author of a brilliant book entitled Singularity, and erstwhile technical adviser on the film Ex Machina, to do something practical and mundane, namely to come up with some basic definitions and terminology within the field of Artificial Intelligence to at least keep the feet of our nomenclatures and terminologies firmly on the ground from the outset. And with that, may I wish you all a truly enjoyable and fruitful conference.

Conference Report

So what is it about these fictional narratives that makes them relevant in examining the future of AI? After all, today's specialist AI is a far cry from the fully developed minds we encounter in literature, from Blade Runner's Nexus androids to Hal in 2001: A space Odyssey. "We don't know what artificial general intelligence will be," said Murray Shanahan, professor of Cognitive Robotics at Imperial College, London and Research Scientist at DeepMind. "So instead, we are projecting forward on the current state of the art." And if no one has technological insight into whether or how AGI will come about, fiction writers are just as qualified as computer scientists to hypothesise.

Indeed, fiction and mythology may be in a unique position to offer guidance, thanks to their role in shaping our world, said Beth Singler of the Faraday Institute for Science and Religion, Cambridge University. This is particularly salient for engineers and computer scientists: science fiction authors conjure the imaginary worlds that these specialists then bring into reality.

What is more, a lot of science fiction, fantasy and mythology stories themselves have roots in ancient parables that are concerned with what it means to be human, especially with respect to ethics and morality. "We have had all these stories for centuries," said Stephen Cave of the Leverhulme Centre for the Future of Intelligence. "Now they're becoming important."

As a broad range of experts, including science fiction authors, AI researchers, futurists, researchers in philosophy, theology, and literature, journalists, critics and publishers debated the most relevant works over the two days of the conference, five major thematic questions emerged.

I. What exactly do we mean when we speak of AI?

90 percent of experts believe AI will happen within a century. But what do we mean by AI? AI already exists. In its current state of the art, AlphaGo can beat world champion Lee Sedol at the game Go. Supercomputers now dominate at chess. Facebook's AI sorts pictures. However, Shanahan pointed out, each individual AI can only do one of these things, whereas "Lee Sedol can also label his own pictures and drive a car by himself". (Sedol also instinctively knows when to do which task.) A putative so-called artificial general intelligence (AGI) would have to have a similarly complex human-level understanding of the world.

But when we speak of AI do we simply mean something that is as intelligent as a human? Or is it something more? The subject of many of our most anxious questions is not AGI - it is the related concept of superintelligence, defined as an AGI that is able to transcend mere human intelligence to leave us in the dust.

This superintelligence is the basis of many assumptions around the future of AI. But why? If we can't even conceive of how technology will achieve generally humanlike intelligence, why take the concept of superintelligence seriously at all? As New Scientist chief strategy editor Sumit Paul-Choudhury pointed out, why is Nick Bostrom's book, Superintelligence, considered a work of prognostication and not science fiction?

The problem is that the "Singularity" upon which so many future AI scenarios depend is defined as the point after which all predictions are moot. It's hard for science to weigh in under such conditions. However, there are plausible paths forward. "We don't have empirical data on the future, so we need to look at technological trends," said Olle Häggström, of the Institute for Future Studies in Stockholm. Thore Husfeldt, professor of theoretical computer science at Lund University, Sweden, did just that, extrapolating from current research projects to identify five pathways to the superintelligence.

- 1) As a result of a breakthrough in current trends of AI or artificial life.
- 2) An emergent intelligence brought about by a massively networked collaboration of humans, machines, and corporations similar to the way the mind emerges from the activity of the brain's network of billions of neurons. "This is the skynet scenario," said Dr. Husfeldt, owing to the sudden unintended shift to self awareness such a network might achieve.
- 3) Via technological augmentation of existing human brains. "Your mind, but gradually augmented and replaced by electronics that make it more intelligent, for example to help you become a chess master," he said.
- 4) In theory, superintelligence could also be achieved organically, for example by evolving everbetter human brains via a programme of eugenics or gene modification.
- 5) Via uploading human brains into the computational ether.

"Ultimately we don't know what AGI will be," said Shanahan. "Maybe it will be something we can barely imagine." For example, why presume a mind or body like our own? Superintelligence doesn't need to be humanlike: Patrick Crogan at UWE says drones and swarms may be a distinctly nonhuman pathway to artificial intelligence. And why stop there? Does AI even need to be defined as a computer? Other AI platforms are possible, suggested science fiction author and former research biologist Paul McAuley. It could be instantiated by microbiomes, forests, viruses, plankton, or even planetary intelligence. It is not only difficult to say what we mean by AI, it is difficult to pin down just what an AI isn't.

One of the main things that distinguishes the organic from the inorganic pathways is the presence of consciousness. For machines, even superintelligence does not equate to consciousness, said Ron Chrisley, director of the Centre for Cognitive Science at the University of Sussex. In his examination of

Destination: Void, he made the case that the latter does not arise automatically from the former. Great pains may need to be undertaken in order to uplift even a highly intelligent machine to a conscious one. But should we? After all, in Destination: Void, a machine's journey to consciousness ends in insanity. Should we take precautions to ensure that AGI or Superintelligences remain "zombies" - intelligent enough to do our bidding but not conscious enough to suffer?

II. What is it like to be an AI?

The essay "What is it like to be a bat?" by American philosopher Thomas Nagel, first published in 1974 in The Philosophical Review, contains relevant warnings for how to think about AI in the 21st century.

Built into many assumptions about the future of artificial general intelligence is an assumption that this AGI will automatically be just like us, possessing a native human understanding simply because humans created it. Missing from this perspective is that such a creature will almost certainly be alien to us in essential ways.

Consider how human understanding is shaped by the physical structure of the human body. For example, the position of two eyes on the front of our head has led us intuitively to the concepts of "in front of" and "behind". However, the AIs we develop should not be shaped or limited by human physical understanding, science fiction author Ian McDonald reminded the audience. A driverless car with a 360-degree field of vision would not only become a better driver than any human could possibly be - its entire understanding of the world would diverge from ours, simply as a result of this perceptual choice. "Humans' entire world view is shaped by the fact that we consider things forwards and backwards," he said. Could an AI designed with a 360-degree field of vision AI even conceive of something being "behind" or "in front of" it? Such seemingly simple design choices could create conceptual gaps between human and AI understanding that may be difficult to bridge.

And that's just the structure of the perceptual organs. There are many more ways machine vision differs from humans'. Kinesic recognition, for example, enables machines to see patterns of movement that humans cannot. However, machine vision is not better in every way: humans can see cyclists, but driverless cars currently cannot, said McDonald. Al has certain blind spots/visual flaws.

The relationship between knowledge and memory is also very different in machines, said science fiction writer Justina Robson. Machines, by the nature of how we design them, will be unable to forget - ever. "What is it like to have idetic memory," asked Robson. "What is it like to remember everything all the time?"

