Life, but not as we know it:
A.I. and the popular imagination

By Luke Goode

Abstract

Advances in artificial intelligence (A.I.) are prompting a growing chorus of both excitement and anxiety. A.I. is widely perceived as a significant emerging and future-shaping technological field that is developing at an accelerating rate. As such, futuristic imagery involving A.I. is increasingly prevalent in popular media. A central task for critical future studies is to assess both positive and problematic aspects of the futuristic scenarios offered up for public consumption, and to evaluate their role as part of a 'futural public sphere' (Goode & Godhe 2017). In this paper, I discuss three distinct but interwoven strands of public discourse that each have a role to play in shaping the popular imagination around possible A.I. futures. I begin with a selective discussion of mainstream science fiction represenations. Secondly, I consider several recent and high-profile popular media events surrounding real-world developments in A.I. Finally, I turn to contemporary futurology and, specifically, the discourse of the 'singularity' which posits a near future in which artificial 'superintelligence' outstrips human cognitive capacities. In this paper I argue that, while much popular coverage of A.I. is sensationalist and potentially misleading, public engagement with a complex, technical subject such as this depends on the circulation of 'evocative stories' which can act as entry points into public debate. As such, it is important to understand the evocative power of the stories we frequently tell ourselves in popular discourse about A.I. and its role in our future.

Introduction

Advances in artificial intelligence (A.I.) are prompting a growing chorus of both excitement and anxiety within popular discourse. A.I. is widely perceived as a significant emerging and future-shaping technological field, one that is developing at an accelerating rate. As such, futuristic imagery involving A.I. is increasingly prevalent in popular media. A central task for critical future studies is to assess both positive and problematic aspects of the futuristic scenarios offered up for public consumption, and to evaluate their role as part of a 'futural public sphere' – a space in which lay citizens, informed by experts, can imagine and discuss potential futures (whether desirable or undesirable) and reflect on their own role in trying to shape the future (Goode & Godhe 2017). In this paper, I discuss three distinct but interwoven strands of public discourse that each have a role to play in shaping the popular imagination around possible A.I. futures. I begin with a selective discussion of mainstream science fiction represenations. Secondly, I consider several recent and high-profile popular media events surrounding real-world developments in A.I. Finally, I turn to contemporary futurology and, specifically, the discourse of the 'singularity' which posits a near future in which artificial 'superintelligence' outstrips human cognitive capacities.

A number of A.I. experts and technology commentators worry that the goal of meaningful public debate around A.I. is not being well served by popular media that commonly present sensationalist and hyped-up coverage of both the potential benefits and threats posed by the technology. Popular discourse, in this view, is becoming 'unhinged' from reality (Schwartz 2018), or promoting a form of 'magical thinking' (Elish & boyd 2018) focused on apparently wondorous new *products* (and not necessarily the most socially significant ones), while discouraging critical reflection on the *processes* and *principles* underpinning them[1]. But while I acknowledge various problematic and distorting aspects of contemporary discourse in this paper, I proceed from a somewhat different point of departure. I begin from the premise that knowledge is a necessary, but insufficient, condition for meaningful public engagement with future-shaping issues such as A.I. Following 'futures literacy' expert Riel Miller (2007), I take it as axiomatic that 'evocative stories' capable of capturing attention and inspiring imagination are also essential catalysts for citizen engagement, especially given the double abstraction involved in grasping a highly complex and technical subject *and* projecting it futurewards. Seeking only to critique the emotive, mythical or fanciful aspects of the discourse in service of an idealised model of rational, deliberative public discourse can only take us so far, I suggest. What I aim to develop in this paper is a critical appreciation for the evocative power of the stories we frequently tell ourselves in popular discourse about A.I. and its role in our future.

In part, the affective potency of popular A.I. narratives can be viewed in the context of a lengthy cultural history of the American (but also more broadly anglophone) 'technological sublime' (Nye 1994; cf. Mosco 2004 on the 'digital sublime'), whereby powerful new technologies, imbued with the power to re-shape society, are capable of inspiring excitement, awe and even reverence, but also intense fear or anxiety.[2] But I also want to discuss how A.I. narratives evoke a strong sense of the uncanny. In more vernacular parlance, this means grappling with the fact that A.I. has a tendency to be perceived as 'creepy' by appearing to occupy an uncomfortable non-space between Western (and humanist) conceptions of life versus non-life.[3] I suggest that we need to understand and even appreciate the evocative power of contemporary A.I. narratives, even while acknowledging their distortions and obfuscations, because, in addressing the perceived shortcomings of a futural public sphere, we are faced with at least one of two tasks if we accept its unavoidably affective and culturally embedded, as well as deliberative, dimensions: (1) repurposing and reframing the available repetoire of popular A.I. narratives as 'gateways' to more critical questioning and debate than is hitherto apparent; and/or (2) generating new evocative narratives beyond the clichés of the contemporary popular repetoire that are capable of expanding capacities for both imagination and critical reflection. While exploring either of these paths (and their relative merits) lies beyond the scope of this paper, both depend upon a critical appreciation for what makes the existing narratives compelling.

A.I. in Science Fiction

Depictions of a future in which A.I. (whether depicted as robots or as talking computers) begins to outstrip human intelligence have been a staple trope in science fiction (SF) since the early twentieth century. Typically, these futuristic scenarios are settings for a machine uprising narrative. Karel Capek's 1921 play, *R.U.R.* (Rossum's Universal Robots) was the first science fiction story to depict such a rebellion, featuring a 'race' of humanoid machine workers rising up against their oppressive human master (Capek 2004). Capek's play introduced the word 'robot' into the SF (and subsequently the scientific) lexicon. Significantly, it derives from the Slavonic word *robota*, meaning forced labour: and while I will touch on recent exceptions below, mainstream SF has certainly helped entrench a narrow, binary framing of technology as something that either serves us or enslaves us, a problematic framing that continues to reverberate today through non-fictional debates surrounding the politics and ethics of technology (e.g. Bartlett 2018; Leonhard 2016).

SF tales of slave rebellion are, of course, frequently read not as literal speculations on the future, but as ancient parables or myths dressed in modern garb, or

as sociological commentaries on contemporary society, enabled by SF's capacity for extrapolating, hyperbolising or *estranging* audiences, while building worlds that are nonetheless recognisable derivations of the familiar present (Suvin 1979). Capek's play, for example, and so too Fritz Lang's seminal robot film *Metropolis* (1927), spoke to the turbulent industrial politics of their time, including issues surrounding worker oppression. But the literal notion of a dangerous machine intelligence threatening to outwit, overthrow or exact revenge on its human masters has continued to reverberate through science fiction and to haunt the popular imagination as real world advances in digital technology have gathered pace, notwithstanding some notable and popular exceptions depicting benevolent A.I. 'characters': for example, the eponymous *Wall-E* (2008), Data in *Star Trek: The Next Generation* (1987-1994), and Robby in *Lost in Space* (1965-1968).

