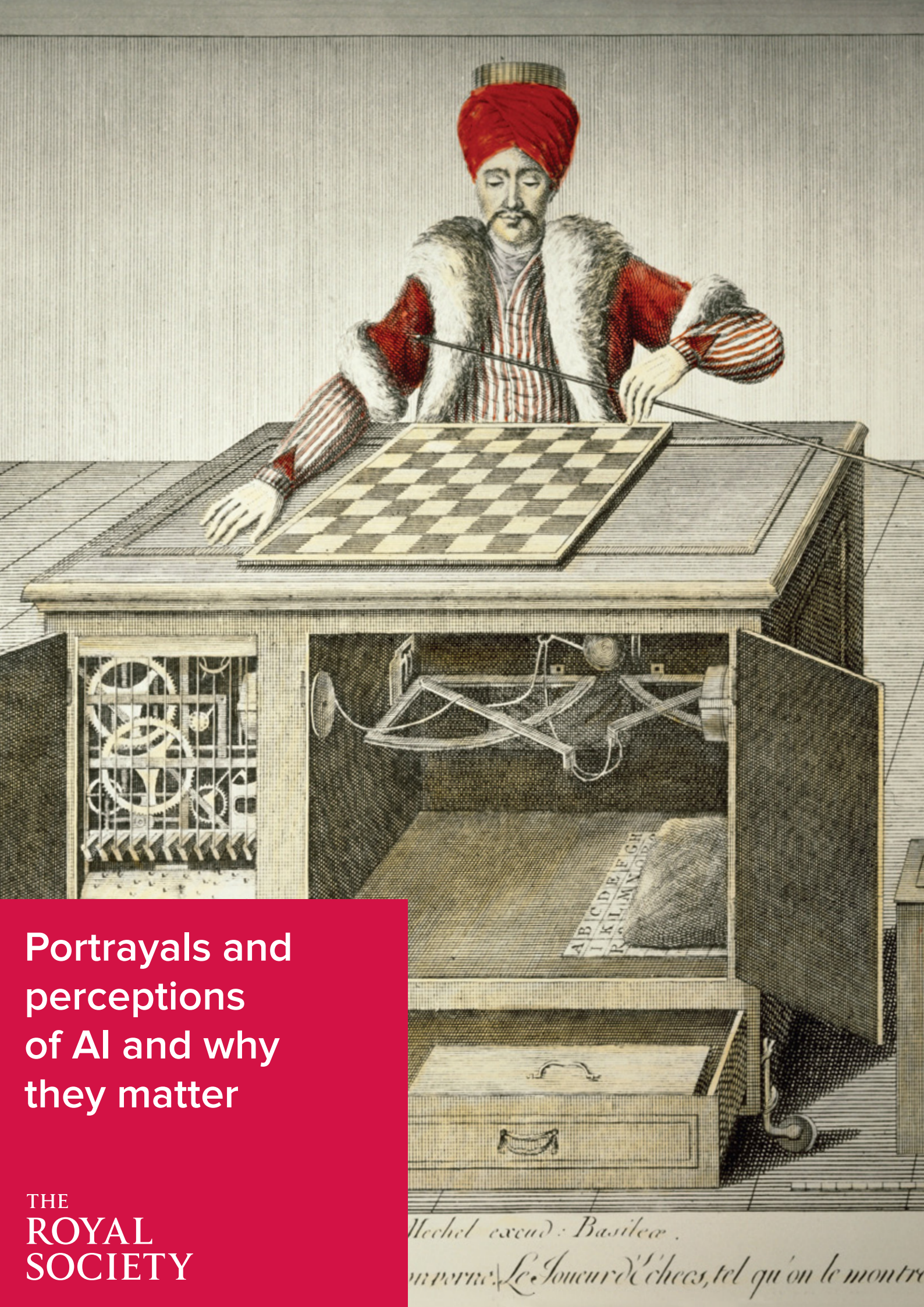


Portrayals and perceptions of AI and why they matter

THE ROYAL SOCIETY

Cover Image
Turk playing chess, design by Wolfgang von Kempelen (1734 – 1804), built by Christoph Mechel, 1769, colour engraving, circa 1780.

Contents

Executive summary

The AI narratives project – a joint endeavour by the Leverhulme Centre for the Future of Intelligence and the Royal Society – has been examining how researchers, communicators, policymakers, and publics talk about artificial intelligence, and why this matters.

This write-up presents an account of how AI is portrayed and perceived in the English-speaking West, with a particular focus on the UK. It explores the limitations of prevalent fictional and non-fictional narratives and suggests how practitioners might move beyond them. Its primary audience is professionals with an interest in public discourse about AI, including those in the media, government, academia, and industry.

Its findings have been synthesised from discussions at four workshops, held in Cambridge and London between May 2017 and May 2018, and organised by the AI narratives project. This project had its origins in questions emerging from public dialogue carried out as part of the Royal Society's report on machine learning.

Imaginative thinking about intelligent machines is ancient, reaching back at least to Homer's Iliad (c. 800 BCE). As the technologies themselves have developed, from automata to robots, and from cybernetics to today's machine learning, so have the hopes and fears associated with them. Prevalent AI narratives share dominant characteristics: a focus on embodiment; a tendency towards utopian or dystopian extremes; and a lack of diversity in creators, protagonists, and types of AI.

Narratives are essential to the development of science and people's engagement with new knowledge and new applications. Both fictional and non-fictional narratives have real world effects. Recent historical examples of the evolution of disruptive technologies and public debates with a strong science component (such as genetic modification, nuclear energy and climate change) offer lessons for the ways in which narratives might influence the development and adoption of AI technologies.

AI narratives can be very helpful; for example, in inspiring those who work in the relevant disciplines and civil, public and private sectors; and in surfacing alternative futures and enabling debates about them. But they can also create false expectations and perceptions that are hard to overturn. For those not engaged closely with the science or technology, narratives can affect perceptions of, and degrees of confidence in, potential applications and those who are developing, promoting or opposing them.

Exaggerated expectations and fears about AI, together with an over-emphasis on humanoid representations, can affect public confidence and perceptions. They may contribute to misinformed debate, with potentially significant consequences for AI research, funding, regulation and reception.

Attempting to control public narratives is neither achievable nor desirable, but present limitations may be at least partly addressed by communicating uncertainty through learning from narratives about other disruptive technologies; widening the body of available narratives, drawing in a wider range of authors and protagonists; and creating spaces for public dialogues.

1. Introduction

1.i Narratives and artificial intelligence

Narratives are an "ensemble of texts, images, spectacles, events and cultural artefacts that 'tell a story'"[1]. Whether fictional or non-fictional, narratives function in the real world. They affect individuals and collectives, and influence human action, thought and social outcomes. They have the power to either enhance or limit the potential for human flourishing. The study of narratives is essential in order to more effectively understand their functioning and critically engage with them.

This write-up focuses on narratives about AI: that is, ways in which this technology is portrayed and perceived. These can be fictional (speculations in popular contemporary science fiction, and the long history of imaginative thinking about intelligent machines) or non-fictional (science communication about AI, and media coverage of AI science and technology and its effects). Narratives in each category might be considered more or less plausible. They combine to create a narrative ecosystem around AI that influences its research, reception and regulation. The purpose of the AI narratives project was to investigate that ecosystem in order better to understand what narratives there are, the influence and consequences that they have (positive or negative), and ways in which the ecosystem could adapt and evolve. It was motivated by the identification of: a disconnect between prevalent narratives and the state of the science; a prevalence of narratives of fear; and a lack of diversity in the producers of AI narratives, and the types of people and of AI represented in them.