Adam Roberts, science fiction author and Professor of English Literature at Royal Holloway, University of London, saw an answer in Mary Shelley's Frankenstein, an example of a non-machine artificial intelligence which nonetheless perceives the world in a way particular to machines, in that it is incapable of forgetting. The human ability to selectively forget is utterly underrated - it underpins behind our ability to remember specifics and forget unimportant details. It underpins our ability to move on from pain, and forgive. A creature that remembers everything that has ever happened in its life - the hallmark of modern computers - can easily become a creature consumed with unforgettable rage and the need for revenge.

If divergence from humanlike characteristics spells trouble for human-AI interaction, should we engineer AI to be more humanlike? As the example of Frankenstein suggests, one way to do so is to make them more fallible. Giving them the ability to forget could stave off one form of insanity - and yet this inbuilt irrationality could also cause suffering. Indeed, anything that makes a machine more able to relate to us would make it inherently irrational, especially subjective consciousness. As John Cornwell

asked: "If we create machine consciousness, might we be creating a being that is in pain? It may spend its existence in unbearable pain, and yet we wouldn't know." And yet, without this irrational underpinning, AI would be alien to us.

Whether or not it becomes possible to relate to AI and understand its motives, we do have rational ways to infer what it will care about, said Hallvard Haug. These can be roughly categorised as: self-preservation; the acquisition of hardware and other resources, for example energy; improving its own software and hardware; and preservation of its final goal.

Irrespective of whether or not the AI is of an aggressive nature, these may become the instrument of our doom.

III. What purpose is AI meant to serve?

What role is AI destined to take in our world? Mythology and science fiction are consistent in their portrayal of the unintended consequences of creating a thing only to enslave it. In the Iliad, Hephaestus' golden maidens were woman-shaped, gold automatons crafted to be attendants for his palace. The modern version is the robot - the word originally coined in 1920 in Karel Capek's Rossum's Universal Robots, and with all the connotations that make us avert our eyes when they are used to describe human experiences: hard labour, serfdom, indentured work.

Indeed, Hallvard Haug pointed out that Capek's original robots were not made of metal at all, but rather of a humanlike shell, in today's vernacular not so much robots as androids. The first AI may have been the Golem, said Yaron Peleg, Reader in Modern Hebrew Studies at Jesus College, University of Cambridge - made of clay, and programmed by symbolic operation to do its creator's bidding.

The fantasy of a "programmable" servant wants its cake and eat it too: To be the perfect servant, a creature must have sufficient conscious awareness and humanity to predict and attend properly to our needs. But there is a danger that if we create something too human-like, it becomes aware - and our very creation of them has condemned them to a life of enslavement.

How can we possibly justify such actions? Many stories, including Blade Runner, show the increasingly convoluted measures we must develop to sort beings into those who are served and those who must serve. And why must those serve? Because of their identity as beings who were manufactured - because of their origins, in short. This too is a familiar justification. History is full of horrors stemming from insistence on sorting people by their origins into exploiter and exploited.

We can't seem to shake that history. Alexa, Siri, Cortana - all these assistants have female voices. Why do people - irrespective of gender - overwhelmingly choose a female voice to be their assistant? We want a traditional secretary in the sense that we want to make it do our bidding, but we want it to be all-knowing. We want "a superintelligence that does all the paperwork," said Beth Singler.

So we want assistants who are all knowing, all seeing and all-powerful over our lives? This is one working definition of a God. Is it a good idea to enslave a God?

Or perhaps we do not want a slave at all - perhaps what we seek from a superintelligence is actually to be enslaved.

Do we have a death wish that we hope to see executed by AI? Murray Shanahan and Kerry Shanahan examined this in "Existential risk and the unconscious", their analysis of the film Forbidden Planet. Kelly Shanahan, at University College London, suggested that we fear that AI will inevitably be too much like

the humans that created it, channeling the monstrous id of its creator. "If you play around with nature too much, you see what's at the real core of your ego - the id," she said. "We are so close to advancing ourselves to the god level, only to realise that our primitive self controls us."

That primitive self may require a godlike figure even if it is housed in an otherwise rational mind: a benevolent AI, set on minimising the suffering of humanity. Because it would be endowed with an intrinsically utilitarian mindset, it would punish anyone insufficiently supportive of its goals. Those goals include coming into existence in the first place.

Roko's Basilisk is such an example of implicit religion - religious like behaviour, belief and ritual in a community that espouses no actual religion. "Al is repeating many of our previous stories, many of them religious parables," said Singler. "It's interesting that this explicitly secular community is adopting religious categories, narratives and tropes."

IV. What does the AI apocalypse look like?

There are obvious visions of AI rebellion. One path is through AI Superintelligence containment failures, said Sci Fi author and publisher Keith Mansfield. These would be driven by a combination of several factors: individual desperation, corporate competition to set an AI loose, an AI equivalent of anti-slavery movement, misplaced confidence, or outright criminality.

However, there are also some indirect paths to the AI apocalypse. God or slave, any AI whose priority is goal preservation can rapidly turn bad - the logical endpoint of a machine that optimises a utility function. This "perverse instantiation" has been the topic of many morality fables: from the genie in the bottle parable to the tale of the Sorcerer's Apprentice, our wish for an AI that solves every human problem is bound up with fears of unintended consequences.

Its most recent incarnation is the "paper clip maximiser" thought experiment, in which an AI tasked with the apparently anodyne goal of maximising paperclips will devour every life-sustaining resource in the solar system in the service of this seemingly trivial goal. In AI, there is particular danger in failing to spell what we do not want in the course of specifying what we do want.

The trouble is, a programmed goal whose implications the human programmers have not thought about in sufficient detail. The film Colossus: The Forbin Project shows that even when AI delivers everything its human creators dreamed of, "they can still end up disconsolate and enslaved," said Mansfield.

The idea of an AI enslaving us because we become incidental to the goal we have set it - or stand in the way of that goal, as humanity's very nature stood in the way of Colossus' quest for world peace - is a frequent theme. Why?

But perhaps it won't be as dramatic as all that. Might the AI automation apocalypse simply bore us to death? After all, if slave AIs do all our bidding, will there be anything left for us to do? It could be a fate worse than enslavement: "We are dying to be free from work but terrified of being out of work," said Stephen Cave. "Some of us struggle with knowing what to do with a rainy Sunday afternoon. Well, now imagine nothing but rainy Sunday afternoons forever."

Alternately, what if the outcome of an AI takeover is beneficial for humans? Apocalypse is not inherently a bad word - Paul McAuley considered the possibility that a godlike AI could uplift humanity, rescuing us from our own id: Could the AI we create become like the monoliths in 2001: A Space

Odyssey, sending apes on their journey to becoming human? Emma Reeves and David Chikwe, creators of the BBC children's show Eve, said: can we build something better than humanity?

From Hal's behaviour in 2001: A Space Odyssey to the benevolent dictatorship of Colossus in The Forbin Project, science fiction is full of AI characters, both sane and not, that seek to dominate or exterminate humanity. But science fiction novelist Lavie Tidhar suggests that we have been conditioned to fear AI by reading science fiction that is too culturally homogeneous. "American science fiction is the product of the start of a rising empire," he said. "British science fiction is a product of a dying empire."

These cultural world views inform our understanding of AI as "inherently dangerous," he said. But what if it's not? What if the AI we create becomes more like us in other important ways? "Most of us don't like to do very much if we don't have to," he pointed out. "Maybe AGI will be a slacker."