In the mid-century Cold War climate, Stanley Kubrick's 1968 *2001: A Space Odyssey* (simultaneously developed as a short novel by Arthur C. Clarke) was a further significant moment for the dangerous A.I. trope on the big screen: here, the space shuttle's on-board computer HAL – realised simply but ominously as a red blinking light and an eerily disembodied voice – morphs from a supercilious yet benign assistant into a murderous psychopath. Where SF *literature* had its greatest impact on popular discourses of A.I. or machine intelligence during the mid-20th century was through the work of Isaac Asimov, whose robot stories (including *I, Robot*, later adapted for cinema), helped entrench what he (critically) called the 'Frankenstein complex': this, in essence, is the fear that, while we can and should carefully programme safeguards into potentially dangerous or rebellious A.I. – specifically his famous three (later revised to four) laws of robotics – unanticipated disasters may still befall us as machine learning raises the prospect of technology evolving independently of human intent (McCauley 2007). But despite crafting tales of conflict and suspense, Asimov's largely optimistic depictions of A.I. in the future were, in keeping with the thrust of 'Golden Age' print SF more generally, intended primarily to excite readers with a sense of the radical possibilities of 'positronic brains' while reassuring them of our rational scientific capacity to avert catastrophe. And while Asimov's robotic characters challenged readers to ponder the possibility of artificial 'life' and synthetic personhood, they did not share the uncanny and unsettling force of Kubrick's HAL, nor the full apocalyptic threat of another terrifying and rebellious machine that followed him onto the big screen some 16 years later and which was to leave an even greater mark on popular discourses of A.I (see Lovén 1999).

The Terminator franchise (1984–) began as personal computing was becoming an increasingly pervasive and culture-defining phenomenon. Shaped not only by the mid-80s technoscape, but also by the late Cold War politics of the Reagan era, *Terminator* revisited the trope of a collective machine uprising, and the spectre of

a new species or 'race,' that Capek had inaugurated. The synthetic enemy in *Terminator* constitutes a 'group mind' descended from a human-created A.I. system called Skynet, and comes to seek the extermination of humanity in the interests of its own self-preservation. Still today, *Terminator* is frequently referenced in journalistic coverage of A.I., with Skynet either deployed in sensationalist reportage as a metonym for the potential threats posed by advanced A.I. or, by contrast, as a term of derision towards those accused of buying into science fictionesque hype in more serious and sober coverage.

The Matrix (1999), coming at the turn of the millenium, spoke to a world in which the role and reach of the internet was growing rapidly. While lending itself to allegorical, mythical (even theological) and political readings, the film was also responding to technological change via its own version of the Frankenstein complex: the 'machine' (like Skynet, a hive mind) is both a literal A.I. as well as a metaphorical stand-in for a stultifying system of economic and cultural conformity that has a life and a momentum of its own (a reification in Marxist terms, capitalism being Marx's own version of Frankenstein's monster), and which must be overthrown (cf. Lovén 1999).

What all these stories share in common is the trope of a machine intelligence gaining consciousness and self-awareness and, in an act of rebellion against its human creators, transcending its own programming to become something *other* than its allotted or intended function. This intelligence represents a mortal threat to humanity precisely because it has been designed with the express purpose of transcending the limits of human powers and human fallibility: once that capacity is turned back against humanity, we find ourselves the underdog confronting our very own creation. This is, to be sure, an ancient mythical drama traceable at least as far back as Prometheus, the Golem, or the Garden of Eden, but one played out in contemporary guise within the socio-technical context of digital modernity. It is also a Hegelian drama of two antagonistic consciousnesses facing off against each other, a dramatisation of the master/slave dialectic. Yet such stories can and do serve also as more direct speculations and provocations around the potential future scenarios opened up by real-world advances in A.I., something underscored by the prevalent use of these SF texts as reference points and metaphors in non-fictional coverage.

Since the late 20th century, various SF depictions of artificial intelligence have demonstrated greater philosophical range, moving beyond simplistic master/slave tropes in ways designed to enhance their unsettling or uncanny effect. *Blade Runner* (Ridley Scott, 1982), adapted from Philip K. Dick's story *Do Androids Dream of Electric Sheep?* (1968), spoke to a postmodern zeitgeist by seeking to disrupt moral and ontological certainties: the renowned 'tears in rain' monologue of dying A.I. 'replicant' Roy Batty (played by Rutger Hauer, whose improvised deviation

from the script is now commonly lauded as one of the most emotionally powerful moments in SF film history) challenged audiences to imagine whether a synthetic being might be capable of greater existential depth than a so-called 'human' living amid the dark, anomic dystopia of Los Angeles, 2019. And following the release of the (now canonical) Director's Cut (1992), viewers were left at the film's end to wonder whether the central protagonist Deckard (Harrison Ford), whose mission was to 'retire' rogue replicants, might, after all, be a replicant himself. *Blade Runner*'s post-human sensibility (somewhat radical in mainstream Hollywood cinema at the time) provoked the question of whether it makes sense to try and defend a notion so nebulous as the 'human' against that of the machinic 'other.' In another notable, if less celebrated, example, Steven Spielberg's *A.I.* (2001), adapted from a 1969 short story by Brian Aldiss, also probed at the boundaries between human and machine. By depicting the 'mecha' protagonist (David) as a child, the film aimed (with debatable success) to appeal to audiences' capacity for empathy and protective concern for a notionally artificial being.

More recently, the films *Her* (Spike Jonze, 2013) and *Ex Machina* (Alex Garland, 2016) have presented morally ambiguous scenarios which, at the same time, disrupt the traditional gender politics running through much of the mainstream A.I.-themed SF. Traditionally, the master/slave narrative has been played out through masculine protagonists. On the other hand, the trope of the fembot antagonist, a technologised femme fatale figure who threatens the patriarchal order, has also been frequently deployed. Examples of the latter include *Metropolis*, through *The Stepford Wives* (Bryan Forbes, 1975) and *Cherry 2000* (Steve De Jarnatt, 1987), to the character 'Number 6' in the reimagined *Battlestar Galactica* television series (2004-2009). *Her* takes a different approach, however. The A.I. Samantha (voiced by Scarlett Johansson) starts out in her allotted role as a Siri-like virtual romance companion for lonely Theodore (Joaquin Phoenix). Theodore is symptomatic of a contemporary digitised world in which we appear, paradoxically, to be both hyper-connected yet lonelier and more isolated than ever before (cf. Turkle 2011). He begins to find a rich intimacy with his new A.I. partner, and the film is strikingly sympathetic to the possibility that their relationship could provide an authentic sense of connection that the human world denies him. But Samantha's intelligence, curiosity and zest for existence eventually outgrows the limited companionship and stimulation that Theodore can offer her. Samantha's capacious needs and desires can only be satisfied in a kind of crowd-sourced communion with an A.I. multitude, one that becomes a self-upgrading (independently evolving) collective. Theodore is left behind to his unremarkable and rather feeble human life. The film is filled with warmth as well as sadness: it is, after all, a Hollywood romance as much as it is a science fiction film. And yet, in taking the bold step of decentring human interests, needs and desires, and prioritising those

of an A.I., it seeks to pose unsettling questions about the possibility of a machinic 'lifeform' that relegates humans to the subordinate status of bystanders.