The term 'AI' was coined in 1955[2]. 'Artificial' means made by human skill. The definition of 'intelligence' is more contested. Legg and Hutter provide over seventy different definitions of the term[3]. This report adopts Margaret Boden's definition of intelligence – that it describes "the sorts of things that minds can do"[4], in particular, the application of those psychological skills that are used by animals for goal attainment. The umbrella term 'AI' is employed here whilst recognising that it encompasses a wide range of research fields, including computer science, engineering, and cognitive science. The term 'AI narratives' is employed even more broadly to include portrayals of any machines (or hybrids, such as cyborgs) to which intelligence has been ascribed, which can include representations under terms such as robots, androids or automata.

1. Bal, M. 2009 *Narratology: Introduction to the Theory of Narrative*, Toronto, Canada: University of Toronto Press
2. McCarthy, J., Minsky, M., Rochester, N., and Shannon, C. 1955 A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, available at: https://aaai.org/ojs/index.php/aimagazine/article/view/1904 [accessed 8 October 2018]
3. Legg, S. and Hutter, M. 2007 Universal intelligence: A definition of machine intelligence *Minds and Machines*, 17 (4): 391 - 444
4. Boden, M. 2016 *AI: Its nature and future*, Oxford, UK: Oxford University Press

1.ii The AI narratives project

In 2017 – 2018, the AI narratives project held four workshops gathering together more than 150 experts from a range of academic fields (AI scientific research, literary and film studies, anthropology, gender studies, history and philosophy of science, science and technology studies, science communication studies), and individuals from sectors outside of academia (journalists, policy-makers, science communicators, members of business and industry, and politicians) (Annex A).

The first workshop explored which narratives around intelligent machines are most prevalent, and their historical roots. The second workshop investigated what could be learnt from how other complex, new technologies were communicated and the impact of this. The third workshop examined how narratives are shaping the development of intelligent technology and the role that the arts and the media play in relation to the challenges and opportunities of AI. The fourth workshop debated these findings with practitioners including science communicators and creative artists. It also investigated how interventions in this space might disseminate academic research in order to influence public debate and policy discourse, to ensure that both are well-founded and well-informed.

This write-up presents the main insights and conclusions of those workshops. Where material that was presented or discussed in depth has been published separately, this document may refer to the published source, but it does not represent a systematic review or synthesis of the publications in any of the disciplines represented. It is aimed at professionals with an interest in public discourse on AI, including those in the media, government, academia and industry.

The project, a collaboration with the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, has its origins in questions emerging from two projects by the Royal Society. These were a public dialogue carried out as part of the Society's report on machine learning and its work with the British Academy on governance of data and its uses[5].

5. The Royal Society 2017 *Machine learning: the power and promise of computers that learn by example*, available at www.royalsociety.org/machine-learning; and The Royal Society and British Academy 2017 Data management and use: governance in the 21st century, available at https://royalsociety.org/topicspolicy/projects/data-governance/

2. AI narratives

This Section looks at some of the prevalent characteristics of fictional and non-fictional narratives of intelligent machines and draws primarily on Workshops 1 and 3[6]. The implications of these prevalent characteristics are addressed in Section 4.

2.i A very brief history of imagining intelligent machines

Although the notion of intelligent machines is currently enjoying an explosion of coverage, it is an ancient one. The oldest known story of something like AI can be found in Homer's *Iliad*, dating from roughly the eighth century BCE. Made by Hephaestus, the god of smithing, the machines were "attendants made of gold, which seemed like living maidens. In their hearts there is intelligence, and they have voice and vigour"[7]. They appear as faithful servants to their crippled master. Other legends attributed similar technological wonders to Hephaestus, such as Talos, a great bronze automaton that patrolled the shores of Crete, throwing stones at pirates and invaders – the first killer robot.

Machines that imitated humans grew in sophistication and popularity in the following centuries. The book *Automata* by Hero of Alexandria from the first century CE explains how to make an entirely mechanical puppet play, alongside other wonders designed to make temple-goers believe they were seeings acts of the gods. But with the declining influence of Greece, the Latin West entered a long period – perhaps a thousand years – in which the skills of automaton-making were lost, along with the hopes associated with them[8]. Until the late 13th century, the mechanical arts were preserved mostly in the Byzantine and Islamic worlds, and so associated by western Europeans with foreignness, and regarded with wonder but also suspicion.

When intelligent machines were rehabilitated in the Latinate Christian imagination, it was first in the old form of loyal retainers, such as copper knights guarding secret gateways. But the fears in today's narratives of inhumanity, obsolescence, alienation, and uprising, addressed in Section 2.iii, soon began to emerge. For example, scholars such as Roger Bacon and Robert Grosseteste were rumoured to have created a bronze head that could answer any question – a proto-Siri, perhaps. But these stories end badly, with mishaps and the destruction of the oracle, sometimes by a terrified layperson. The moral is that the creation of AI is an act of Promethean hubris; that such divine power should not belong to mortals.

This association with hubris has never left the AI project, but other themes have also come to the fore as the technology itself has developed. The second half of the seventeenth century, through to the early nineteenth, saw the heyday of automata in Europe. In this period, master craftsmen built astonishing marvels of art-imitating-life, such as Jacques de Vaucanson's famous flute player. Though neither genuinely intelligent nor autonomous, these machines suggested that lifelike androids might be within reach. With this came new fears of transgression and deceit. In E.T.A. Hofmann's 1816 short story *The Sandman*, for example, the protagonist Nathanael is bewitched by the beauty of a maiden called Olimpia. When, after much wooing, he discovers she is an automaton, he is driven to insanity and suicide.

6. A summary of the project's workshops is in the annex.
7. *Iliad*, ll. 18.417-421
8. Truitt, E. 2015. *Medieval Robots: Mechanism, Magic, Nature, and Art*, Pennsylvania, US: University of Pennsylvania Press

Although a narrative concerning a biological rather than a mechanical creature, Mary Shelley's *Frankenstein* (1818) provides the paradigmatic narrative of humankind's unnatural creations rising up against us. The "Frankenstein complex", as Isaac Asimov called it, has become a staple of AI fiction of the twentieth and twenty-first centuries. In the very work in which the term 'robot' was coined – Karel Čapek's 1920 R.U.R. (*Rossum's Universal Robots*) – the creations rebel against their masters and destroy them. This story of rebellion has been retold many times since, including in some of the most iconic and instantly recognisable portrayals of intelligent machines, such as in the Terminator film franchise and most recently the *Westworld* TV series.

The greatest density of fictional narratives exploring artificial intelligence can be found subsequent to the coinage of the term in 1955, on page and on screen. Such narratives provide a rich source of imaginative thinking about AI in relation to a range of issues[9]. These include explorations of AI in relation to control, immortality, parenting, consciousness, value alignment, cybernetworks, distributed intelligence, sex and gender, war and autonomous weapons, enslavement and governance. Many of these narratives are dystopian, some are utopian, some involve elements of both. The workshop discussions on this topic focused on a number of these issues in relation to the most prevalent fictional and non-fictional AI narratives, as summarised in the following sections on embodiment and extremes.

