Artificial intelligence in fiction: between narratives and metaphors

Isabella Hermann[1]

Abstract

Science-fiction (SF) has become a reference point in the discourse on the ethics and risks surrounding artificial intelligence (AI). Thus, AI in SF—science-fictional AI—is considered part of a larger corpus of 'AI narratives' that are analysed as shaping the fears and hopes of the technology. SF, however, is not a foresight or technology assessment, but tells dramas for a human audience. To make the drama work, AI is often portrayed as human-like or autonomous, regardless of the actual technological limitations. Taking science-fictional AI too literally, and even applying it to science communication, paints a distorted image of the technology's current potential and distracts from the real-world implications and risks of AI. These risks are not about humanoid robots or conscious machines, but about the scoring, nudging, discrimination, exploitation, and surveillance of humans by AI technologies through governments and corporations. AI in SF, on the other hand, is a trope as part of a genre-specific mega-text that is better understood as a dramatic means and metaphor to reflect on the human condition and socio-political issues beyond technology.

Keywords AI narratives · Artificial intelligence · Robots · Science-fiction · Science communication

1 Introduction: AI narratives and science-fictional AI

In 2018, German Chancellor Angela Merkel met the humanoid robot *Sophia* produced by Hanson Robotics for a conversation event; the year before Sophia became citizen of Saudi Arabia as the first robot being granted the right of citizenship of a country. Another year earlier, in 2016, the software *Alpha Go* beat the world champion Lee Sedol in the board game Go—20 years after *Deep Blue* won against the then world champion in chess Gary Kasparov in 1996/1997. The progress in the broad field of artificial intelligence (AI) seems to be catching up with many films of the science-fiction (SF) genre, in which we have been watching humanoid robots and powerful computers for decades. The fictional android *Data* from the *Star Trek* franchise, for example, is a valuable member of the Enterprise crew and earned 'his' right to personal freedom before a trial; in the same vein, the board computer *HAL 9000* from *2001: A Space Odyssey* appears to have its own will and by playing chess with the human astronauts "depict[s] the future birth of a superior

intelligent being" (Bory 2019: 628)—nowadays, this certainly applies to a software mastering the much more complicated game Go. It is, therefore, not surprising that SF has become a reference point not only in the media, but also for humanities scholars and social scientists on the ethics, opportunities, and risks around AI.

There are two ways to look at SF featuring AI, which I call science-fictional AI: First, it can be viewed as being a substantial part of a larger corpus of AI narratives. Narratives in general are cultural artefacts of various kinds that tell stories, which convey particular points of view or sets of values (Bal 2009). The term AI narratives applies to narratives featuring intelligent machines (The Royal Society 2018: 5), they can be analysed as a reflection of our hopes and fears towards these technologies and thus may shape the development of AI by influencing developers, public acceptance, and policy makers (Cave et al. 2020; Cave and Dihal 2019: 74). In this sense, AI narratives are understood as a serious representation of the potential of real AI and its possible consequences—like foresight or technology assessment. However, science-fictional AI—like the genre of SF in general—is not only about the hopes and fears of the particular technology, but about human dramas for a human audience and readership. From this perspective, it is not AI per se that inspires dramatic stories, but—quite the other

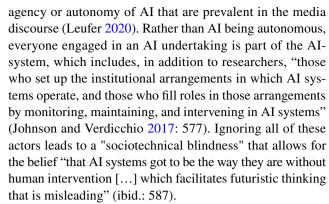✉ Isabella Hermann
   mail@isabella-hermann.com

1   Technical University Berlin, Berlin, Germany

way round—the desire to tell dramatic stories requires certain types of AI, for example humanoid robots or almighty systems. Thus, second, science-fictional AI is not necessarily about the technology but can be a metaphor for other social issues.

This second perspective is relevant, because if it is the case that AI narratives exert influence on AI research, public uptake and political regulation then taking AI in fictional stories too literally can be misleading, because it paints a distorted image of the present potential and functionality of the technology. The UK House of Lords Select Committee on Artificial Intelligence writes in its report AI in the UK: ready, willing, able? (House of Lords 2018: 22):

> The representation of artificial intelligence in popular culture is light-years away from the often more complex and mundane reality. Based on representations in popular culture and the media, the non-specialist would be forgiven for picturing AI as a humanoid robot (with or without murderous intentions), or at the very least a highly intelligent, disembodied voice able to assist seamlessly with a range of tasks. […] this is not a true reflection of its present capability, and grappling with the pervasive yet often opaque nature of artificial intelligence is becoming increasingly necessary for an informed society.