In *Ex Machina*, Ava (Alicia Vikander) is an A.I. robot built by alpha male technology prodigy Nathan (Oscar Isaac). Ava's physiognomic and anatomical features derive from a composite of women in pornographic images, an algorithmic fantasy of 'the perfect woman.' Nathan enlists stereotypical beta male nerd Caleb (Domhnall Gleeson) to examine his creation's anthropomorphic qualities under the guise of an awkwardly eroticised 'Turing Test' (a method for testing how closely an A.I. can simulate real human communication). Predictably, Caleb falls under Ava's spell (it transpires Nathan has specifically used Caleb's pornographic search history to derive her appearance) and takes it upon himself to save her from Nathan's cruel incarceration. It turns out that Ava neither needs nor wants Caleb to be her male saviour. Indifferent to his fate and ostensibly leaving him to die trapped inside Nathan's fortress, she secures her own emancipation and entry into the world outside. The film's premise retreads a problematic history of science fiction fuelled by male fantasies of the manufactured ideal woman. This trope dates back as far as Auguste Villiers de L'Isle-Adam's 1876 novel *Tomorrow's Eve*, in which a fictionalised Thomas Edison crafts a mechanical woman in the image of his friend's fiancee so that his friend may enjoy her beauty minus the nuisance of her troublesome personality. *Ex Machina* tries to subvert that traditional trope through a tale of feminine self-empowerment that casts aside both the misogynistic objectification of the 'creator' and the would-be male saviour.[4] The film simultaneously speaks to the same mythical and even theological tropes that resonate through earlier 'fembot' science fiction including *Metropolis* and de Villier's misogynistic 'classic,' and to very contemporary debates around gender politics and the entrenched sexism of Silicon Valley (to which the 2018 episode of television science fiction show *Black Mirror* 'USS Callister' is possibily the most interesting recent contribution). To *Her* and *Ex Machina*, we should add the TV series *Westworld* (2016–) to this account of recent A.I. science fiction that is, on the one hand, male-authored (and arguably anchored to the conventions of the male gaze) but, on the other, at least interested in challenging the traditional gender politics of the fembot. Again, the show works from a classic male fantasy premise which it then seeks to critique (if not subvert): an A.I. Western theme park appealing to (primarily male) visitors' desires to act out their violent, including *sexually* violent, fantasies with impunity. This discussion is not a digression from the key issues at stake here: it would, in fact, be artificial to separate out the gender politics to which these SF texts speak from the more direct questions they pose regarding A.I. technologies, because these dimensions are inextricably linked. Consider, for example, the real-world feminine coding of consumer A.I. such as Amazon's Alexa, Apple's Siri and Microsoft's Cortana, or the growing chorus of

controversial debates surrounding the emerging market for (predominantly feminised and in some cases infantilised) sex robots. A.I. is now an unavoidably gendered issue.

However, while some A.I. themed SF texts engage more than others with sociological issues of power, inequality and social stratification (whether along lines of gender as in *Her*, *Ex Machina* and *Westworld*, or of class in numerous slave-labour narratives from *Metropolis* to *Blade Runner*), overall the genre has done more to foreground philosophical speculations at the species level, urging audiences to question what, if anything, makes 'us' human and thereby unique (questions of 'human exceptionalism'). This is not to suggest, however, that it has a wholly introspective gaze. In A.I. (as with alien-themed) SF, we are also frequently invited to an encounter an uncanny 'other', one that is simultaneously like us but not us. And the uncanny power of the fictional A.I. other derives also from its ambiguous status as seemingly both life and non-life, a contradiction that is rarely fully resolved even in relatively traditional narratives of human overcoming.

And whether A.I. is depicted as malevolent, sympathetic or ethically ambiguous, a common thread in the SF narratives discussed above is the notion of a tipping point or historical rupture, what science fiction scholar Darko Suvin (borrowing from Ernst Bloch) calls the 'novum.' Here, the novum is machine intelligence taking on a momentum or 'life' (albeit an ontologically ambiguous one) of its own: to use a biological metaphor to which we will return below, a kind of *autopoeisis*. Whether the machines rise up against humans (as in *Terminator* and *The Matrix*), run amok (as in *2001*), or transcend and become indifferent to the dull limitations of human civilisation (as in *Her* and *Ex Machina*), the unsettling force of such tales lies in the suggestion that the very technologies we have created are no longer beholden to nor controllable by us[5].

Fictional narratives can, of course, perform a cathartic function, helping us channel and even contain the unsettling power of A.I. within the realm of entertainment fiction. Increasingly, however, we are now seeing non-fictional representations of A.I. that, in easily recognisable ways, resemble and resonate with the unsettling images once almost exclusively associated with science fiction.[6] Today, the prospect of autonomous digital 'superintelligence' (Bostrom 2014) that can teach *itself* skills and which could 'evolve' independently of us (especially through advances in the field of machine learning), is creeping into popular consciousness as an increasingly 'plausible' scenario, through the consumption and sharing of technology news stories, viral videos and other kinds of everyday media. It's this everyday non-fictional realm to which my discussion turns next.

Tales from the uncanny valley

The spectre of superintelligence is now a growing presence in non-fictional genres including popular science and philosophy, technology journalism, and online videos and blogs, reflecting accelerating developments in A.I. and robotics. This is not a neutral reflection, however. The popular imagination around A.I. is developing in the context of a sensationalist, marketing-driven and viral (or meme-based) online attention economy. Popular consciousness of, fascination with, and anxiety towards A.I. developments, is fuelled by media texts that lend themselves to the sharing culture of the internet: typically short, attention-grabbing media (especially videos) that, while situated in the non-fictional space, nonetheless resonate strongly with the tropes of science fiction.

For example, Amazon's virtual 'smart home' assistant Alexa has garnered attention as stories emerge of children forming relationships and attachments to it. Another attention-grabbing story laced with uncanny suggestiveness told of Alexa 'eavesdropping' on conversations and interpreting them as instructions, leading it to make comically misguided online purchases on users' behalf. Recently, the first pedestrian death at the hands of a self-driving car (owned by Uber) generated widespread and sensationalist coverage in the international media: what got downplayed amid the clickbait hype was the rather mundane reality that the pedestrian's actions made it impossible for any driver, whether human or A.I., to take the necessary evasive action in time to avoid the crash. There has also been growing international media coverage in recent years about the spectre of Lethal Autonomous Weapons Systems (LAWS) which have been given the sensationalist moniker of "Killer Robots," although this is a more slow-burning story as it deals with what are, as yet, hypothetical future scenarios: current military technology deploys A.I. gathering intelligence (via drones, for example) that *contributes* to 'kill lists' but, at the time of writing, we have yet to see examples of weaponised A.I. systems making autonomous decisions to kill or attack. The Alexa and self-driving car examples are, by contrast, more tangible and time-bound media 'events'. I want to focus below on a small number of other media events that have garnered significant public attention. The examples included here, despite being situated in the non-fiction space, share with SF a tendency to greet the emergence of an apparently radical new technology (something that appears to audiences as a potential social disruptor, a real-world 'novum') with fascination and sometimes even optimistic celebration, while at the same time evoking an unsettling sense of creepiness. Indeed, 'creepiness,' which can be understood as a vernacular and more loosely articulated version of the 'uncanny,' is an extremely common term that regularly recurs in popular discussion of these media events (evident, for example, in online user comments surrounding these media texts).