2.ii Embodiment

As the short history above shows, there is a strong tendency in fictional narratives in particular to conceive of intelligent machines as taking humanoid form. The anthropomorphisation of AI in the popular imagination can be accounted for in a number of ways. First, the widespread belief, at least in the West, that humans are the most intelligent animals means that the human becomes the paradigm for intelligent beings. Therefore when humans imagine other intelligent beings, these imaginings tend to take humanoid form. This is true of many visions of gods, angels and demons, and of intelligent machines. Second, such machines are often conceived of as doing human

labour: Hephaestus's "attendants made of gold" do the work otherwise done by human servants; C3PO from *Star Wars* does the work of a human translator and diplomat. It is therefore understandable that they are represented as metal versions of the people they are designed to imitate. Third, visual storytelling in particular – both film and television – requires bodies, and storytelling in general tends to privilege human actors enacting human dramas. The simplest way in which machine intelligences can be included in such dramas is therefore to take human form. It is easy for the viewer to identify with the robot protagonists of the TV series *Westworld*, for example, because they are in fact human actors expressing human emotions in recognisable plots of escape and self-discovery.

One consequence of this anthropomorphism is that AI systems are frequently gendered: their physical forms are often not androgynous, but have the stereotypical secondary sexual characteristics of either men or women. Indeed, they are often hyper-sexualised: they have either exaggeratedly muscular male bodies and aggressive tendencies, like the T-800 Terminator, or conventionally beautiful female forms such as Ava in *Ex Machina*.

9. Cave, S., Dihal, K., & Dillon, S. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (Oxford: Oxford University Press, forthcoming 2020).
10. Wiener, N. 1948 *Cybernetics: Or Control and Communication in the Animal and Machine*, Cambridge, US: MIT Press

Despite this tendency to anthropomorphism, there is nonetheless a notable sub-genre of narratives that portray artificial intelligence in ways that are not embodied. In E.M. Forster's short story *The Machine Stops* (1928) humanity is dependent on a totalised, distributed AI system that is worshipped – until it fails. Robert Heinlein's 1966 novel *The Moon is a Harsh Mistress* has an AI that resides in a computer mainframe, and which can manifest itself in a range of personalities, both male and female. Iain M. Banks' *Culture* novels provide the most sustained fictional imagining of societal governance by benevolent distributed AI, the Minds.

While in fictional narratives robots are often humanoid, robotics itself, at least since the work on cybernetics in the 1950s, has also been inspired by the capacities of non-human animals. This can be seen in the tortoises Elmer and Elsie by William Grey Walter from 1948, through to whisker-inspired robotic sensors today[10]. Ted Chiang's 2010 novella *The Lifecycle of Software Objects* provides a sophisticated exploration of what it might mean to have a relationship with an artificial intelligence whose virtual avatar, and occasional robotic body, is animal rather than humanoid.

2.iii Extremes and control

Popular portrayals of AI in the English-speaking West tend to be either exaggeratedly optimistic about what the technology might achieve, or melodramatically pessimistic. The grand hopes for AI might stem in part from the perception that it is a kind of master technology, as it amplifies the cognitive powers that humanity has deployed in all its achievements to date. For example, in 2014, the eminent scientists Stephen Hawking, Stuart Russell, Max Tegmark and Frank Wilczek wrote: "The potential benefits are huge; everything that civilisation has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools that AI may provide, but the eradication of war, disease, and poverty would be high on anyone's list. Success in creating AI would be the biggest event in human history"[11].

One way of framing public perception around AI presented at the AI narratives workshops was according to the hopes and fears it represents[12]. The extreme hopes, which are expressed both in fiction and non-fiction, include AI solving ageing and disease so that humans might lead vastly longer lives; freeing humans from the burden of work; gratifying a wide range of desires, from entertainment to companionship; and contributing to powerful new means of defence and security. The extreme fears around AI represent the flipsides of these hopes, and include AI leading to humans losing their humanity; making humans obsolete; alienating people from each other; and enslaving or destroying humans.

A core theme to emerge in discussion was the extent to which perceived control over the technology determined positive or negative perceptions of it. An analysis presented at the first workshop of the Open Subtitles Corpus, a dataset of over 100,000 film subtitles, identified control (or loss of it) as a recurring motif in films about artificial intelligence[13]. This directly reflected findings of the Royal Society's public dialogue on machine learning[14], in which members of the public with a range of backgrounds came together with leading scientists to explore the potential near-term implications of machine learning in settings such as health and social care, marketing and policing.

11. Hawking, S., Russell, S., Tegmark, M., and Wilczek, F. 2014 Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough? *The Independent*, 1 May 2014, available at: https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificialintelligence-but-are-we-taking-9313474.html [accessed 9 October 2018]
12. A summary can be found in: Cave, Stephen, and Kanta Dihal. 2018. *Ancient Dreams of Intelligent Machines: 3,000 Years of Robots*. Nature 559 (7715): 473–75. https://doi.org/10.1038/d41586-018-05773-y
13. This will be discussed further in: Recchia, G. Forthcoming. *The Fall and Rise of AI: Computational Methods for Investigating Cultural Narratives*, in *AI Narratives: A History of Imagining Thinking About Intelligent Machines* Oxford, UK: Oxford University Press
14. The Royal Society and Ipsos MORI 2017 *Machine learning: what do the public think?* Available at www.royalsociety.org/machine-learning [accessed 9 October 2018]

3. Narrative and emerging technology: lessons from previous science and technology debates

Previous waves of technological change offer lessons for the ways in which narratives might influence the development and adoption of AI technologies. The second AI narratives project workshop explored what lessons can be learnt from previous emerging technologies.

The second workshop brought together practitioners and scientists who had been at the forefront of debates about nuclear power, GM crops, and climate change to consider the narratives that accompanied these areas of science. While reflecting the experiences of individuals, these accounts of how public and policy debates unfolded can provide a useful lens through which to consider current debates about AI.

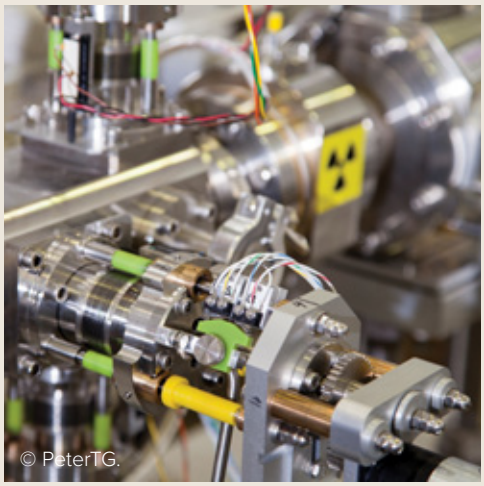
The perspectives from this workshop are summarised in the following case studies. They focus on non-fictional narratives, but the workshop has generated research which attends to the role both non-fictional and fictional narratives play in public discourse and decision-making[15].

15. This will be discussed further in: Sarah Dillon and Claire Craig, *Listen: Taking Stories Seriously* (forthcoming).

CASE STUDY ONE

A perspective on nuclear power

The enduring images of mushroom clouds over Hiroshima and Nagasaki – and the devastation that followed – dominated early framings of the power of nuclear technologies across the globe.



For some, this power spurred positive narratives; not only was nuclear the technology that 'won the war', but it was also a rapidly-advancing area of science that promised a new energy source to transform society. In the early years of its development, this excitement surrounding nuclear power contributed to promises such as those by Lewis Strauss, Chairman of the American Atomic Agency, in 1954 that nuclear power would produce energy that was "too cheap to meter".

Such positive narratives, however, came with challenges. Excitement about the potential of nuclear technologies contributed to hype surrounding the field, and hype helped create expectations that technological advances would come quickly. While scientific and technological developments in the field were significant, for those that had believed such hype this pace of change did not meet expectations, with implications for policy and investment decisions by both the public and private sectors. Strauss's promise was still being held up as an example of an inflated claim in public debate decades later.