V. Whichever AI we create, do we have any hope of controlling it?

In case it is not, "at some point we should expect the machines to take control", said Alan Turing in 1951. Only over the past decade did legitimate researchers like Nick Bostrom begin to take this idea seriously, said Olle Häggström. What are the consequences for us? Husfeldt, professor of theoretical computer science at Lund University, Sweden, explained Max Tegmark's AI scenarios: It could become anything from a zookeeper to an enslaved god.

What will make the difference? In terms of creating constraints on AI to keep humans safe, literature has not got far beyond Isaac Asimov's three laws of robotics, Kanta Dihal, of the Leverhulme Centre for the Future of Intelligence, pointed out. Neither has the real world - AI rules are popular and popping up everywhere, from German driverless car laws to the UK and the IEEE. At the Asilomar conference, attendees developed 23 principles, priorities and precautions to guide the development of artificial intelligence, ensuing safety, ethics and benefit to society.

But are such laws enough to keep AI in check? Dihal revealed that the 3 laws of robotics was never intended as a prescription against misbehaving robots - indeed Asimov himself admitted to creating them specifically to create a conflict to motivate the narrative.

And yet as Häggström said, "we need to make sure AI's goals are aligned to our own."

Perhaps the way to make sure our goals align with the machines is to literally merge with them. Thore Husfeld explored the different scenarios within which this could take place: neural implants for augmentation; storing human consciousness on hard drives to use when "sleeving" and "resleeving" them into a series of human bodies. In his workshop, he asked participants whether they were interested in uploading themselves in this way. Most were unsure.

What would be the consequences of such a merger? The science fiction writer Ann Charnock recalled the futurist Ray Kurzweil's predictions, in 1999, for the year 2099: "Among those human intelligences still using carbon-based neurons, there is ubiquitous use of neural implant technology, which provides error correction and augmentation of perceptual and cognitive abilities. Humans who do not utilise such implants are unable to meaningfully participate in dialogues with those who do." The potential schism between machine-augmented humans and regular humans, she said, gave her nightmares.

However, Lavie Tidhar and Justina Robson saw a better way to ensure machine-human cooperation. Tidhar wondered whether fears of AI taking over were overblown. "AI will not be able to live without us," he explained. We will be needed for routine maintenance. If AI becomes humanlike, it will need a human therapist to guide it through the experience, said Justina Robson. AI will need teachers. "They will be so desperately dependent on us. There is no reason to fear AI," said Tidhar.

Concluding Comments and Recommendations from Participants

"As we are now thinking so hard about implications of AI and Machine Learning, it makes sense to be going back to some of the early myths and legends and ideas to see what there is in there that might explain our current obsessions or indeed help us to do things better."

"In the 1920s in both the West and under Communism, people believed that machines would be central to the delivery of a better and new society. People began to think of the body as a machine, the city as a machine, society as a machine, human beings as machines. Many of these ideas as applied to AI are relevant now."

"The Golem is one of the earliest manifestations in literature of a creature which is part-human, partmachine. It has become one of the most articulate expressions of this dichotomy of construction and destruction, and of the interference of man with nature, and the attempts to create a human being or human-like creature which would have super powers."

"Sci-Fi has had a tendency to be very white male dominated... and now many more women and more minorities and people from other parts of the world are involved, and hopefully it won't be so monocultural – some of the concerns about AI are white, western concerns, about control, loss of control, about power, economics, and politics."

"I'm hoping that research into AI will feed back into engineering systems that we can use within ourselves to become more contented individuals who will live to their full potential."

"There are echoes in old stories (*Golem* and *Frankenstein*) and, more recently, films (*Forbidden Planet*) that are important to us because they speak to deeper human concerns and deeper forms of story telling.... those current working in AI should find ways of linking in to those older narratives."

"Sci Fi is the art of the spectacle - and this is both a blessing and a curse. It's a very visual, in-your-face, emotional medium - it has the power to terrify. Science Fiction (rather than actual AI) is mostly responsible for this. Sci Fi is guilty as charged.. Sci-Fi would like to more responsible but it doesn't sell books!"

"Al researchers and companies have a different interests from sci-fi authors. Unless the Al community speaks to this creative Sci-Fi community in a language they can understand and which resonates with their need to tell stories, then Al wont be listened to. Al researchers have a responsibility to give Sci-Fi creators more to work with. Al researchers shouldn't expect Sci-Fi somehow to reflect the science, or to get it right, or even to to be socially responsible.. Sci-Fi writers aren't there for that.. they're not necessarily socially responsible."

"There is no reason emerging new AI technology shouldn't be seen by Sci-Fi as an exciting and positive thing, especially for young audiences. Sci Fi can show that seemingly evil machines can become your best friends and help us. Ideally though, Sci-Fi will be nuanced so that the stories can tell how AI could be used for ill and for good."

"Public perception of AI is sometimes shaped by Sci Fi - maybe this happens rather too much. People may be misled by Sci-Fi into thinking that AI is more capable that it actually is, that it's more threatening that it really is, or maybe that it's more human than it really is. So often, the motivation for the story-teller is to make something that is entertaining and draws you in as a reader and viewer. Sci-Fi is a great source of inspiration and ideas and a spur to thinking about a lot of these important issues, about the philosophy of AI, the ethics of AI, but maybe we shouldn't take it too seriously. Real AI is not going to look like the Terminator, even if that film does present an exciting image of AI - more so than, say, driverless cars!"

"This seems like it could turn into a what-if fest. Who is funding the projects that may lead to the different approaches? What are their priorities? Which vision is winning and losing?" asked an audience member.

"the unintelligent processes of the market are already sending us into a paper clip scenario."

So what are the current directions in AI development, and how might they manifest in terms of the future they create? To go back to Murray Shanahan's observation, "we don't know what AGI will be. Instead we are projecting forward on the current state of the art."

The next conference in the Science & Human Dimension Project's AI and the Future of Humanity series, The Singularity Summit – Imagination, Memory, Consciousness, Agency, Values, will examine why projects now underway seeking to simulate human "intelligence" must engage in dialogues with scholars of philosophy of mind, literary and cultural studies, and philosophical theology.

Artificial Intelligence in Sci-Fi Film and Literature Imagining AI - the View from Mythology, Fantasy, and SF

Science & Human Dimension Project Conference 15-16 March 2018 Jesus College, Cambridge

Overview

Since at least as early as Mary Shelley's Frankenstein, artificial intelligence in the form of fiction has been a constant theme in the modern period. In addition to an explosion of Sci Fi novels, in recent decades there has been a spate of popular TV shows and movies imagining a variety of scenarios depicting intelligent robots and super-intelligent machines: *Blade Runner* (1982), *Blade Runner* 2049 (2017), *Battlestar Galactica* (2004-), *Wall-E* (2008), *Black Mirror* (2011-), *Ex Machina* (2015), *Her* (2015), *Westworld* (2016). At the same time there has long been a cross-fertilisation between hypotheses, goals and conceptual understanding between Sci Fi and AI research. For example, the term "singularity" (or technological singularity) was introduced by the science fiction writer Vernor Vinge in a 1983 Omni magazine article; it was picked up by Ray Kurzweil in his popular 2005 book *The Singularity is Near*.