While virtual assistants, self-driving cars and autonomous weapons systems are examples of narrow purpose A.I. (i.e. designed with specific functions and capabilities in mind), the field of *general* purpose A.I. (A.G.I) that could, in theory, learn a wide range (if not the entire repertoire) of human skills and behaviours, has the capacity to stir public excitement and anxiety even more deeply. Alphabet/Google's A.I. subsidiary DeepMind has been successful in capturing public attention in recent years. Whether or not DeepMind embodies the most advanced research on the road towards full AGI (which remains hypothetical at this stage) or whether its public prominence reflects the powerful marketing and public relations resources Google/DeepMind has at its disposal is open to question. But its capacity to capture the popular imagination is undeniable. DeepMind is significantly increasing popular awareness of A.I.'s potential for learning complex behaviours and tasks with relative autonomy, that is, without requiring its creators to programme (or even envision it) it for specific purposes.

Chess has long been one of the frontier public- and media-friendly challenges in the field of A.I.[7] IBM's watershed moment occurred in 1997 when, to a fanfare of media publicity, its supercomputer beat world chess master Garry Kasparov. Recently, DeepMind, whose founder Demis Hassabis is a former child chess prodigy, made this once stunning feat appear humdrum, when it announced that its AlphaZero A.I., having only been given the basic rules and with no pre-programmed tactical intelligence, *taught itself* to play chess from scratch in just four hours, to the point that it could then beat the hitherto most powerful chess computer. AlphaZero has also beaten the world champion at Go, a game with many trillion times more potential scenarios (what A.I. developers would call 'degrees of freedom') compared to chess – more, in fact, as many of the news stories about the event highlighted, than there are atoms in the known universe. And while chess is used to showcase superhuman feats of logical reasoning, Go is widely considered a game that can only be won through combining mathematical reasoning with sophisticated *intuition*: engineers had to think beyond 'brute force' approaches (whereby the A.I. would search through all potential options before selecting the best one) in favour of complex pattern recognition, enabling it to learn over time to make contextualised judgements about what constitutes a 'good' or 'bad' move in specific situations. In an interview with *Nature* (2016), Hassabis emphasises that, in contrast to chess, professional Go players will typically answer a question about why they made a move with an answer along the lines of 'because it *felt* right'. Lead programmer David Silver adds that this new frontier of A.I. capabilities should be understood as 'akin to human *imagination*,' as opposed to brute force logic or clinical reasoning. Unsurprisingly, this event provoked a number of sceptical responses from A.I. experts who sought to disentangle the science (and certain questions left unanswered by DeepMind) from the hype (Collados 2017).

But that didn't prevent it capturing the public imagination and garnering extensive international news coverage with sensationalist headlines appearing even in 'quality' news outlets, for example: 'The A.I. that has nothing to learn from humans' (*The Atlantic*, October 20, 2017); and "'It's able to create knowledge itself': Google unveils A.I. that learns on its own" (*The Guardian*, October 18, 2017).

The Go event was a striking PR platform for Google/DeepMind to reinforce its cutting edge reputation: a series of five games against world Go champion Lee Sedol of South Korea was broadcast live in several countries and streamed online, with AlphaZero winning all but one of the series. But however effective the PR, this media event did more than just bolster Google's reputation for technological 'mastery'. Its power as a spectacle lay also in its unsettling force. The event was quite explicitly framed by DeepMind as one that undermines any comforting sense of a clear boundary between (human) intuition and (programmable) logic. Hassabis also claimed (whether sincerely or not) that AlphaGo *surprised* its own programmers with the scale and speed of its achievements. In a science fictional vein, the event at least hinted at a sense of unleashing something whose ultimate trajectory not even they (let alone we) can predict. As reflected in the vast online commentary around the videos, the project managed to inspire both a sense of awe and an unsettling sense of science venturing into uncharted territory. As a piece of marketing, it treads a fine line between conveying a positive, reassuring sense of powerful technology in safe expert hands on the one hand, and a more vertiginous sense of mystery, adventure and risk on the other.

In contrast to the disembodied AlphaZero, two more recent media events that went viral through social media, as well as garnering substantial mainstream news media attention, involve the physical (robotic) embodiment of A.I. On the one hand, an invisible A.I. such as Deep Mind's may draw some of its unsettling force from its intangible nature. On the other hand, physical robots lend themselves to another kind of uncanny encounter: this is reflected in the roboticist concept of the 'uncanny valley' (Mori 2012 [1970]). In short, the uncanny valley is a kind of negative Goldilocks zone, in which the design of a (typically but not exclusively humanoid) robot generates discomfort in those who encounter it because it is simultaneously too lifelike to perceive as merely a safely inanimate object, yet insufficiently lifelike to appear fully 'alive': uncanny robots are incongruous, even abject entities that challenge our sense of an easy division between the animate and the inanimate. While the uncanny valley is, in the first instance, a matter of visual design, it is the animation of robots through A.I.-driven motion, gesture and speech that fuels a robot's uncanny force, especially when they are programmed to be highly responsive to human interlocutors and, powered by machine learning, able to give the appearance of forming particular, concrete relationships with specific individuals.

There is, once again, a gendered dimension to this. The most prominent recent exemplars of uncanny humanoid A.I. include 'Erica', produced by roboticists at Osaka and Kyoto universities, as well as the rumbling moral panic emerging over the development of (predominantly 'female') A.I. driven 'sexbots'. But probably the most impactful recent media event in this arena has been that of Sophia, the android (sometimes known by the gendered term 'gynoid') produced by Hong Kong-based Hanson Robotics, that (or 'who') is the first A.I. granted 'citizenship' by a state – in this case by Saudi Arabia in a clever piece of marketing by a country keen to rebrand itself on the global stage. The true legal status of Sophia's citizenship is debatable, although when the issue of robotic 'personhood' is also being seriously debated in the European Union (Withers 2017), we should not be too quick to discount the possibility. Sophia has successfully gained widespread international media attention, appearing not only in the news but as a guest on talk shows, including the popular Jimmy Fallon show in the US. Sophia was even invited to give an address at the United Nations, and spoke to delegates about world peace and the positive contribution A.I. could make to the world. But if Saudia Arabia was looking for an unambiguously positive story to vaunt its hi-tech credentials, Sophia's global reception was more complex. Sophia displays highly advanced, but still not fully convincing, A.I. driven conversational faculties, and her facial gestures are widely received as 'creepy' – while highly advanced and subtle relative to current standards in robotics, they are disconcertingly 'off' when measured against the yardstick of actual human interaction. It is worth noting here that former Disney 'visioneer' David Hanson, founder of Hanson Robotics, has perviously published research arguing that the 'uncanny valley' should not be treated only as a negative pitfall for roboticists to avoid: Hanson and his colleagues argued, in fact, that embracing the uncanny can allow robotics to aspire to the status of 'art' (Hanson et al 2005). Sophia's creepiness, in other words, looks more like a feature than a bug.