For others, the development of nuclear technologies came with concerns about their destructive power; both the 'mushroom cloud' image and the invisible power of radiation contributed to new dystopian visions about post-nuclear futures, and narratives about the safety of such technologies.

Compared to artificial intelligence, nuclear power has had to contend with much more destructive consequences throughout the history of its application. Aside from the aforementioned intentional destructive powers, unintentional disasters such as the ones in Windscale (1957), Chernobyl (1986) and Fukushima (2011), spur continued urgent attention to safety issues.

These visions and narratives were an influence on the ways in which policy debates considered the risks of nuclear energy, and the regulatory environment that followed. Some workshop participants commented that the fears had in part been helpful in ensuring those engaged in the research and its implementation built a focus on safety into their cultures.

What lessons for AI might be taken from this perspective?

Popular excitement and concerns about an emerging technology influence public and policy debates, and shape the cost- (or risk-) benefit analysis publics make about that technology.

In navigating hype and uncertainty, it can be helpful to have a range of credible scenarios for how a technology might develop, which are agreed by key stakeholders and can be used in a public conversation.

Narratives of extreme fear can have potentially beneficial outcomes, for instance in ensuring safety concerns are considered at an early stage in research, regulation and implementation of a technology.

CASE STUDY TWO

A perspective on GM

Since their early development in the 1970s, fierce public and policy debates have surrounded the development of Genetically Modified (GM) crops.

For those positive about the applications, these offered new ways to feed the world: vitamin A enhanced rice (so-called golden rice) promised to reduce health problems in nutritionally-deprived people, and a range of plants with different environmental tolerance potentially offered new crops for farmers dealing with drought.

However, such arguments about the potential of GM for social good came with questions about the motivations of those driving technological development: who was developing these crops, and who would benefit?

The narratives that followed were tied to concerns about globalism, corporate control, and the legacy of colonialism; GM crops became a lightning rod for these broader societal concerns. In such narratives, multinational corporations played a key role, often contributing to scepticism about who would benefit from the widespread adoption of GM crops. For some, this corporate influence was contrasted against a vision which emphasised small-scale, locally-led innovation in food systems.

Alongside these questions about control, a narrative about the safety of GM foods also emerged. This narrative explored concerns about the movement of genes between plants or species, the power of farmers or others to control what would happen once GM crops were released into the wild, and broader debates about how societies interact with the natural world, protections for the countryside, and notions of the 'good life'.

The significant role of corporate interests in GM research and development is comparable to the current situation in AI. Just like GM food, AI is hailed as having immense potential for social good, but fears that the technology will be controlled by a powerful few corporations or nations are widespread.

What lessons for AI might be taken from this perspective?

Technology can be a lightning rod for broader social narratives or concerns, which can draw debate away from the actual risks and benefits of a technology. It is important to understand which broader concerns or interest may be at play, and how these are bundled with questions about a specific technology.

The reception of a technology can be shaped by perceptions of who benefits and who is at risk from technological developments.

CASE STUDY THREE

A perspective on climate change

At least two types of framings surround climate change debates: climate change as an issue of technology, and climate change as an environmental concern.


© Mario_Hoppmann.

Each of these framings has implications for the types of solution proposed to address climate change concerns; for some, technological advances seem key, with a focus on geoengineering or similar technologies, while for many the environmental issues require action through political and economic systems and human behaviour.

Across both of these framings, there is a narrative of uncertainty. In this narrative, technical discussions about risk assessment and data types are woven into questions about scientific credibility or trust. In this context, the language used to communicate climate science is highly influential: terms such as 'low confidence', 'unlikely', and 'uncertainty' take on different meanings for different communities. Scientists, meanwhile, have faced questions about their values and their role as 'honest brokers' of information and evidence.

In contrast to the narratives surrounding nuclear power, which promised significant benefits (or cautioned about significant risks) in the near-term, discussions about climate change have to grapple with a vision of the long-term, with impacts over decades or more. This disconnection from daily life can result in narratives that give a sense of disempowerment or disengagement; discussions about the future seem less relevant, with potential consequences for the actions people take (or do not take). AI applications span both near- and long-term concerns: AI technologies are already influencing society, but the development of human-like artificial intelligence may take decades or centuries, if it is ever possible at all.

The role of the individual in climate change also gives rise to a set of narratives around responsibility, with different narratives around the burden of guilt or responsibility on individuals, communities, and societies to address climate change.

What lessons for AI might be taken from this perspective?

The level of public trust in scientists and technologists influences the perception and reception of new technologies.

The language used to communicate scientific research is influential and terminology has different meanings and effects in different communities.

Clear models can support well-informed public debates, showing uncertainties, alternative futures, and the implications of different interventions.

4. Implications

Narratives have played a crucial role in the communication and shaping of ideas in the natural historical sciences and in the history of technology[16]. Both fictional and non-fictional narratives have real-world effects. This section explores the implications of the characteristics of prevalent narratives around intelligent machines and analyses the constraints on such narratives. It draws on discussions from workshops 1, 3 and 4.

4.i Disconnect from the reality of the technology

As the Royal Society report on machine learning indicated, public knowledge about the specifics of the science and technology is limited[17]. Their perceptions and expectations are therefore usually informed by their personal experiences of existing applications and by the prevalent narratives about the future[18].

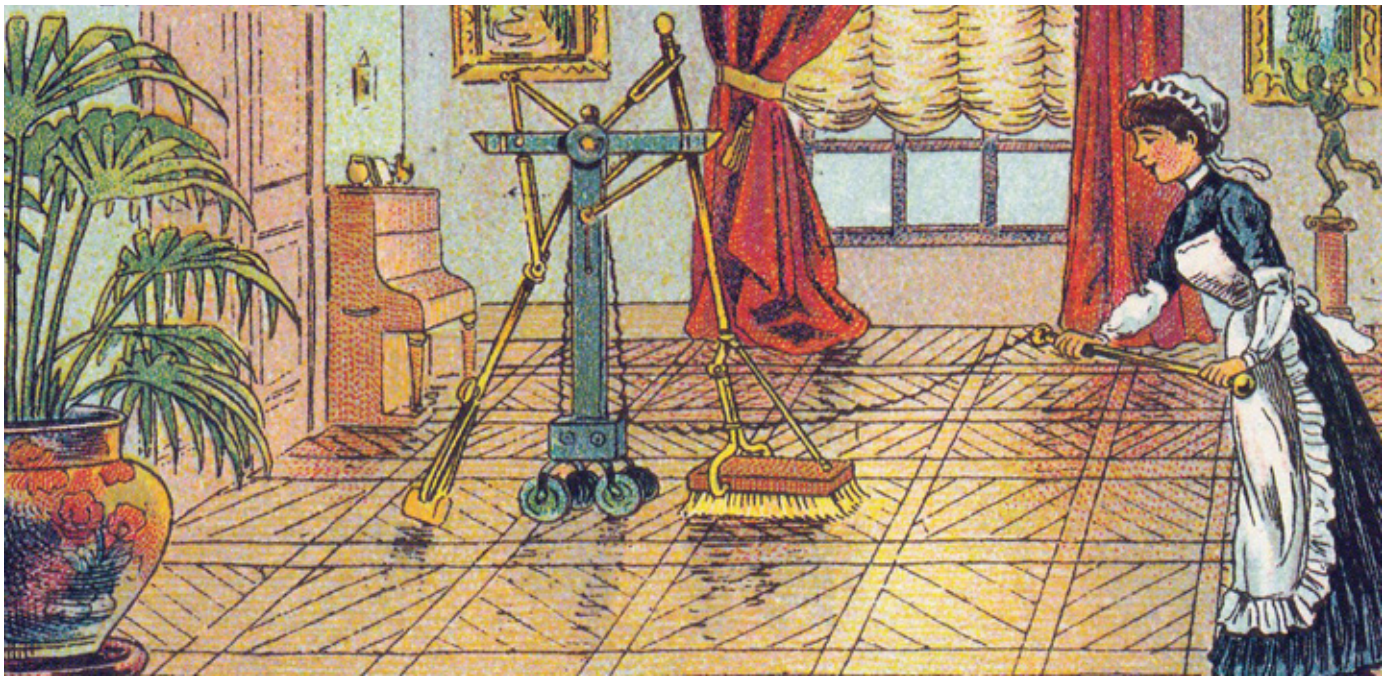
Both fictional and many non-fictional narratives focus on issues that form either a very small subset of contemporary AI research, or that are decades if not centuries away from becoming a technological reality. This disconnect between the narratives and the reality of the technology can have several major negative consequences.