At many other points we similarly find fiction in all its forms driving ideas in AI, and vice versa. Film and literary fictions have thus become catalysts for the drama of ideas that circulate around AI developments and its future, including its moral and spiritual dimensions. There are also powerful religious themes in the history of Sci Fi representations of machine intelligence, including the achievement of immortality, notions of Omega point futures, transhumanism, and the prospect of androids outstripping humans in virtue. Clearly the hopes, fears, ambitions, dangers, and hubris of future AI are here being given dramatic and imaginative expression.

The aim of this conference, the first in the SHDP's AI and the Future of Humanity project, is to discover what we can learn about ourselves in relation to AI by exploring fictional narratives. By initiating our project with a conference that interrogates and critiques the significance of future AI in imaginative terms, we bring together a wide constituency that crosses many boundaries of interest and specialism. Discussions around specific texts and films will also give rise to consideration of the impact of AI on religious beliefs and practice.

This conference convenes a broad group of experts including SF authors, film and tv writers, AI researchers, futurists, academic researchers in philosophy, theology, literature, gender studies, film studies, SF, creative writing, as well as journalists, critics and publishers.

Filming

Please note that the sessions will be filmed and some of the footage may be used in a short film about the conference. A report will be made available, written by conference rapporteur is Sally Adee. If you have any questions about the film or the report please contact Jonathan Cornwell.

Acknowledgements

The AI and the Future of Humanity Project will run for two years from August 2017. It is an initiative of the Science & Human Dimension Project, based at Jesus College, Cambridge since 1990. We thank the Master and Fellows of Jesus College, Cambridge for their support of this project. For their contributions to the AI in SF film and literature conference we thank Prof Adam Roberts, Rev'd Dr Paul Dominiak, Dr Tim Jenkins, Dr Beth Singler, the Leverhulme Centre for the Future of Intelligence, Dr Adrian Weller, Prof Murray Shanahan, Keith Mansfield, Dr Tudor Jenkins, conference rapporteur Sally Adee, Sumit Paul-Choudhury, and the Cambridge University Science Fiction Society. This conference was made possible through the support of a grant from Templeton World Charity Foundation, Inc.. The opinions expressed here are those of the Science & Human Dimension Project and do not necessarily reflect the views of Templeton World Charity Foundation, Inc..



Science & Human Dimension Project Jesus College, Cambridge

Al in Sci Fi Film and Literature

Imagining AI - the View from Mythology, Fantasy, and SF

Science & Human Dimension Project Jesus College, Cambridge

Day 1: Thursday, 15 March 2018

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

11.30-11.40Welcome and Introduction - Frankopan Hall, West CourtJohn CornwellScience & Human Dimension Project - AI and the Future of HumanityProf Murray ShanahanDefining terms in AI

11.45-13.00 Session 1 - Monsters and AI: Frankenstein and the Golem

Chair: John Cornwell Prof Adam Roberts Frankenstein Dr Yaron Peleg The Golem: from Legend to Metaphor

13.00-13.50 Lunch - Dining Room, West Court

13.55-15.15 Session 2 - Explaining the Future through Sci-Fi

Chair: Sumit Paul-Choudhury Keith Mansfield Bostrom's Superintelligence through Sci-Fi Film: Colossus - The Forbin Project; Ex Machina; and Star Trek - The Motion Picture Dr Patrick Crogan Techno-Science Fictions of AI: Swarming AI in Visions of Future War Christopher Markou Sci-Fi Depiction of Future and Alternative Legal Systems

15.20-16.35 Session 3 - Anxiety, Apocalypticism and AI

Chair: Professor Kathleen RichardsonDr Beth SinglerRobopocalypses: Introducing an Anthropology of AnxietyAnne CharnockWhy I'm afraid of the Super Machine - Sci-Fi author inspirationProf Olle HäggströmApocalyptic Scenarios: Science Fiction or Real Risk?

16.35-17.00 Tea - Bawden Room

17.05-18.15 Session 4 - AI in Film and TV

Chair: Zoe Wible

Dr Hallvard Haug Artificial Intelligence in Film and Television - an Overview Emma Reeves and David Chikwe Fearing and embracing "the other"; creating an AI for BBC Children's TV

- 18.50-19.30 Drinks Prioress's Room, Cloister Court
- 19.30 Dinner Upper Hall
- 21.15 Screening of Forbidden Planet Frankopan Hall

The College bar in West Court will be open from 6.00pm-11.00pm



Science & Human Dimension Project Jesus College, Cambridge

Al in Sci Fi Film and Literature

Imagining AI - the View from Mythology, Fantasy, and SF

Science & Human Dimension Project Jesus College, Cambridge

Day 2: Friday, 16 March 2018

08.45-09.15 Registration - Bawden Room, West Court, Jesus College Refreshments served

09.15-10.00 Session 5 - Forbidden Planet and AI

Chair: Prof Adam Roberts Prof Murray Shanahan Kerry Shanahan

10.00-10.50 Session 6 - Machine Messiah: Lessons for AI in Destination: Void Chair: Dr Tim Jenkins Dr Ron Chrisley

10.55-11.25 Break - Bawden Room

11.25-12.25 Session 7 - Leverhulme Centre for Future of Intelligence Chair: Dr Stephen Cave On the dichotomies that shape our hopes and fears for AI Dr Kanta Dihal The influence of Isaac Asimov's fictional laws of robotics on public policy

12.30-13.25 Lunch - Dining Room, West Court

13.25-14.35 Session 8 - Al in Sci-Fi Author Session and Workshop

Chair: Prof Thore Husfeldt Justina Robson Life Finds A Way; creating AGI and the stories that shape the future Dr Paul J. McAuley A brief history of encounters with things that think Lavie Tidhar Greek Gods, Potemkin AI and Alien Intelligence Ian McDonald The Quickness of Hand Deceives the AI

14.35-15.25 Workshop - AGI Scenarios in Sci-Fi Prof Thore Husfeldt

15.25-15.40 Concluding Remarks

Chair: John Cornwell Dr Beth Singler Prof Kathleen Richardson Prof Murray Shanahan Prof Adam Roberts

15.40-16.15 Conference Close and Tea - Bawden Room

#SHDP



Science & Human Dimension Project Jesus College, Cambridge

Daedalus on Wheels

Jesus College, Cambridge is home to an impressive art collection, including works by John Bellany, Barry Flanagan, Antony Gormley, Richard Long, Sarah Morris, Cornelia Parker, Alison Wilding and Eduardo Paolozzi. Sir Eduardo Paolozzi (1924-2005), the Scottish sculptor and artist, had a long-standing relationship with Jesus College, where he



Daedalus on Wheels by Eduardo Paolozzi 1994 Chapel Court, Jesus College Cambridge

became an honorary fellow in 1994. Among the prints and sculpture owned by the College is the bronze Daedalus on Wheels. In this and many of his works, Paolozzi reflects on the future of humanity and our collective angst in great technological change. the face Incorporating metal, nuts, bolts and scrap to enhance humanoid limbs and facial features, Paolozzi's fascination with the relationship between humans and machines is expressed in the fusion of both - our destinies seemingly inseparable; some claim it to be a self-portrait. Paolozzi viewed advances in technology, and our reliance on it, with some scepticism. He said that his art is not intended to help people escape from the world, but rather to remind them of it.