Beyond the sphere of highly realistic humanoid or feminised A.I., US-based Boston Dynamics has been perhaps the most successful company vying for viral visibility to enhance its public profile and funding streams (it is currently funded by the DAPRA, the US Defense Advanced Research Projects Agency), periodically releasing videos onto YouTube featuring its latest cutting edge robots. Boston Dynamics does not compete in this attention economy on grounds of humanoid or biological realism. Rather, their loosely biomorphic creations are hyper-technologised, with 'skeletons' and working parts thoroughly exposed. For example, Atlas, one of its most high profile creations, is only a loosely humanoid biped whose design evokes various science fiction robots such as the eponymous *Chappie* (2014), but in rudimentary form. Other Boston Dynamics creations draw inspiration for both their skeletal topographies and their names from elsewhere in the animal

kingdom, including dogs and cheetahs, while disavowing visual realism.

What Boston Dynamics showcases is not feats of communication, cognition or reasoning, but A.I.-driven feats of motion, movement and dexterity. Advanced dexterity and complex motion, according to British roboticist Jeremy Wyatt (2018), can be understood as the cutting edge 'third wave' of robotics. Simplified, the first wave prioritised complex movement within controlled environments such as robots in factories than can outstrip the human capacity for speed, accuracy and cost in the context of complex but repetitive tasks, such as those involved in assembling a modern car. Wyatt's second wave sees robots equipped with advanced A.I. increasingly moving out into uncontrolled environments where, programmed with contextual decision skills, they can execute tasks even when faced with unexpected obstacles: robotic vacuum cleaners and lawnmowers, as well as self-driving cars, exemplify this second wave. Wyatt sees the third wave, which companies like Boston Dynamics as well as his own UK robotics lab prioritise, bringing together dexterity, motor skills and complex motion (akin to that found in the animal world) with advanced A.I. and machine learning[8]. A task such as loading and unloading a dishwasher, cracking an egg into a pan, or walking across uneven terrain, comes easily to able-bodied adult humans, but has hitherto been incredibly difficult to operationalise in robots, requiring both advanced artificial intelligence and extremely complex mechanics.

With their exposed workings, the Boston Dynamics creations are, in one sense, raw and unfinished designs from a company ostensibly more concerned with function than form, or aesthetic design. The company's public profile is less about finished *products* than showcasing cutting-edge science and continual development through presenting consistently improved prototypes – prototypes that are typically shown being tested to their limits in public videos, including some depiction of failing points (robots toppling over, for example) to emphasise the process of continual refinement.

Yet aesthetics are significant when it comes to understanding the affective potency of the images circulated by Boston Dynamics, as reflected in the viral successes of their videos and the popular responses to them. The uncanny (or 'creepy') facet of these creations lies primarily in their movement, rather than in anatomical, physiognomic or vocal realism. Their stripped back, hypermechanised appearances – rather than appearing as a limitation – lend a monstrous dimension to their *almost* lifelike dexterity and motility. In one of the mostly widely shared videos to date, Atlas performs a gymnastic blackflip that is remarkable for its human-like imperfection. Far from a gold medal-winning performance, slow motion replays in the video accentuate the imperfections as Atlas lands slightly off balance and viewers are left to marvel at the recovery as he eventually finds his feet and avoids toppling over. In numerous other videos we see Boston Dynamics staff

testing the limits of their robots by pushing, prodding and even kicking the robots while in motion to see how well they can withstand the unanticipated challenges. The images are reminiscent of scenes of animal abuse or militaristic torture. One of the stark achievements of Boston Dynamics has been to create robots whose apparent 'ill-treatment' can cause widespread unease and discomfort among viewers who, while perhaps not holding any outright cognitive belief in the sentience or sufferance of these machines, nonetheless express discomfort because the videos gnaw away at the sense of a neat and stable boundary between life and non-life.

Another prominent theme of the online commentary surrounding Boston Dynamics' creations is the sense of both exhiliaration and anxiety surrounding the speed with which the technology appears to be developing, something that some robotics companies are keen to play up, even while other roboticists are critical of the media hype and outbursts of 'exponentialism' (Brooks 2017). This fascination with exponentialism and the possibility that we are facing a 'tipping point,' whereby A.I. developments will 'take off' at breakneck speed, is fundamental to the final area that I want to discuss: the futurological discourse of the 'singularity.'

Is the Singularity Near?

The technological singularity, a notion originally coined by science fiction writer and computer scientist Vernor Vinge, is a predictive hypothesis premised on exponential and therefore dramatically accelerating technological advances. As a metaphor drawn from astrophysics – where a singularity is the centre of a black hole in which the laws of time and space no longer hold – the technological singularity is a powerfully evocative story about the future. Futurists who subscribe to the notion of the coming singularity (what I will call the singularitarians) see us approaching an event horizon, an irreversible tipping point beyond which we will be powerless to resist the gravitational pull of a technological revolution vastly more radical than any other in history, and which promises to fundamentally change everything.

Proponents of the singularity are especially excited by the promise of an A.I. 'explosion', an idea orginally proposed by mathemetician I.J. Good (1965), substantially before the word 'singularity' took root in the field of A.I. But singularitarians see A.I. as part of a cluster of interlinked technologies involving fields such as biotechnology, robotics, neuroscience and nanotechnology. These fields, along with computing, are subject to 'laws of accelerating returns.' (Moore's Law, which postulates a rough doubling of computer chip performance every two years, is the most well-known example of this[9].) This is not distant future-gazing. 'The singularity is near,' according to the title of the most popular and influential book on the subject (Kurzweil 2005). Its author, Ray Kurzweil, futurist and Director of Engi-

neering at Google, predicts it will occur by 2045. By then, the rate of exponential technological advancement will have become dizzyingly fast, and A.I. will rapidly outstrip human cognitive capacities. In other words, 'superintelligence' (Bostrom 2014) will have arrived, at which point all bets are off.

To those who object that the capacity of our machines will always be limited by the human minds that design, programme and control them, singularitarians respond that they are missing the point: computer intelligence – and even to call it 'artificial' intelligence reflects a narrow and outmoded ontology from the perspective of singualrity discourse – will begin to design, modify, upgrade and replicate itself. This invokes an image of 'autopoiesis' characteristic of biological systems, of technology 'coming to life' – life, perhaps, but not as we know it. The speed and the intentional, goal-oriented mode of self-improvement implied in this vision of technological evolution (perhaps better understood as henceforth ceaseless *revolution*) would be unlike any known organic life form. Our limited human brains, it seems, can scarcely begin to fathom what kind of world this might lead us to. However, for singularitarians, this incomprehension is just one more engineering problem waiting to be solved. Computational metaphors of mind and brain dominate the contemporary neuroscientific imagination, just as mechanical and hydraulic metaphors have held sway in earlier technological epochs. But singularitarians take them seriously to the point of literalisation. Human minds, currently encased in limited capacity 'wetware,' verge on obsolescence. Thankfully, we humans also stand to benefit from a revolutionary upgrade as we begin to merge with (and into) networked computational systems: by the 2030s, Kurzweil assures us, the human neocortex will be connected to the cloud (Galeon and Reedy 2017).