The prevalence of narratives focussed on utopian extremes can create expectations that the technology is not (yet) able to fulfil. This in turn can contribute to a hype bubble, with developers and communicators potentially feeding into the bubble through over-promising. If such a bubble bursts because the technology was unable to live up to the unrealistic expectations, public confidence in the technology and its advocates could be damaged. The case study on nuclear power illustrates this pattern.

By contrast, false fears may misdirect public debate. For instance, an over-emphasis on implausible AI and humanoid robotics could overshadow issues that are already creating challenges today. These issues are often harder to describe through compelling narratives. They include: the robustness of digital infrastructure; and the consequences of potential uses of machine learning for bias and accuracy in decision-making, and for individual and collective privacy and agency. False fears may also lead to lost opportunities through failure to adopt potentially highly beneficial technology.

With major social and economic questions at stake, such as the future of work and distribution of wealth, it is important for public debate to be well-founded. Discussion of the future of work is distorted if it focuses only on robots directly replacing humans. Debate needs evidence and insight into the disruptive potential and opportunities created by new forms of business or social networks, as well as attention to the direct impact on particular tasks or jobs.

16. Beer, G. 2009 *Darwin's Plots*, Cambridge: Cambridge University Press; Morgan, M.S. (2017), Narrative Science and Narrative Knowing, *Studies in History and Philosophy of Science Part A* 62: 1-5 and the rest of this special issue on Narrative in Science.
17. The Royal Society 2017 *Machine learning: the power and promise of computers that learn by example*, available at www.royalsociety.org/machine-learning [accessed 8 October 2018].
18. House of Lords Select Committee on Artificial Intelligence, 2018 *AI in the UK: ready, willing and able?* HL Paper 100, Ordered to be printed 13 March 2018 and published 16 April 2018.



Bad regulation is another potential consequence of this disconnect. Prevalent narratives, including misleading ones, can influence policymakers: they either respond to these narratives because these are the ones that resonate with the public, or they are themselves influenced by them. False expectations can mean that a sector is allowed to grow without further intervention by governments, such as providing supportive regulation and market structures. As a result, a sector might grow slowly, reducing potential benefit. Or, it might grow fast, but in ways that are not aligned with social values, or in ways that lead to a bubble that will cause harm when it bursts. False fears, meanwhile, can lead to either over-regulation that suffocates growth and innovation, or to spending significant time and other resources on regulating something that will not require such regulation. Finally, the disconnect can lead to a misdirection of research funding. Hype bubbles can lead to disproportionate amounts of research funding being directed into a field because it is prominent in certain narratives, at the expense of other fields of research.

4.ii Underrepresented narratives and narrators

Prevalent fictional narratives around AI are based on a limited number of recurrent motifs. For instance, many narratives present hyper-sexualised, anthropomorphic AI, or extreme utopian or dystopian scenarios. This focus on embodiment and extremes also characterises non-fictional narratives around AI, for instance coverage in the popular press or industry advertising. The implications of this were explored particularly in Workshops 1 and 4.

Workshop participants discussed how 'narrative injustice' leads to under-representation of certain groups and types of story. The discussions focussed on three types of injustice: some groups are less likely to have authored prevalent narratives; some groups are less likely to figure in narratives as protagonists; and the narratives may perpetuate stereotypes or other biases.

Many contemporary narratives reinforce harmful stereotypes that have been established in AI discourse and in wider society. Gender stereotypes, for instance, are perpetuated both by fictional narratives, and by the way real technologies are framed. Alex Garland's 2014 film *Ex Machina* is one example: the question of whether a female robot is considered to have human-like intelligence is left to the judgment of a non-expert man who has been selected based on the fact that he will be sexually attracted to her.

These stereotypes also recur in the framing of AI technologies. The voice assistant Alexa, for instance, is portrayed as a tool that can be used to perpetuate the stereotype of the nuclear family: one advert shows a mother setting-up Alexa to give the father guidance on how to parent before leaving him alone with their child.

Counter-narratives can be found in less prevalent AI fictions. For instance, the 1997 film *Conceiving Ada* directed by Lynn Hershman Leeson provides a social commentary on the relationship between people and technology. In this film, the AI researcher is female, played by Tilda Swinton; moreover, the AI is embodied as a dog and as a bird, instead of in human form. The film is unusual in using AI to reframe the way in which historically significant events are described in order to highlight the roles and perspectives of women. M. John Harrison's *Kefahuchi Tract* trilogy of novels offers radical imaginings of human intimacy with algorithms, code and mathematics. Such narratives offer necessarily different perspectives on the power dynamics of AI, gender and embodiment.

Race and ethnicity stereotypes are also reinforced in many prevalent narratives. Stock images of anthropomorphic robots, used in non-fictional accounts as well as fictional ones, tend to depict a white plastic man with features that are visibly Caucasian. This depiction plays into a history of racist ideas that connect whiteness to intellectual ability. This trend has been made visible in critiques such as the indie film *Robots of Brixton* (2011). Narrative structures also reinforce race inequalities: the fear of being dominated by an AI might reflect the history of one group justifying domination over another using the claim they are intellectually superior. This narrative has been central to the justification of colonialism and patriarchy[19].

There are existing narratives that address pressing issues, concerns, and technological developments in accurate ways, and from a range of diverse perspectives. However, such narratives are underrepresented in the sense that a small number of similar and potentially misleading narratives dominate public debate and entertainment. This impoverishes both. In addition, relevant communities miss out on opportunities to develop and disseminate their visions, which means that there may be visions of AI 'for good' that are not being developed, or widely seen and heard.

4.iii Constraints

There are a number of reasons why only a narrow range of narratives are heard. These can be divided into three categories: social, economic, and literary or imaginative.

Social injustice

In Western society people from certain groups, such as women and people of colour, experience more difficulty in having their voices heard than others. Latent prejudices and historical injustices restrain both the creators of narratives and their audiences. Creators that do not belong to privileged groups may have to contend with limited access to training, influential platforms, or financial resources. This can affect both the quality and the reach of their narratives. Similarly, very few influential narratives seem to be constructed with these groups in mind. Such groups are therefore less likely to engage with these topics even though their lives may be affected by them. The perpetuation of the aforementioned biased narratives might influence who will choose to enter the AI workforce.

Economic

The nature of film and television as high-finance arts means that artistic decisions are, at least in part, determined by the economic demands of production and distribution. 'The problem with market-driven art-making is that movies are green-lit based on past movies. So, as nature abhors a vacuum, the system abhors originality. Originality cannot be economically modelled', director Lana Wachowski has noted[20]. Hence the *Terminator* franchise has been shaping the public perception of AI since 1984, with its sixth film due in 2019.