The mythical Daedalus (literally 'cleverly wrought' in Greek) is a skilled craftsman, innovator and sculptor. He is also father of lcarus and creator of the Cretan Labyrinth, home to the Minotaur. Daedalus's inventions are infamous for their unintended and harmful consequences: lcarus falls from the sky to his death when, forgetting his father's advice, he flies too close to the sun, melting the wax in his wings; and the Labyrinth is so impenetrable as to make finding and slaying the Minotaur almost an impossibility.

Paolozzi's concerns about the relationship between humans and machines continue to vex and fascinate in equal measure. The unintended consequences of advances in Artificial Intelligence, both real and imagined, raise many questions, some of which will be discussed at the Science & Human Dimension Project conference: what can we learn about ourselves in relation to AI by exploring sci-fi narratives, and to what extent will the ambitions of AI match, challenge, demoralise, or perhaps even aid the human faculties of consciousness, imagination and agency?

Daedalus on Wheels is on permanent display in Chapel Court, Jesus College.

Artificial Intelligence in Sci-Fi Film and Literature

Imagining AI - the View from Mythology, Fantasy, and SF

Science & Human Dimension Project Conference

15-16 March 2018 Jesus College, Cambridge

Speaker bios

Dr Stephen Cave is Executive Director of the Leverhulme Centre for the Future of Intelligence, Senior Research Associate in the Faculty of Philosophy, and Fellow of Hughes Hall, all at the University of Cambridge, UK. Stephen earned a PhD in philosophy from Cambridge, then joined the British Foreign Office, where he served as a policy advisor and diplomat. He has subsequently written and spoken on a wide range of philosophical and scientific subjects, including in the New York Times, The Atlantic, and on television and radio around the world. His research interests currently focus on the nature, portrayal and governance of Al.

Anne Charnock's writing career began in journalism and her reports appeared in New Scientist, The Guardian, Financial Times, International Herald Tribune and Geographical. She has written three novels and a novella. Her latest novel Dreams Before the Start of Time imagines the impacts of advances in fertility science, and is currently shortlisted for the British Science Fiction Association 2017 Award for Best Novel. Her recent novella The Enclave is shortlisted for the BSFA 2017 Award for Best Short Fiction. This novella is written in the same world as her debut novel A Calculated Life, which was a finalist for the 2013 Philip K. Dick and The Kitschies Golden Tentacle Awards. These works imagine a near-future when the state decides who can, and cannot be, cognitively enhanced. Anne was educated at the University of East Anglia where she studied Environmental Sciences, and at The Manchester School of Art where she gained a MA in Fine Art. Anne is 'interviewer-inresidence' for the Ada Lovelace Conversations, a collaboration between the Arthur C. Clarke Award and the Ada Lovelace Day.

http://www.annecharnock.com http://www.twitter.com/annecharnock

David Chikwe Prior to becoming a screenwriter, David was an award-winning TV drama and film producer with over twelve years industry experience. David co-created and produced 36 episodes of BAFTA nominated and RTS-winning series Eve, and continues to produce.

Dr Ron Chrisley 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 Patrick Crogan is Associate Professor of Digital Cultures in Arts and Cultural Industries at UWE and teaches across several media programmes. He wrote Gameplay Mode: War, Simulation and Technoculture (2011) and has published work on video games, cinema, digital animation and simulation, drones, automation and AI in journals and collections. An expert in the work of philosopher of technology, Bernard Stiegler, Patrick has published extensively on Stiegler's film and media theory. He has led funded collaborations working on video game experimentation and supporting independent game production.

Dr Kanta Dihal is a Postdoctoral Research Associate and the Research Project Coordinator of the Leverhulme Centre for the Future of Intelligence. As a researcher on the AI Narratives project she explores the public understanding of artificial intelligence as constructed by fictional and nonfictional narratives. Kanta's work intersects the fields of science communication, literature and science, and science fiction. She recently obtained her DPhil in science communication at the University of Oxford: in her thesis, titled 'The Stories of Quantum Physics,' she investigated the communication of conflicting interpretations of quantum physics to adults and children.

Dr Sarah Dillon is University Lecturer in Literature and Film in the Faculty of English at the University of Cambridge. She is a feminist scholar specialising in the analysis of intersection and interconnection, especially between contemporary literature, film, philosophy and science. She is author of The Palimpsest: Literature, Criticism, Theory (2007), Deconstruction, Feminism, Film (2018), and editor of David Mitchell: Critical Essays (2011) and Maggie Gee: Critical Essays (2015). Sarah is presently Chair of the British Association for Contemporary Literary Studies and a Senior Research Fellow at the Leverhulme Centre for the Future of Intelligence. There, she is co-Project Lead for the AI Narratives project, in collaboration with the Royal Society. She is co-editor (with Stephen Cave and Kanta Dihal) of AI Narratives: Imagining Intelligent Machines (forthcoming 2020) and is researching a monograph provisionally entitled Artful Intelligence: AI, Narrative and Knowledge. Sarah is committed to engaging with audiences outside of the academy. In 2013 she was selected as an AHRC BBC Radio 3 New Generation Thinker and now broadcasts regularly on BBC Radio 3 and BBC Radio 4. She writes and presents the BBC Radio 3 documentary series, Literary Pursuits.

Prof Olle Häggström earned his Ph.D. in mathematical statistics in 1994 at Chalmers University of Technology, where since 2002 he serves as professor of mathematical statistics. He is currently also at the Institute for Future Studies in Stockholm. He is elected member of the Royal Swedish Academy of Sciences, and served as chairman of the Swedish Mathematical Society during 2005-2007. The bulk of his research qualifications are in probability theory and statistical mechanics, with more than 80 scientific publications in these fields. During the last decade, however, he has broadened his research interests considerably, towards statistical inference, risk theory, climate science, artificial intelligence, futurology and several branches of philosophy including the philosophy of science. He is the author of four books, most recently the wide-ranging Here Be Dragons: Science, Technology and the Future of Humanity (Oxford University Press, 2016). He has served as invited speaker at more than 70 conferences, and on panel discussions at the European Parliament in Brussels twice. He is a prolific popularizer of science (in magazines and on blogs), and appears frequently in Swedish radio, television and newspapers to discuss mathematics, science and emerging technologies.

Dr Hallvard Haug is a writer and academic with an interest in the intersection between culture, society, science and fiction. As well as researching the history of human enhancement technologies, he has written on topics such as the early-response team responsible for preparing the recently deceased for cryonic storage and the people working out the knotty politics of running a lunar colony. He is currently working on a book on transhumanism.