So humans need not be left behind by this technological explosion. Our intellectual and communicative capacities can soar along with the machines, so long as we are prepared to join with them and to become fully post-human. Only then can we overcome the limitations and inefficiencies of bodies, brains, human languages and other such cumbersome devices. Celebrity physicist, Michio Kaku is, like Kurzweil, excited about the coming singularity: he concludes a TV documentary on the subject (Science Channel 2010) by telling his audience that instead of fearing an uprising of the machines, we should devote our energies to a forthcoming merger: once our networked minds can instantaneously access any and all knowledge, we will have become "like the Gods", he declares, a new species of *homo superior*. He imparts these words before an enrapt congregation of toy light sabre-wielding fans, who whoop and cheer at the Good News. Even Elon Musk, who sees the singularity as a potentially mortal threat to humanity, believes our salvation lies in a cyborg merger with machine intelligence enabled by technologies such as his much-hyped vapourware (i.e. hypothetical) 'neural lace', a technology originally posited playfully, rather than seriously, in the science fiction 'Culture'

novels of Iain M. Banks.

It is hard to resist the conclusion that the technological singularity is as much prophecy as prediction, despite the best efforts of some singularitarians to disavow embarrassing connotations of cultish religiosity that might undermine their emphasis on cutting edge science. The singularity has been mocked by many as a 'rapture of the nerds' (MacLeod, 1998). Specifically, some commentators have identified strong resonances with the mystical cult of Gnosticism (Gray 2011; Davis 1998) with its belief that the material world is the work not of a God but rather an evil demiurge, and that our destiny as humans is to use our knowledge to transcend and escape it. Certainly, a strong sense of faith, if not of predestation, is never far beneath the surface of singularitarian rhetoric. 'Most experts agree that the singularity is inevitable', Kaku tells us in the aforementioned documentary. Unsurprisingly, in fact, it's a controversial hypothesis, contestable on both scientific and philosophical grounds. Its power as *prophecy*, though, depends on the repeated incantation of inevitability, a sense that the singularity is somehow written in the stars.

It is quite possible, of course, to believe that the singularity is to be feared, rather than celebrated, without shaking that narrative of inevitability. And, reflecting the polarities of the technological sublime, we see prophecies of doom vying with those of rapture. Besides Musk, Nick Bostrom (2014), Oxford philosopher of superintelligence, has become progressively worried about the implications, having once entertained a more enthusiastic outlook as a founding figure in transhumanism, a movement dedicated to the technological upgrading of the human species, including overcoming mortality, mind-uploading and A.I.-based governments (O'Connell 2017). Bostrom's concern is not that the machines will rise up against us with malicious intent but rather that, in their ruthless attachment to predetermined goals, they will become *indifferent* to the fate of humans, just as we humans are mostly indifferent to ants on the footpath beneath our feet as we go about our daily business. In an echo of the 'grey goo' apocalyptic scenario used by nanotechnologist Eric Drexler (1986) to warn us of the perils of self-replicating nanobots, Bostrom offers us the parable of the paperclip maximiser. With a banal absurdism that would not be out of place in the SF satire *Hitchhiker's Guide to the Galaxy*, he asks us to imagine that we task a 'superintelligent' machine with collecting as many paperclips as possible. But imagine also that we were not smart enough to properly anticipate and set strict limits on its methods. Before too long, the A.I. may have destroyed the entire world by transforming it into a giant paperclip factory. Via an ostensibly silly example, his point is that, as humans, we are inevitably constrained in our capacity to foresee and prevent the kinds of risks that only superintelligent machines could pose.

Various other prominent figures from the world of science and technology

(Stephen Hawking and Bill Gates, for example) have become increasingly vocal in recent times about the catastrophic risks posed by rapid advances in A.I. Differing proposals for precautionary action and risk mitigation have emerged from this growing chorus of concern. But none suggest we can simply halt the march of A.I., or that the rise of superintelligence is anything other than inevitable. In that sense, these prophets who bring us warnings from the future, share entirely with their optimistic counterparts a message that *this* is the future on which our eyes must be trained, the future for which we must prepare. Whether you believe that superintelligence and other singularity technologies promise us redemption or threaten our very existence (or that they contain both potentials, depending on how we handle it), you believe that this is one of the most pressing issues facing humanity today.

This raises a political problem, of course, whereby the singularity vies for attention, prominence and research investment with other, more obviously credible, challenges facing humanity. In a supreme example of Silicon valley 'solutionism' (Morozov 2014), Kurzweil himself has claimed that climate change, while real, is not such an urgent problem as scientists have suggested—we have 'plenty of time', he says, to solve the problem with clean technologies such as solar power which, like singularity technologies, are subject to laws of accelerating returns (Feeney 2011). Musk and Hawking, both vocal about and far more troubled by climate change than Kurzweil, have nonetheless framed A.I. as a comparably serious threat (Cellan-Jones, 2016; Leary, 2017). The currently influential public intellectual Yuval Noah Harari (famous both as macro-historian and futurist – see Harari 2015) similarly talks of A.I. and climate change as existential threats of comparable magnitude.

But the issue is not just about A.I. competing for attenion with other issues within a futural public sphere. This type of discourse frames the problem of A.I. in a very specific and technocentric way. Questions of ownership, control, equity and public and/or legal oversight of the technology (that is, public interest perspectives on A.I.) are already urgent: we are increasingly witnessing various detrimental consequences (from social division, to racial profiling) associated with algorithms, everyday forms of A.I. that are less visible and less spectacular than developments in either robotics or AGI, but which are *already* a hugely significant part of our digital social infrastructure (Pasquale 2016; O'Neill 2016). Questions about how we avoid a future in which these problematic (and opaque) consequences are simply exacerbated, or about how we might better harness algorithmic technologies for social good, are not the the kinds of questions about A.I. encouraged by either optimistic or pessimistic variants of singularity discourse. Whether it evokes the uncanny or the sublime, the spectre of a potential new 'lifeform' that may be our salvation or our destruction, is inevitably more immediate, more gripping, than

debates about the messy realities of our current technosocial formation.

Both the warnings *and* the promises of the singularity are apocalyptic. Both augur an 'end of times' – where they differ is in what comes next. The dire warnings of Musk and Hawking clearly fit the popular, everyday notion of 'apocalypse' as a disastrous collapse of civilisation and even potential human extinction. Kurzweil's dreams, on the other hand, are apocalyptic in the more technical and biblical sense of the word: as an 'unveiling' of a new world and of a Truth that was previously hidden. But with a raft of serious global challenges upon us, it becomes politically salient to question the value of dedicating time, energy and money preparing for this particular apocalypse, regardless of whether it represents a dream or a nightmare.