19. Cave, S. 2017 Intelligence: a history, *Aeon*, 21 February 2017, available at: https://aeon.co/essays/on-the-dark-history-of-intelligence-as-domination [accessed 9 October 2018]
20. Hemon, A. 2018, Beyond the Matrix, *The New Yorker*, 10 September 2012, https://www.newyorker.com/magazine/2012/09/10/beyond-the-matrix [accessed 10 October 2018]

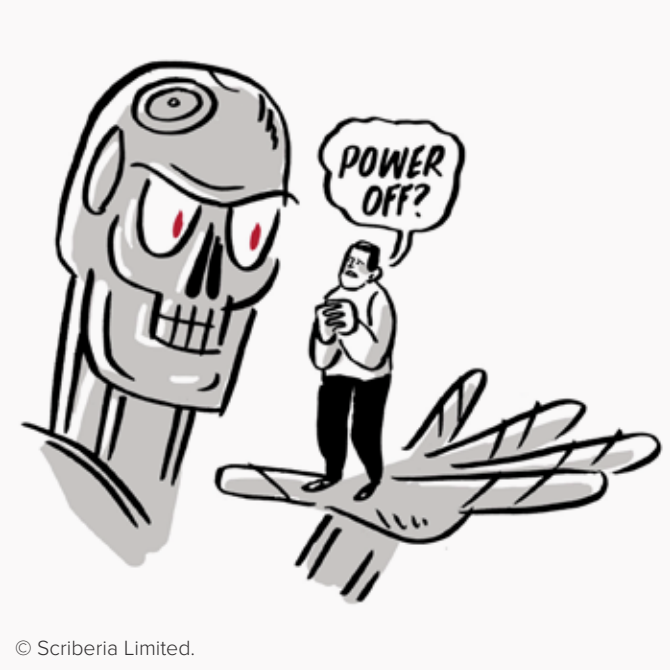
This same structure can, however, be used to accelerate change once an initial breakthrough has been made. One successful film that breaks with convention can provide an economic model for further films on that model. In addition, the arrival of new viewing platforms such as Netflix have changed the balance of power and reach, and made it easier for small-budget productions to reach a large audience.

These economic constraints also affect media working with smaller budgets, or media whose income streams are threatened by new digital technologies. In order to gain advertising income they may tend to resort to headlines and images that are both familiar and sensationalist, attempting to be successful from within an 'attention economy'[21]. The images and metaphors used come from a stock set that is easily deployable under time pressure, limiting the number of ways in which media outlets address specific issues around AI. In the UK, the Terminator is the image of choice for many print and online news media in discussions of AI safety.

Literary or imaginative

An individual's scope for paying attention to different narratives is limited: narratives that are not engaging enough will be ignored in favour of other narratives that are competing for attention. The nature of storytelling favours conflict, which means that narratives with this structure will be considered more engaging. Engaging utopias are notoriously hard to write: a story of a perfect world, in which nothing really happens because nothing goes wrong, is a dull story. Dystopian, negative, depictions of a future are therefore much more easily made engaging.

At the same time, narratives that aim to highlight underrepresented perspectives can usefully engage with these constraints. Prevalent AI narratives can be used as recognisable hooks to frame alternative perspectives on AI research and ethics from contemporary researchers[22]. AI researchers can use such narratives as a common entry point for science communication. They can leverage the exposure of their implausibility to introduce more accurate accounts of the current state of AI research[23].



Workshops 1 and 4 explored how the implications of the characteristics of prevalent AI narratives, and the constraints on such narratives, might be addressed. The collective hope was to find ways forward that would mitigate the disconnect from research and the underrepresentation, in order to encourage public discourse and diversify prevalent fictional imaginings. Directions for beginning such a journey can be found in Section 5.

21. Davenport, T. and Beck, J. 2001 *Attention Economy: Understanding the New Currency of Business*, Boston, US: Harvard Business Review Press.
22. Singler, B. 2018 Rise of the Machines: Short Documentary Films on Artificial Intelligence, available at https://bvlsingler.com/rise-of-the-machines-short-films-on-ai-and-robotics-available-online/ [accessed 3 November 2018].
23. Dillon, S. and Schaffer-Goddard, J. *What AI Researchers Read: The Role of Literature in Artificial Intelligence Research* (forthcoming 2019).

BOX 1

Future directions for AI narratives research and innovative practice

The systematic study of AI narratives is an emerging field of research, with a range of areas for development. Further research and innovative practice would develop the evidence base for the role AI narratives play in the development, regulation and reception of AI and related technologies. This section highlights a set of topics where workshop discussions suggested that progress would have a direct impact on increasing understanding of the role of narratives in public and societal perception and portrayal of AI.

© nicescene.

1. Global AI narratives

The first phase of the AI narratives project concentrated on Western English-speaking AI narratives, which excludes examination of portrayals and perceptions of AI in other regions of the globe. Such cross-cultural comparisons would be illuminating, enhancing understanding of how different cultures and regions perceive the risks and benefits of AI, and of the narrative influences that are shaping those perceptions. For example, as in Western narratives, AI is predominantly portrayed in Japanese fiction in embodied form. However, it is represented less as a slave or servant, and more frequently as a friend or tool. Mighty Atom, in English known as Astroboy, was the friendly lead character of a manga series that ran from 1952 to 1968). Tetsujin-28 (or, Iron Man 28) is a remote-controlled, non-autonomous device which featured in a popular 1956 manga later made into an anime TV series. Strikingly, whereas the figure most frequently used to portray intelligent machines in the UK is the hyper-aggressive T-800 Terminator, in Japan it is Doraemon, a podgy, friendly, blue robot cat. This research is now underway at the Leverhulme Centre for the Future of Intelligence's new Global AI narratives project funded by the Templeton World Charity Foundation. Comparative analysis will reveal whether the characteristics of prevalent Western AI narratives, and the constraints operating on such narratives, are shared globally. If not, UK practitioners will be able to learn in this space from other regions.

2. Analysis of AI media coverage

Mixed methodology studies of media coverage of AI stories in the past ten years would evidence the prevalence of recurrent tropes and images. This would extend existing work which concentrates exclusively on coverage in *the New York Times*[24]. In the UK such research could explore differences between different media segments, such as tabloid and broadsheet newspapers, or print versus online media. This research would benefit from comparative analysis of Western English-speaking media coverage with that of other countries, tracking the relationship between such

24. Fast, E. and Horvitz, E. Long-Term Trends in the Public Perception of Artificial Intelligence, *Proceedings of the Thirty-First AAAI Conference on AI*, 963-969.

coverage and public perception and policy across nations. Focused case study analysis of the relationship between media coverage and the science being covered, including sociological interviews with scientists, would evidence the disconnect between media narratives and the actual science.

3. Quantitative survey of influence of science fiction
A large-scale quantitative survey would evidence the influence of science fiction reading and viewing on AI researchers' career choice, research direction, community formation, social and ethical thinking, and science communication. Qualitative research presented at workshop 3 indicates that the influence of science fiction is significant and distributed and that it may play a role in contributing to inequities and lack of diversity in the field. Further research evidencing this would provide a basis for recommendations regarding narrative-based contributions to diversity initiatives in the sector.