In popular SF, precisely for the sake of dramatic storytelling, AI is often anthropomorphized and given human or even superhuman qualities that exceed the actual capabilities of the technology and can even become magical (Hermann 2020).[1] Unlike *Data* from *Star Trek*, for example, who possesses agency, the actions of *Sophia* are limited and scripted (Estrada 2018). And in contrast to *HAL 9000* from *2001: A Space Odyssey*, which is presented as a faultless assistant, if not even a conscious being, *Alpha Go*—not to mention *Deep Blue*—is a specialised, albeit highly complex, computer programme (Silver et al. 2016) that has nothing of what we would call personal interests or affects, let alone consciousness. Taking science-fictional AI that can develop a will of its own for real obscures the fact that machines do not have intentions and reinforces existing misperceptions about the

agency or autonomy of AI that are prevalent in the media discourse (Leufer 2020). Rather than AI being autonomous, everyone engaged in an AI undertaking is part of the AI-system, which includes, in addition to researchers, "those who set up the institutional arrangements in which AI systems operate, and those who fill roles in those arrangements by monitoring, maintaining, and intervening in AI systems" (Johnson and Verdicchio 2017: 577). Ignoring all of these actors leads to a "sociotechnical blindness" that allows for the belief "that AI systems got to be the way they are without human intervention […] which facilitates futuristic thinking that is misleading" (ibid.: 587).

Hence, sociotechnical blindness obscures the fact that AI systems follow human interests and are embedded in social power structures set up by humans. This is problematic both in terms of a competent and realistic assessment of the opportunities, such as optimizing processes, and the challenges associated with the technology, such as algorithmic biases. Science-fictional AI can serve as techno-scientific inspiration and techno-philosophical thought experiment but taken as foresight or technology assessment it rather distracts from the chances and risks around AI in the real world (Giuliano 2020: 1019). However, SF serves as a distorting mirror and metaphor to reflect on the human condition and socio-political issues in relation to and beyond technology. In this way, *Data* stands for what it means to be accepted as an equal human being (Barrett and Barrett 2001: 87), not for robots becoming human. And among these lines, *HAL 9000* "can be seen as a metaphor for those organizations and societies that cannot admit their flaws, and instead revert to the 'human error' explanation for what may be weak signals of systemic problems" (Shorrock 2013).[2] Thus, even though SF unfolds against the background of technological development, the genre tells stories about current and timeless social issues, which do not necessarily have to do with technology, but find their expression through it.

In this article, I offer alternative interpretations of science-fictional AI, moving away from literal readings towards understanding it as dramatic means and metaphor. In the following Sect. 2, I provide an overview of dramatic and metaphorical readings of AI as a trope and mega-text of science fiction, before discussing modern SF films featuring AI in more detail in Sect. 3. The focus is on films because not only do they visualize AI, but we can also assume that AIs from science-fiction films are generally known to a broader global audience than AIs from literature which is relevant

---

[1] There are, of course, other examples of AI in SF that are not portrayed as human-like or superhuman, but what is referred to here as "popular SF" is mainstream and commercially successful SF—mostly films and series – known to a wider range of people who are not necessarily genuine SF fans. For example, in a representative survey in 2019, the German Informatics Society asked Germans to name the ten best-known and the ten most influential science fiction AIs. Coming out on top in both categories were the Terminator, R2D2 from Star Wars, KITT from Night Rider, Data, Agent Smith from Matrix; and Sonny from I, Robot—all of them either humanoid, human-like with human traits, or endowed with superhuman abilities and/or consciousness (GI 2019).

[2] In this sense, not only can AI be interpreted as systemic, flawed structures detached from humans, but conversely, opaque, and unaccountable systems can be interpreted as black-box AI systems, such as the bureaucratic judicial apparatus in Franz Kafka's Der Process (The Trial) (Hermann 2021).

for the argument of this article.[3] Next, a brief excursus in Sect. 4 critically explores the use of the science-fictional AI trope in science communication films. I conclude with further real-world issues concerning AI and a note not to confuse science fiction with reality in other areas as well such as climate change, space travel, or mega/smart cities. This article contributes to a more nuanced reading of AI in SF, in order not to be distracted from serious socio-political issues regarding AI in the real world, but also not to miss the metaphorical richness of the SF genre when it comes to robots and machines.

2 AI as a mega-text and trope in science-fiction films

SF as a genre emerged in the era of modernity with its social upheavals and belief in technological progress. Basically, SF tells stories about and through fictional technology, but within the prevailing paradigm of scientific thought. Thus, the departure point of SF is a fictional but scientifically explainable novelty, a "novum" (Suvin 1979), which establishes a new world different from the one we know. The scientific foundation, however, does not imply that the novum must be able to truly exist in the real world, but that it is cognitively imaginable within the story world (Roberts 2010: 31, 32). In this way, the novum enables "what if"-questions to speculate about the present and alternative futures in various, but cognitively plausible constellations (Mehnert 2019). In SF, the novum of AI has become a common trope, which is generally understood as a theme or device that is used in a figurative sense, but it can also be overused and become a stereotype or cliché (Merriam-Webster 2021). The different nova and tropes of SF—including also for example space ships or futuristic cities—form part of the SF mega-text, which is composed of the intertextual references and relations of all SF works over time, and understood by the inducted creators and recipients of SF—the "native speakers"—in the "full semiotic density of a given text, most of which will overflow or escape the 'realistically'-sanctioned definitions of the words in the fiction […]" (Broderick 2017: 147). Under the AI trope, I subsume intelligent computer systems, smart machines as well as humanoid robots, in accordance with today's use of AI in the public discourse as well as both in the SF genre itself and in the field of AI narratives research.[4] Generally, against the backdrop of the

mega-text, AI and robots can be analysed in two ways, as dramatic means and as metaphor.