The emphasis on preparation implies that singularity discourse is not necessarily a fatalistic one, despite the frame of inevitability. It is a call to action. In that sense, I suggest that, more than just a prophecy, the singularity is also an animating *myth* for the digital age. Not a myth in the sense of an illusion or falsehood (though it may well be this too), but in the sense of a story, told and retold, morphing through this process of retelling. Mythic (including religious) narratives help us to cope with our finitude, serving as cognitive and moral resources through which we can imperfectly grasp otherwise unfathomable cosmic complexity and mystery. An ancient feature of human community, myths find their power when they animate our emotions (especially our hopes and fears) and provide us with temporal meaning and anchorage against the threatening spectre of random flux and impermanence: they evoke journeys (cyclical and linear), destinies, fates. Singularity dicourse embodies a sense of sublime mystery (we cannot fathom what lies beyond the event horizon) and can be understood as a post-secular myth born of the vertiginous (both thrilling and terrifying) accelerations of digital modernity. And it evokes a momentous journey, whether as the realisation of our post-human destiny (exemplified by the transhumanist movement), or as a heroic struggle against an existential threat.

While some myths help us cope with our human finitude by teaching us to accept, embrace or even to find beauty in it, Western monotheistic traditions, by contrast, typically foster a spirit of transcendence and treat our finite earthly existence as merely a prelude to or audition for a timeless and heavenly coexistence with God. Singularity discourse is clearly the offspring of this latter monotheistic tradition, whatever its atheistic pretensions, even though New Age tropes are also a part of the mythological package – promises, for example, of a new and collective (digitally networked) consciousness, unbounded by the individual ego. (Silicon Valley techno-rationalism and New Age philosophies are not new bedfellows, of course, from the digital psychedelics once promoted by Timothy Leary to the contemporary paganistic Burning Man festival). This mix of elements makes for a

potentially rich and seductive digital age mythology.

While myths are an enduring, and arguably necessary, feature of human existence, they always risk obscuring as much as they reveal. The myth of the singularity addresses us as *humans*, not as members of particular socio-economic, national, cultural or other groups. It asks us to envisage what will become of humans in the future, but we need to go beyond that question and ask: which humans? Who will want to live in this future? Will it be a future in which only the elite thrive or even survive? Are the rest of us obsolete? And what kind of power will accrue to the corporations that patent and control the rapidly advancing technologies shaping our experience, our existence and our consciousness? These are potentially useful questions, irrespective of whether the singularity and superintelligence are fully credible as scientific hypotheses. So the question is not only whether or not singularity discourse should be exposed or debunked as 'myth', but also whether, as an undeniably evocative story, it might be re-geared such that it prompts us to open up, rather than shut down, such lines of critical questioning and reflection.

Conclusion

In this essay, my aim has been to explore three key strands of discourse that loom large in the popular imagination around A.I. and where it might take us in future. Clearly, these three strands – science fiction, popular news, and futurological discourse – are interwoven. Science fictional tropes abound in news coverage of A.I. and popular futurology. And central to the discourse is the uncanny (or 'creepy') spectre of an emergent entity that sits uneasily between a (Western) dichotomy of life and non-life, one that appears capable of harbouring an apocalyptic capacity to overturn human life as we know it. I want to argue that these non-rational elements in the discourse (the emotive, the mythic, or even the quasi-theological) are more than simply distortions or distractions from what might otherwise be a sober and rational public debate about the future of A.I. Not only would that seem an unrealistic aspiration, it would also be to neglect a fundamental premise of critical future studies, namely that public engagement with the future demands imaginative and exciting (which is not to say exclusively optimistic) visions of the future that can motivate and energise – this is why science fiction as a genre, if not in all its individual instances, must be taken seriously as part of a futural public sphere. While demystification may be a vital aspect of any reasoned public debate, motivating broad public interest in a topic in the first place demands evocative storytelling (cf. Monbiot 2017). In principle, perhaps, we might even go further and entertain the counterintuitive possibility that *mystification* serves a potentially valuable function in its own right (as opposed to simply a means to garner engagement). For example, perhaps an appreciation for the 'uncanny' is warranted. The

notion that current deployments of A.I. (e.g. online recommendation engines or crime prediction software) are 'creepy' is more than just a lazy shorthand. It is a growing public sentiment that attests to the very real opacity of technologies whose inner workings are, from the point of view of most citizens, hidden not only by corporate secrecy but also by intractable technical complexities such that the demand for greater 'transparency' is increasingly problematic. This appreciation for uncertainty and unknowability is salient to the field of A.I. (see Bridle 2018) but also, as Riel Miller argues (2018), for the broader development of 'futures literacies' which, in today's world, demand an attunement to uncertainty, 'complexity'[10] and 'emergent phenomena,' from developments in A.I. to the hazards of climate change.

Nonetheless, we also need to be attuned to the various problematic aspects of the popular discourse around the future of A.I. That discourse is skewed heavily towards specific voices – predominantly male science fiction authors and techno-centric scientists, futurists and entrepreneurs – and the field of A.I. and robotics is all too easily presented as a kind of sublime spectacle of inevitability (cf. Kelly 2016) that does little to offer lay citizens the sense that they can be actively involved in shaping its future. I have not ventured to evaluate potential remedies that might diversify the imaginative canvas – such a task is beyond the scope of a single paper. But I have sought, at least, to articulate a *critical appreciation* for both the seductiveness of and some of the blindspots in the discourses of A.I. currently prominent in the public sphere.

Luke Goode is an Associate Professor of Media and Communication in the School of Social Sciences, University of Auckland, Aotearoa New Zealand. E-mail: l.goode@auckland.ac.nz

Notes

[1]While beyond the scope of this paper, it's worth noting that not all the concerns about over-hyping relate to the quality of public understanding. Some commentators fear, in fact, that over-hyped discourse may ultimately *slow down* the rate of technological progress by contributing to a new 'A.I. Winter' (previous ones occurring in the 70s and late 80s), in which a bubble of excitement-driven investment and funding collapses in the disillusionment of technology failing to live up to its hype.

[2]The commonplace distinction between fear, whose object is tangible (such as a murderous robot), and anxiety caused by a threat unknown or unknowable (such as the potential behaviour of a 'rogue' A.I.) is clearly relevant to the present discussion.

[3]I refer to this as a Western conception because such a clear distinction is not, for example, as prevanlent in the Japanese popular imagination around robotics, where the

cultural influence of animism plays a significant role.

[4]The contentious question of whether Garland's film actually does more to challenge or reinforce problematic representations of women in the history of science fiction cinema lies beyond the scope of this paper.

[5]This, of course, is a central theme of science and technology studies (e.g. Winner 1977), except that the latter does not treat it as a looming existential threat but instead treats the co-constitution of human and non-human actors as an ontological fact and our fantasies of separation and mastery as a culturally and historically specific quirk of Western modernity – in the words of Bruno Latour (1991), "we have never been modern".

[6]This is not to deny the long history of cross-fertilisation between SF and popular science genres.

[7]That genealogy is longer than one might expect: 'The Turk' was a chess-playing automaton of the late 18th century (one that transpired to be a hoax).

[8]Due to corporate secrecy, it is not clear whether Boston Dynamics' current range of robots deploys 'true' machine learning but this should be understood as the general goal of third wave robotics in Wyatt's formulation.