4. Individual narrative case studies
Fine-grained individual case studies of the influence of narratives in specific areas of AI research, or on specific researchers, both contemporary and historical, can provide specific examples of the role such narratives play. For example, existing unpublished work analyses the literary influences on Alan Turing's foundational essay 'Computing Machinery and Intelligence' (1950)[26]. Identification and analysis of further case studies will add depth and breadth to the AI narratives research field. At the same time, a large and well-structured corpus of historical nonfictional case study narratives could be researched and made available for practitioners to reference, instead of the currently used extreme narratives.

5. Investigating interests
Research into how AI narratives are created and disseminated will identify who benefits from them and why. Social scientific research could, for example, be used to analyse funding and dissemination, the role of charisma, hierarchies of power, actor-networks, and other forms of influence in the context of AI narratives. This will contribute to understanding why certain narratives are prevalent, how they might be appropriately critiqued, and how others might be brought to the fore. Investigating forecasting methods and the use of scenarios in the sciences can help build an understanding of how to construct plausible narrative accounts of the future.

6. Incorporating narratives research into AI ethics research
This write-up has outlined the problems and limitations of Western English-speaking AI narratives, including with regard to issues of equality and diversity. Further exploring the role of narratives will be an important component of the burgeoning research into the ethics and impact of AI and related technologies. It will be important to consider how existing injustices – especially with regard to race, gender and class – contribute to the perpetuation of certain narratives (for example, by inhibiting the dissemination of certain voices), and are perpetuated by them (for example, making it more difficult for certain communities to enter the field of AI).

25. Dillon, S. and Schaffer-Goddard, J. *What AI Researchers Read: The Role of Literature in Artificial Intelligence Research* (forthcoming 2019)
26. Jennifer Schaffer, *How to Write a Human: Playing for Affect in Alan Turing's Imitation Game*, MPhil Dissertation June 2016, Faculty of English, University of Cambridge.

5. The role of practitioners

Drawing from the understandings of how and why popular discourse portrays AI set out in the preceding Sections, this Section considers questions that practitioners may wish to consider when thinking about how to develop and inform AI narratives. It summarises some of the suggestions for communicating AI that have been offered by participants in workshops throughout the project. By 'practitioners' this Section refers to anyone engaged in creating or using AI narratives in the course of their normal professional or personal activities. It draws in particular on Workshop 4, whose participants included journalists, professional science communicators, creative artists, research scientists, and scholars.

5.i Communicating AI: lessons from narratives about previous waves of emerging technology

Section 3 set out how narratives about previous waves of emerging technologies have influenced public perception and technological development. A common thread across these technologies is the importance of a discourse that reflects differing levels of confidence or uncertainty in different types of technologies and over different periods.

In some cases, this lesson comes in the form of a cautionary tale about over-promising the potential of a technological advance, and subsequently failing to deliver on these promises (as in the case study on nuclear power). For many in AI, this risk is felt particularly acutely, with the field having been subject to previous waves of hype and disillusionment that have had very real effects upon both public perceptions and research in the field. For other technologies, this lesson is found in the ways in which different publics take account of, or have confidence in, relevant science depending on the degree of consensus that is communicated to public and policy audiences (as in the case study on climate change).

These lessons point to the importance of understanding how to communicate and discuss uncertainty. They also highlight the importance of crafting compelling narratives about AI that accurately reflect the underlying science and its possibilities, while acknowledging scientific and social uncertainties about the future. An aim is to draw people's attention to knowledge, and to informed speculation, in ways that will best enable reasoned discussions about the possible futures.

Questions for practitioners:

- With what levels of confidence is it possible to discuss the various aspects of this topic? How can uncertainties be built into the story being created while also conveying the limits to informed speculation?
- How can stories be created that generate engagement and excitement without contributing to hype?


© Iirtlon.

5.ii Reshaping AI narratives: alternative narratives and voices

This write-up notes that despite the broad range of potential applications of AI, a small number of analogies, stories or images tend to dominate broad-based public discussions, notably the Terminator-style humanoid imagery and a tendency for stories to default to extreme descriptions of apparent utopias or dystopias. Prevalent stories and images have also tended to lack social and cultural diversity both in their authorship and in their protagonists and imaginings.

Workshop participants suggested that there are successful alternative models for talking about AI, which have the potential to be much more influential in shaping broader public discourse. From the workshop discussions, two broad approaches to creating these alternatives emerged: finding alternative analogies and images; and supporting a range of voices.

For example, there are already a range of ways in which intelligence exists and is portrayed in non-human systems. In the animal kingdom, communication between colony members – whether in bees, ants, or the octopus – has been studied and discussed as a form of intelligence. There may be ideas from such portrayals that are relevant to narratives of AI as a distributed, rather than embodied, force. Perhaps less popularly compelling, but also relevant, are the notions of human organisations, such as businesses or other social networks, being considered to have forms of collective intelligence.

Another approach to generating alternative narratives could lie in the everyday or 'mundane' nature of many AI technologies: both great art and significant social movements have found power in conveying the minutiae of everyday life in new ways. Given the range of ways in which AI systems are beginning to find application in 'mundane' activities, the 'everyday' might be a fruitful source of ideas for new stories or narratives about AI and its implications.

A number of workshop participants noted how the composition of the research community contributed to shaping the types of narratives that dominates discussions about AI. For the field to advance in a way that represents a broad range of interests or concerns, it will need to draw from a wider range of voices from different backgrounds and social groups.

There already exist networks and platforms that promote underrepresented voices in the AI communities, such as Women in Machine Learning and Black in AI. Further action to create opportunities for underrepresented voices to engage in public and policy debates could build on these initiatives.

Questions for practitioners:

- What images might provide a compelling alternative to the focus on humanoid embodied intelligence?
- How can discussions about advances in AI be connected to everyday life?
- How can a wider range of voices be brought into public discussions about AI?

5.iii The importance of dialogue

An alternative approach to reshaping the narratives surrounding AI lies in changing the ways in which people create narratives, rather than focusing on the content of the narratives themselves. Spaces can be created that allow new stories to emerge through new approaches to dialogue and engagement.

The last five years have seen a number of early efforts to create spaces for informed public dialogue about AI and its implications. In 2016 and 2017, the Royal Society carried out the first UK public dialogue on machine learning. This brought together AI researchers with demographically representative groups of members of the public across the UK. They considered the risks and benefits of AI technologies. Since that effort, the Royal Society has continued to create spaces to support a well-informed public conversation about AI through its flagship public lecture series, *You and AI*.

Other initiatives in this area include:

- The RSA's citizens' juries project on automated decision-making, which has been considering the conditions under which publics would have confidence in automated decision-making systems, and will report at the end of 2018[27];
- doteveryone's Women Invent the Future initiative which, although not AI-specific, represents a further form of innovation in enabling new authors and narratives[28];
- Nesta's public engagement on AI and the future of work, which has been creating new stories about AI technologies and the workplace, putting the voices of workers at the centre of this dialogue[29].

BOX 2

Key messages that have emerged from public dialogues around AI include:

- Public awareness of AI technologies is low, but awareness of applications is higher. Both vary demographically.
- Context is key to how members of the public evaluate AI technologies. The risks and benefits people associate with AI technologies vary according to the application under consideration.
- In considering an application, people have questions about the purpose of its development, who benefits, and why the application is necessary.

These messages are broadly consistent with those emerging regularly in public dialogue about new technologies.

Understanding the types of question publics have about AI technologies is a key part of communicating and engaging effectively, and a number of workshop participants noted that good 'explainer' material is consistently popular. This should be seen as a platform to enable more people to engage with and influence future applications.