Prof Thore Husfeldt is professor of theoretical computer science at Lund University, Sweden and associate professor at IT University of Copenhagen, Lund, and a leading researcher in algorithms. He is a core researcher in the center for Basic Algorithms Research, Copenhagen (BARC, barc.ku.dk) and a council member of the European Association for Computer Science.

Outside of his research on the mathematical foundations of algorithms, Thore is a prolific teacher and active science communicator. He hosts a video podcast Cast IT at castit.itu.dk, which contains long-form conversations with academics on foundations and applications of information technology, including its societal impact. He regularly gives public lectures on artificial intelligence, including its concepts, limitations, applications, and current and future threats. Thore has been a voracious consumer of speculative fiction since before it was cool. He lives in Lund, Sweden, with his wife and three children.

Dr Timothy Jenkins is Reader in Anthropology and Religion at the University of Cambridge. He was trained at The Oxford Institute of Social Anthropology and has carried out fieldwork in Britain and France. His interests include theoretical approaches to the social sciences, European ethnography, especially concerning politics and language, and the study of religion, particularly moral uses of science discoveries. He spent 2016-17 at the Centre of Theological Inquiry in Princeton on the project funded in part by NASA, 'The Societal Implications of Astrobiology'. He is the author of Religion in English Everyday Life, Berghahn 1999, The Life of Property, Berghahn 2010, and, most recently, Of Flying Saucers and Social Scientists, Palgrave Macmillan 2013.

Keith Mansfield is a writer, publisher and broadcaster. Having studied mathematics and physics at Trinity College, Cambridge, he began a career in science publishing with Robert Maxwell's Pergamon Press. Along the way he was the editor for AI textbook publishing at Pearson Education while at Oxford University Press he created a book programme in existential risk, including publishing Nick Bostrom's Superintelligence.

His Johnny Mackintosh series of children's sci-fi books (published by Quercus) include highly intelligent spaceship minds and androids, as well as a sentient quantum computer.

Mansfield has also been the book publisher at the British Film Institute and developed various shows for ITV, most recently It's Not Rocket Science. He is currently writing books about the Fermi paradox and also colonizing Mars, while founding Herschel Publishing which will focus on science and science fiction books.

Christopher Markou is a PhD Candidate in the Faculty of Law, University of Cambridge, and a member of Jesus College. His doctoral research, entitled 'Law and Artificial Intelligence: A Systems Theoretical Analysis', is supervised by Professor Simon Deakin (Peterhouse). Christopher's research is generously funded by a Doctoral Fellowship provided by the Social Sciences and Humanities Research Council of Canada. He is on the Legal Expert Committee of Responsible Robotics, the Advisory Board of ADA-AI, and his writing on issues related to law, technology and society has been featured in Newsweek, Scientific American, The Independent and others.

Paul McAuley worked as a research biologist (plant/animal symbiosis) and a university lecturer in botany before becoming a full-time writer. He is the author of more than twenty novels, several collections of short stories, a Doctor Who novella, and a BFI Film Classic monograph on Terry Gilliam's film Brazil. His fiction has won the Philip K Dick Memorial Award, the Arthur C. Clarke Award, the John W. Campbell Memorial Award, the Sidewise Award, the British Fantasy Award and the Theodore Sturgeon Memorial Award. His latest novel, Austral, a novel about post-global warming Antarctica, was published in 2017. http://www.unlikelyworlds.co.uk/

Twitter: @UnlikelyWorlds

Ian McDonald is a writer of (mostly) science-fiction, living just outside Belfast in Northern Ireland, and now viewing the future with some apprehension. His first novel, Desolation Road was published in 1988 and was the winner of the Locus Best First Novel Award and nominated for the Arthur C Clarke Award. He's been translated into fifteen languages and been nominated for every major in the genre and won the prestigious Hugo for Best Novellete in 2007 for his story The Djinn's Wife. Ian also worked in television for 16 years, in programme development from documentary to animation to children's television. His most recent publications are Luna: New Moonand Luna: Wolf Moon, from Gollancz UK and Tor US, part of the Luna series set, inevitably, on an industrially developed moon. Volume three is in the works, and forthcoming in April 2018 is a novella Time Was from Tor.com.

Sumit Paul-Choudhury is strategy director and editor emeritus at New Scientist, the world's most popular science weekly, having served as editor-in-chief from 2011-2017. Sumit trained as a physicist at Imperial College before turning his hand to journalism, working in London and New York; he spent fifteen years writing about finance and technology before joining New Scientist in 2008. In addition to the day job, he was editor-in-chief of Arc, an acclaimed digital publication dedicated to the future, between 2012 and 2014; and in 2016 he served as the founding creative director for New Scientist Live, the world's most exciting festival of ideas and discovery. He is currently on the Sloan Fellowship at London Business School, serving as a judge on this year's Wellcome Book Prize, is a Fellow of the RSA, the founder of Chantepleure Productions and a founding member of the Bishopsgate Experimental Noise Theatre, among other side projects. He lives, works and fails to sleep in London.

Yaron Peleg is Kennedy-Leigh Reader in Modern Hebrew Studies at the University of Cambridge. His publications include, Directed by God, Jewishness in Contemporary Israeli Film and Television (2016), Israeli Culture Between the Two Intifadas (2008), Orientalism and the Hebrew Imagination (2005), and Derech Gever (2003) a study of Homoeroticism in Hebrew Literature. He is also co-editor of an anthology of articles on contemporary Israeli cinema, Identities in Motion (2011). Dr. Peleg has also published articles on a number of topics, including literary critiques, which examine the concept of Land in modern Hebrew prose, attitudes toward militarism, homoeroticism in biblical as well as more modern Hebrew literature and various articles about Israeli cinema that focus on gender, masculinity, ethnicity and religiosity.

Emma Reeves TV work includes The Worst Witch (Lead Writer), Eve (Lead Writer and co-creator), The Dumping Ground, Young Dracula, Tracy Beaker Returns, Belonging, The Murder of Princess Diana, Spirit Warriors and Doctors. Theatre work includes CAMEO award winning and Olivernominated Hetty Feather, West End adaptions of Carrie's War, Little Women and Cool Hand Luke. She has also written adaptations of Mary Barton, 1984, Jekyll & Hyde, Sherlock Holmes and Anne of Green Gables, and two plays for younger audiences, Snow Child and Ugly Duckling. Audiowork includes plays for Radio Four, Radio Three, BBC Wales and Big Finish including Torchwood: Forgotten Lives which won Doctor Who Magazine's Favourite Audio Drama award 2016. She has won Best Children's TV Episode at The Writers Guild of Great Britain Awards twice, and been nominated on two other occasions. Her work has won a national RTS award, an RTS Scotland award, an RTS Scotland award, been nominated for two BAFTA awards and two Broadcast awards. In 2017, both Eve and The Worst Witch were nominated for Best Witch won. Emma has a first class Honours degree in English from Magdalen College, Oxford, and an MA in Creative Writing from the University of East Anglia.