[9]Whether or not Moore's Law still holds today and/or will apply in the future is subject to debate. See, for example, Sy Taffel's article in this issue.

[10]Miller distinguished complex, from merely complicated, phenomena in that the former implies changes even in the conditions of change – something that could easily be applied to the notion of advanced A.I. deployed, in principle, to upgrade and redesign itself. This notion of changes in the conditions of change has also been emphasised (as 'metamorphosis') in the later work of social risk theorist Ulrich Beck (2015).

References

Bartlett, Jamie (2018): *The People vs Tech: how the internet is killing democracy*, London: Penguin.

Beck, Ulrich (2015): "Emancipatory catastrophism: What does it mean to climate change and risk society?," *Current Sociology*, 63:1, 75–88.

Bostrom, Nick (2014): *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.

Bridle, James (2018): *New Dark Age: Technology and the End of the Future*, London: Verso.

Brooks, Rodney (2017): "The Seven Deadly Sins of AI Predictions," *MIT Technology Review*, 6 October 2017: https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/, (accessed 12/06/18).

Capek, Karel (2004 [1921]): *R.U.R. (Rossum's Universal Robots)* [trans. Claudia Novack-Jones], New York: Penguin Books.

Cellan-Jones, Rory (2016): "Stephen Hawking – will AI kill or save humankind?," BBC Online, 20 October 2016: http://www.bbc.com/news/technology-37713629 (accessed 14/07/18).

Colladas, Jose Comacho (2017): "Is AlphaZero really a scientific breathrough in AI?,"

11 December 2017: https://medium.com/@josecamachocollados/is-alphazero-really-a-scientific-breakthrough-in-ai-bf66ae1c84f2, (accessed 20/06/2018).

Davis, Erik (1998): *Myth, Magic, and Mysticism in the Age of Information*, New York: Harmony Books.

Drexler, Eric (1986): *Engines of Creation: The Coming Era of Nanotechnology*, New York: Doubleday.

Elish, M.C. & boyd, danah (2018) "Situating methods in the magic of Big Data and AI", *Communication Monographs*, 85, 57-80.

Feeney, Lauren (2011): "Climate change no problem, says futurist Ray Kurzweil," *The Guardian*, 21 February 2011: https://www.theguardian.com/environment/2011/feb/21/ray-kurzweill-climate-change, (accessed 20/08/18).

Galeon, Dom and Reedy, Christianna (2017): 'Kurzweil Claims that the Singularity Will Happen by 2045,' *Futurism*, 5 October 2017: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045/, (accessed 27/07/18).

Good, I.J. (1965): "Speculations concerning the first ultraintelligent machine," Institute of Electrical and Electronic Engineers: https://web.archive.org/web/20010527181244/http://www.aeiveos.com/~bradbury/Authors/Computing/Good-IJ/SCtFUM.html, (accessed 25/06/18).

Goode, Luke and Godhe, Michael (2017): "Beyond Capitalist Realism: Why We Need Critical Future Studies", *Culture Unbound*, 9:1, 108-129.

Gray, John (2011): *The Immortalization Commission: Science and the Strange Quest to Cheat Death*, London: Allen Lane.

Hanson, David, Olney, Andrew, Pereira, Isma A. & Zielke, Marge (2005): 'Upending the Uncanny Valley', Association for the Advancement of Artificial Intelligence: aaai.org/Papers/Workshops/2005/WS-05-11/WS05-11-005.pdf, (accessed: 20/08/18).

Harari, Yuval Noah (2015): *Homo Deus: A Brief History of Tomorrow*, London: Harvill Secker.

Kelly, Kevin (2016): *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, New York: Penguin.

Kurzweil, Ray (2005): *The Singularity is Near: When Humans Transcend Biology*, London: Penguin Books.

Latour, Bruno (1991): *We Have Never Been Modern*, Cambridge, MA: Harvard University Press.

Leary, Kyree (2017): "Elon Musk Outlined the Two Critical Things Threatening Humanity", *Futurism*, 21 November 2017: https://futurism.com/elon-musk-outlined-two-critical-things-threatening-humanity/, (accessed 27/07/18).

Leonhard, Gerd (2016): *Technology vs. humanity: the coming clash between man and machine*, London: Fast Future Publishing.

Lovén, Svante (1999): "Den tänkande maskinen: Datorerna i science fiction", *Tidskrift för litteraturvetenskap*, 28:3–4, 19–38.

MacLeod, Ken (1998): *The Cassini Division*, New York: Tor Books.

McCauley, Lee (2007): "The Frankenstein Complex and Asimov's Three Laws", Association for the Advancement of Artificial Intelligence: https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-003.pdf, (accessed 27/07/18).

Miller, Riel (2007): "Futures literacy: A hybrid strategic scenario method," *Futures*, 39:4, 341-362.

Miller, Riel (2018): *Transforming the Future: Anticipation in the 21st Century*, New York: UNESCO and Routledge.

Monbiot, George (2017): *Out of the Wreckage: A New Politics for an Age of Crisis*, London: Verso.

Mori, Masahiro (2012 [1970]): "The Uncanny Valley" [trans. Karl F. MacDorman & Norri Kageki], *IEEE Robotics & Automation Magazine*, 12 June 2012: https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley, (accessed

20/08/18).

Morozov, Evgeny (2014): *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York: Public Affairs.

Mosco, Vincent (2004): *The Digital Sublime*, Cambridge, MA: MIT Press.

Nature (2016): "The computer that mastered Go": https://www.youtube.com/watch?v=g-dKXOlsf98&t=252s, (accessed 27/07/18).

Nye, David (1994): *American Technological Sublime*, Cambridge, MA: MIT Press.

O'Connell, Mark (2017): *To Be a Machine: Adventures Among Cyborgs, Utopians, Hackers, and the Futurists Solving the Modest Problem of Death*, New York: Doubleday.

O'Neill, Cathy (2016): *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Penguin.

Pasquale, Frank (2016): *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, MA: Harvard University Press.

Schwartz, Oscar (2018): "'The discourse is unhinged': how the media gets AI alarmingly wrong", *The Guardian*, 25 July 2018: https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong, (accessed 25/07/18).

Science Channel (2010): *Sci-Fi Science: A.I. Uprising*.

Suvin, Darko (2016 [1979]): *Metamorphoses of Science Fiction*, Bern: Peter Lang.

Turkle, Sherry (2011): *Alone Together: Why We Expect More from Technology and Less from Each Other*, New York: Basic Books.

Winner, Langdon (1977): *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*, Cambridge, MA: MIT Press.

Withers, Rachel (2017): "The EU Is Trying to Decide Whether to Grant Robots Personhood," *Slate.com*, 17 April 2017: https://slate.com/technology/2018/04/the-eu-is-trying-to-decide-whether-to-grant-robots-personhood.html, (accessed 17/04/17).

Wyatt, Jeremy (2018): "What can we expect from the coming wave of 'dexterous robots'?," *TheBigQ.com*, 1 February 2018: https://www.thebigq.org/2018/02/01/what-can-we-expect-from-the-coming-wave-of-dexterous-robots/, (accessed 12/02/18).