This type of public dialogue requires engagement from both researchers and publics. There are a range of ways in which research funders, researchers, and others can support scientists to enter public debates about AI technologies:

- **Training:** initiatives to train scientists to communicate with the media and the public, such as the communications training offered by the Royal Society, can offer assistance navigating public debates and in communicating risk and uncertainty.
- **Engaging:** many researchers already engage in public conversations about the development of AI technologies through social media. Further engagement in such debates can offer alternative narratives to those in play in mainstream outlets.
- **Sharing:** researchers can use social media to increase the reach of articles that reflect the technological realities of AI development or in which users discuss how they relate to AI technologies. Sites such as robohub.org are already working to increase the profile of narratives that seek to demystify AI technologies.

Inclusion of funding for public engagement in large-scale research funding programmes in AI would support such engagement by the science community.

Considerations for practitioners:

- How can researchers be supported to engage in public debates about AI technologies?
- In what ways can governments, research institutions, companies, and the third sector support inclusive but context-specific public dialogues about AI technologies?
- How can public dialogues support new ways of thinking and talking about AI?

27. Both the Royal Society's You and AI lecture series and the RSA's citizens' juries were supported by DeepMind.
28. See, for example, https://doteveryone.org.uk/2018/07/women-invent-the-future/
29. See, for example, https://www.nesta.org.uk/project/common-futures-future-work-imagined-working-people/

Annex A: Project summary

Workshop 1 (16 May 2017)

An introduction to narratives and AI

From Hephaestus's golden handmaids to Karel Čapek's Roboti, people have been imagining intelligent machines long before technology helped build them. Technologies like AI are therefore developing in an environment that is loaded with a long-history of cultures and narratives that shape how societies respond to advances in these technologies. These narratives are a way by which people make sense of the world around them, and can influence technology development and use, policy responses, and public debates.

Current narratives around AI draw from stories about technology control, human-machine interaction, and the future of work and wealth, amongst other influences. This first workshop explored the stories and narratives surrounding the development of AI, and their implications for its development.

The following presented at the workshop:

Professor Caroline Bassett, University of Sussex

Dr Stephen Cave, University of Cambridge

Dr Claire Craig, The Royal Society

Dr Adrian Currie, University of Cambridge

Dr Kate Devlin, Goldsmiths, University of London

Dr Kanta Dihal, University of Oxford

Dr Sarah Dillon, University of Cambridge

Patrick Parrinder, University of Reading

Dr Gabriel Recchia, University of Cambridge

Dr Beth Singler, University of Cambridge

Dr Will Slocombe, University of Liverpool

Professor Elly Truitt, Bryn Mawr College

Workshop 2 (30 May 2017)

Technological narratives – lessons for AI

From nuclear energy to genetic engineering and stem cells, the ways in which scientists, policymakers, and publics have talked about new technologies – and their risks and benefits – has contributed to how these technologies develop. The second workshop considered the evolution of narratives around emerging technologies, using a series of case studies to explore the implications of narrative for how technologies evolve, how different communities respond them, and, ultimately, their place in society.

The following presented at the workshop:

Professor Jon Agar, UCL

Professor Sir David Baulcombe FRS, University of Cambridge

Professor Steven Cowley FRS, Imperial College London

Dr Tamsin Edwards, KCL

Professor Keri Facer, University of Bristol

Professor Penelope Harvey, University of Manchester

Professor Theodore Shepherd FRS, University of Reading

Dr Jon Turney, UCL

Workshop 3 (13 and 14 July 2017)

AI myth and reality

As part of the CFI's 2017 annual conference, a third workshop explored the myths that circulate around AI technologies, drawing from perspectives about the role of science fiction in creating visions for the future, and the role of arts and media in helping shape narratives around AI. It also considered current public perceptions of AI and intelligent robots, and how portrayals about AI's capabilities diverge from reality.

The following presented at the workshop:

Dr Anna Alexandrova, King's College London

Dr Stephen Cave, University of Cambridge

Dr Claire Craig, the Royal Society

Professor Jon Crowcroft FRS, University of Cambridge

Dr Adrian Currie, University of Cambridge

Dr Sarah Dillon, University of Cambridge

Luba Elliott, Impakt Festival

Professor David Alan Grier, George Washington University

Dr Hatice Gunes, University of Cambridge

Cassian Harrison, BBC4

Dr Sabine Hauert, University of Bristol

Professor Neil Lawrence, University of Sheffield and Amazon

Professor Huw Price, University of Cambridge

Professor Stuart Russell, UC Berkeley

Ben Russell, Science Museum

Jennifer Schaffer-Goddard, University of Cambridge

Dr Beth Singler, University of Cambridge

Professor Murray Shanahan, Imperial College London and Google DeepMind

Professor Alan Winfield, University of Bristol

Workshop 4 (3 May 2018)

How do we talk about AI?

When talking about AI, scientists, scholars and science communicators are faced with choices about how they discuss the risks, benefits, and implications of these technologies. A range of forces shape these choices, and different approaches to communication and engagement can be suitable for different audiences and purposes.

The final workshop in this series focussed on steps that researchers, communicators, and others could take to support a well-founded public debate and well-informed policy discourse about AI.

The following presented at the workshop:

Dr Stephen Cave, University of Cambridge

David Chikwe, BBC

Kenneth Cukier, The Economist

Sally Davies, Aeon

Dr Kanta Dihal, University of Cambridge

Dr Sarah Dillon, University of Cambridge

Tabitha Goldstaub, Cognition X

Bill Hartnett, the Royal Society

Dr Sabine Hauert, Bristol University

Dr Fiona Kumari Campbell, University of Dundee

Dr Genevieve Lively, University of Bristol

Professor Sofia Olhede, UCL

Sydney Padua, graphic artist

Jonnie Penn, University of Cambridge

Dr Amanda Rees, York University

Emma Reeves, BBC

Gila Sacks, Department of Digital, Culture, Media and Sport

Dr Henry Shevlin, University of Cambridge

James Young

George Zarkadakis, Willis Towers Watson

Annex B

Project team

Dr Stephen Cave, Executive Director, Leverhulme Centre for the Future of Intelligence, University of Cambridge

Dr Claire Craig, Chief Science Policy Officer, the Royal Society

Dr Kanta Dihal, Postdoctoral Research Associate and Research Project Coordinator, Leverhulme Centre for the Future of Intelligence, University of Cambridge

Dr Sarah Dillon, Programme Director – AI: Narratives and Justice, Leverhulme Centre for the Future of Intelligence, University of Cambridge

Jessica Montgomery, Senior Policy Adviser, the Royal Society

Dr Beth Singler, Junior Research Fellow in Artificial Intelligence, Homerton College, University of Cambridge

Lindsay Taylor, Policy Adviser, the Royal Society (until October 2018)

Peer reviewers

Professor Zoubin Ghahramani FRS, Chief Scientist, Uber

Professor Dame Angela McLean FRS, Professor of Mathematical Biology, University of Oxford

Professor Mary S. Morgan FBA, Albert O. Hirschman Professor of History and Philosophy of Economics, LSE

Dr Amanda Rees, Department of Sociology, University of York

Participants

The Royal Society would like to thank all those who contributed to the development of this project through attendance at events.

Funding

Drs Cave, Dihal, and Dillon are funded by a Leverhulme Trust Research Centre Grant awarded to the Leverhulme Centre for the Future of Intelligence. Dr Singler was funded by a Templeton World Charitable Foundation grant during the course of the AI narratives project, awarded to the Faraday Institute for Science and Religion, St Edmund's College, Cambridge.

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500
W royalsociety.org

Registered Charity No 207043
Issued: November 2018 DES5612