Prof Kathleen Richardson is Professor of Ethics and Culture of Robots and AI at De Montfort University in Leicester. She is also founder of the Campaign Against Sex Robots and author of An Anthropology of Robots and AI: Annihilation Anxiety and Machines (2015) and Challenging Sociality? An Anthropology of robots, autism and attachment (2018) and Sex Robots: the end of love (2018).

http://www.dmu.ac.uk/about-dmu/academic-staff/technology/kathleen-richardson/kathleen-richardson.aspx

Prof Adam Roberts is Professor of Nineteenth-Century Literature and Culture at Royal Holloway, University of London, and the author of seventeen science fiction novels. His Palgrave History of Science Fiction (2nd ed 2017) is a comprehensive account of the genre from classical Greece to the twenty-first century. He is presently working on a literary biography of H G Wells and a new novel.

Justina Robson has written 12 novels and a book of short stories, many of them featuring AI and all of them concerned with the inner engineering of human beings and how they experience themselves in relation to their own technologies. Awards shortlists, Transformers Lore and philosophical debates have ensued at various points between 2000-present. She is happy to attend this conference. http://justinarobson.co.uk/

Kerry Shanahan is a final-year BA student in English Literature at University College London.

Prof Murray Shanahan is Professor of Cognitive at Imperial College London and a Senior Research Scientist at DeepMind. As well as many scientific papers he has published several books, including "Embodiment and the Inner Life" (Oxford University Press, 2010) and "The Technological Singularity" (MIT Press, 2015). He was scientific advisor to the film Ex Machina, and regularly appears in the media to comment on artificial intelligence and robotics.

Dr Beth Singler is the research associate on the 'Human Identity in the Age of Nearly Human Machines' project at the Faraday Institute, St Edmund's College, Cambridge. Her anthropological research considers the social, ethical, and religious implications in developments in Artificial Intelligence and robotics, as well as popular and religious re-imaginings of science and technology. Pain in the Machine, a short documentary on whether robots should feel pain made as a part of the Faraday project won the AHRC Best Research Film of the Year Award, 2017, and is the first in a four part series made by Beth and Little Dragon Films on the rise of the thinking machines and their implications for human identity and society. She is also an associate research fellow at the Leverhulme Centre for the Future of Intelligence, University of Cambridge. In 2017 Beth appeared at the Hay Festival (2017) as one of the Hay 30 to watch, and she has also spoken at the Cheltenham Science Festival, the Science Museum, the Barbican, and on BBC Click, BBC Radio 4's Start the Week, the Today Programme, and the Sunday Programme. In 2017 she was one of the Evening Standard's Progress 1000. In October 2018 she will join Homerton College, Cambridge as the junior research fellow in Al.

Lavie Tidhar is the author of the Jerwood Fiction Uncovered Prize winning and Premio Roma nominee A Man Lies Dreaming (2014), the World Fantasy Award winning Osama (2011) and of the critically-acclaimed and Seiun Award nominated The Violent Century (2013). His latest novel is the Campbell Award winning and Locus and Clarke Award nominated Central Station (2016). He is the author of many other novels, novellas and short stories.

Zoe Wible is a PhD student in Film at the University of Kent. Her research interests include sciencefiction and cognitive film theory. Following her master's dissertation on the reception of androids in contemporary television show Westworld, she is now researching the relationship between imaginary creatures and spectator engagement in visual narrative media. She also draws on recent developments in interactive media and forms of engagement, including video games and online fandom spaces. The provisional title for her thesis is: "Monster schemas and the space of possible minds: A cognitive approach to science fiction characters in contemporary cinema".

Science & Human Dimension Project

John Cornwell is Director, Science & Human Dimension Project and Fellow Commoner, Jesus College, Cambridge. In 1990, after 12 years on the editorial staff of The Observer, he was elected Senior Research Fellow and Director of the Science and Human Dimension Project at Jesus College, Cambridge. In this role he has brought together many scientists, philosophers, ethicists, authors and journalists to debate a range of topics in the public understanding of science. His edited books include Nature's Imagination, Consciousness and Human Identity, and Explanations (all OUP); 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; and received the Independent Television Authority-Tablet Award for contributions to religious journalism (1994). His journalism has been published in a variety of outlets including 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, especially on culture, science and religion, including "Hard Talk", "Choice", "Start the Week," "The Moral Maze", "Today" (debate with Richard Dawkins); "Beyond Belief", "Thought for the Day", "Sunday", and various programmes in the BBC's World Service.

Jonathan S. Cornwell has a background in academic and digital publishing in Europe and the Middle East. From 2010-17 he was co-director of the Rustat Conferences and is currently executive director of the Science & Human Dimension Project (SHDP). He has produced conferences and edited reports on a wide range of topics including AI, cybersecurity, blockchain and bitcoin, energy security, food security, inequality, north-south divide, future of work, and ageing. He also works with curators and patrons to produce exhibitions, including *Houghton Revisited: Masterpieces from the Hermitage; James Turrell: Lightscape;* and Beyond Beauty: Transforming the Body in Ancient Egypt. He studied at UCL, Cambridge and Imperial.

Sally Adee is an award-winning science and technology writer and editor. She was a technology features and news editor at New Scientist for seven years, writing and commissioning articles about medical technology, artificial intelligence, and the Venn diagram of the human mind and the machines we create. Before that she was on the microchips beat at *IEEE Spectrum* magazine in New York. She has received awards from the National Press Club and BT, and has reported rom China,

DARPA headquarters, and the Estonian cloud. In her spare time Sally writer speculative fiction and nonfiction at *The last Word on Nothing*, an independent science blog dubbed a "must follow" by Wired. Sally is currently working on a book about an obscure fight between Luigi Galvani and Alessandro Volta. Everything she learned, she learned from her father's dog eared copies of Analog magazine. Sally is the conference rapporteur for SHDP's Al in Sci-Fi Film and Literature Conference.

Dr Tudor Jenkins' background is in Artificial Intelligence and Artificial Life, where he used a situated adaptive behavioural approach to understand conditions under which grammars can evolve in social agents controlled by neural networks. Tudor is currently interested in how adaptive systems can be better used to understand and improve human performance in sport. He holds a PhD in AI from Sussex University and did research at the École Normale Supérieure, Paris. He is an adviser to Science & Human Dimension Project.

Dannielle Cagliuso Paramedic, historian and philosopher of science, post-graduate student, Faculty of Law, University of Cambridge Faculty of Law; assisting Science & Human Dimension Project at the AI in Sci Fi conference.

Olivia Morley Physics student at the Cavendish Laboratory, University of Cambridge; President, Cambridge University Sci-Fi Society; assisting Science & Human Dimension Project at AI in Sci-Fi conference.

Artificial Intelligence in Sci-Fi Film and Literature

Imagining AI - the View from Mythology, Fantasy, and SF

Science & Human Dimension Project Conference

15-16 March 2018 Jesus College, Cambridge

Participants list

- Sally Adee SHDP Conference Rapporteur; author, science journalist, fmr Technology Features Editor, New Scientist
- Azeem Azhar entrepreneur, editor of AI newsletter The Exponential View
- Prof Jean Bacon Professor of Computer Science, Fellow, Jesus College; Computer Laboratory, University of Cambridge
- Tony Ballantyne SF author of Dream London, Penrose and Recursion series
- Chris Beckett lecturer; SF novelist and short story writer; winner Philip K. Dick Award
- Andrew Brown editor, broadcaster, journalist, The Guardian
- Prof Andrew Briggs Professor of Nanotechnology, University of Oxford; Fellow, St Anne's College

Dannielle Cagliuso Paramedic, historian and philosopher of science, post-graduate student, Faculty of Law, University of Cambridge Faculty of Law; assisting Science & Human Dimension Project

- Dr Aifric Campbell SF author; creative writing, Imperial College London
- Geoff Carr Science & Technology Editor, The Economist
- Dr Stephen Cave Executive Director, Leverhulme Centre for the Future of Intelligence, Cambridge
- Anne Charnock SF author and Philip K. Dick Award Finalist for Calculated Life
- David Chikwe author, producer, co-creator and writer of BBC TV series Eve

Dr Ron Chrisley Reader in Philosophy, Informatics, Sackler Centre for Consciousness Science, Centre for Cognitive Science, Evolutionary & Adaptive Systems, Sussex University

Dr Dave Clements Senior Lecturer, Physics, Imperial College London; SF author

Dave Coplin author, Rise of the Humans; Founder and CEO, the Envisioners; former Chief Envisioning Officer, Microsoft

John Cornwell Founder and Director, Science & Human Dimension Project, Jesus College, Cambridge; author, journalist

Jonathan S. Cornwell Executive Director, Science & Human Dimension Project

Dr Claire Craig Director of Science Policy, The Royal Society

Prof Patrick Crogan Professor of Digital Cultures in Arts and Cultural Industries, University of the West of England UWE

Dr Adrian Currie Philosophy of Science, Centre for the Study of Existental Risk, University of Cambridge

Tracy Darnton YA writer, lawyer; member, Jesus College, Cambridge

Rhidian Davis Programme Director, British Film Institute

Rev'd Dr Andrew Davison Starbridge Lecturer in Theology and Natural Sciences, University of Cambridge

Dr Matthew De Abaitua SF and non-fiction author; Head of Creative Writing, Department of Literature, Film and Theatre Studies, University of Essex

Dr Kanta Dihal Postdoctoral Research Assistant, Research Project Coordinator, Al Narratives Project, Leverhulme Centre for the Future of Intelligence, Cambridge

Dr Sarah Dillon University Lecturer in Literature and Film, Cambridge; co-Lead, Al Narratives project, Centre for the Future of Intelligence CFI and Royal Society

Rev'd Dr Paul Dominiak Dean of Chapel, Director of Studies in Theology, Religion, and Philosophy of Religion, Jesus College, Cambridge

Dr Moshe Freedman Rabbi, The New West End Synagogue, London

Dr Milica Gašic Lecturer in Spoken Dialogue Systems, Machine Intelligence Laboratory, Department of Engineering, University of Cambridge; Fellow of Murray Edwards College

Alberto Garcia PhD candidate; Research project: The Emotional Machine: Representations of Emotions in Intelligent Machines in 20th Century Science Fiction, UCSD, University of California San Diego

Vadim Grigoryan Creativity Consultant, Curator

Prof Olle Häggström Professor of Mathematical Statistics, University of Gothenburg; author, Here be Dragons: Science Technology and the Future of Humanity

Dr Marta Halina Lecturer in the Philosophy of Cognitive Science, University of Cambridge

Dr Hallvard Haug academic; fmr Research Fellow, Birkbeck, University of London, and author, Engineering humans: cultural history of the science and technology of human enhancement

Prof Stephen Heath Professor of English and French Literature and Culture, Cambridge University; Fellow, Jesus College

Dr Thore Husfeldt Professor of Computer Science, Lund University; Associate Professor IT, University of Copenhagen

Dr Timothy Jenkins Reader in Anthropology, University of Cambridge; Fellow, Jesus College

Dr Tudor Jenkins Technologist; Member, Advisory Board, Science & Human Dimension; PhD, AI, University of Sussex

Abby Kidd Sci-Fi Film Studies, University of East Anglia

Vishal Maini Research Content Manager, DeepMind

Keith Mansfield writer, publisher broadcaster; author, Johnny Mackintosh series; founder, Herschel Publishing; publisher of Nick Bostrom's Superintelligence

Christopher Markou Law and technology, Jesus College, Cambridge

Dr Paul J. McAuley SF author; botanist; Winner, Philip K. Dick and Arthur C. Clarke awards

Dr Una McCormack SF author; Lecturer in Creative Writing and Co-Director, Anglia Ruskin University Centre for Sci-Fi

Ian McDonald SF author, winner of Philip K. Dick and Hugo awards

Conrad McDonnell barrister

Dr Ken Moody Computer Science; Fellow, King's College, Cambridge

Dr Glyn Morgan Independent researcher, Science Fiction

Olivia Morley student; President, Cambridge University Sci-Fi Society; assisting Science & Human Dimension Project

Oliver Morton Science writer; Briefings Editor, The Economist

Prof John Naughton CRASSH, University of Cambridge; Professor of public understanding of technology, Open University; Observer technology columnist

Sumit Paul-Choudhury Strategy Director and Editor Emeritus, New Scientist; former Editor, Arc

Dr Yaron Peleg Kennedy-Leigh Reader in Modern Hebrew Studies, University of Cambridge; Fellow, Jesus College

Colin Ramsay Director, Little Dragon Films

Emma Reeves author, co-creator and writer of BBC TV series Eve

Prof Kathleen Richardson Professor of Ethics and Culture of Robots and AI, Centre for Computing and Social Responsibility, De Montfort University

Prof Adam Roberts SF author; Professor of 19th Century Literature, Royal Holloway University of London; Winner, BSFA Award for Best Novel

Justina Robson SF author; nominee Arthur C. Clarke, BSFA, and John W. Campbell awards

Prof Dale Russell Visiting Professor, School of Design, Royal College of Art

Dr Elisabeth Schimpfössl Lecturer in Sociology and Policy, Aston University

Prof Murray Shanahan Professor of Cognitive Robotics, Imperial College London; Senior Research Scientist, DeepMind

Kerry Shanahan English literature, University College London

Dr Beth Singler Human Identity in an age of Nearly-Human Machines Project, Faraday Institute, University of Cambridge; Centre for Future of Intelligence, Cambridge

Lindsay Taylor Policy Adviser - New Horizons (future and emerging technologies), The Royal Society

Bill Thompson technology journalist, broadcaster, critic

Lavie Tidhar SF author; Winner, World Fantasy Award, and John W. Campbell Award

Dr Liesbeth Venema Editor in Chief, Nature Machine Intelligence

Richard Watson Futurist-in-Residence, Tech Foresight, Imperial College, London

Prof Ian White Master, Jesus College, Cambridge; Van Eck Professor of Engineering, University of Cambridge

Margaret White St Faith's School; Jesus College, Cambridge

Zoe Wible Sci-Fi Literature, University of Kent; author, Story telling machines of Westworld

Prof Tim Wilkinson Professor of Engineering, University of Cambridge; Fellow, Jesus College

Prof Rowland Wymer Emeritus Professor of English, Anglia Ruskin Centre for Science Fiction and Fantasy