2.1 Dramatic means

SF, like any other genre, with its many formats including literature, comics, games, or movies, conveys dramatic stories that people can identify with. To fulfil its role in a narration, science-fictional AI often possesses qualities that go beyond real-world technological capabilities of a technical artefact operated by algorithms. One can roughly distinguish between two basic storylines in films: AI with a body trying or simulating to be more human, and AI at the level of computer systems that yearn to rule over humans/humanity (Irsigler and Orth 2018).[5] Apparently, AI in the form of a robot is often embodied by real human actors, because—independent of the production budget—if the plot dictates that a humanoid robot should be indistinguishable from real humans, the robot must consequently be played or voiced by a human actor to make the illusion perfect. Examples include next to the aforementioned android *Data*, *Andrew* from *Bicentennial Man*, the robot boy *David* from *A.I.*, the *Cylons* in *Battlestar Galactica*, the *Hosts* in *Westworld*, the *Hubots* in *Real Humans*, the *Synths* in *Humans*, the *Replicants* (even though not AI in the strict sense) and the virtual girlfriend/hologram *Joi* in *Blade Runner 2049* as well as the operating system *Samantha* from *Her*; all are played and/or voiced by real people and undergo human dramas (The Royal Society 2018: 8).

A broadly discussed female film robot—or fembot—of recent years is the character *Ava* from *Ex Machina*. In the film, the Silicon Valley entrepreneur/programmer Nathan develops his latest version of a humanoid robot named *Ava* and brings in his rather shy employee Caleb to test how human-like 'she' is with a "reversed Turing test". *Ava* makes Caleb fall in love with 'her' and in the end fools both men and escapes. The difficult balance of the film is that "[…] the audience has to understand she's a robot, but for the movie to work that idea then needs to fall away, in the same way it does for Caleb" (Bishop 2015). Therefore, *Ava* was not

---

[3] For the same reasons of global prominence of the films and the AI depicted, also mainly AI representations from western culture were used.

[4] It can be criticized that AI and robots are confused, since in the media as well as other reporting and communications AI-software applications are illustrated as quite unrealistic humanoid robots, which is deceptive as to what AI is and is not capable of. There exist

---

Footnote 4 (continued)

projects collecting these misleading illustrations like https://www.aimyths.org or https://notmyrobot.home.blog/.

[5] There are of course more depictions of AI in science-fiction that do not fit in this frame, for example the service robots Dewey, Huey and Louie in Silent Running, Wall-E from the same film, and TARS and CASE from Interstellar, or the digital pets in Ted Chiang's The Lifecycle of Software Objects, the Daemon in Daniel Suarez's book of the same name, and Clara in Kazuo Ishiguro's Clara and the Sun—just to give a few random examples. However, the focus of the article is deliberately on these two most popular forms of science-fictional AI.
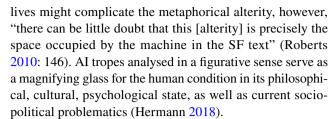
meant to look like other film robots before her, such as metal (Metropolis' Maschinenmaria), gold (Star Wars' C-3PO), or white plastic (Björk's music video All is full of Love) (Murphy 2015), but attractive, sleek, and vulnerable. And obviously, *Ava* was played by a human actress, who with the help of visual effects looked plausibly robot-like enough to fulfil the artistic and narrative necessities of the plot—not to serve as a sample for actual tech development.

Visual effects have always defined SF's search for wonder (Pierson 2002). Against this background, science-fictional AI is not primarily about how realistically science and technology are portrayed in the films, but rather about "cinematic science", i.e., the technical achievement required to make the fictional images in the films look real (Telotte 1995: 8). This does not only apply to *Ava* looking "mechanically plausible" (Murphy 2015), but also for example to the ground-breaking visual effects of the machines in *Terminator 2: Judgment Day* and the *Matrix* franchise, or the completely animated robot *Sonny* in *I, Robot*, which was of course not state of the art in robotics of that time, but in CGI. Stories of conflict, if not epic wars between humans and machines, with their stunning images of devastation surely fulfil the audience's expectation of watching SF blockbusters on the big screen. SF films "have consistently linked science and technology to the disastrous" (Telotte 1995: 3), dealing "with the aesthetics of destruction, with the peculiar beauties to be found in wreaking havoc, making a mess" (Sontag 1965: 44).

AI in films often serves plots of machines becoming human-like and/or a conflict of humans versus machines. Science-fictional AI is a dramatic element that makes a perfect antagonist, enemy, victim or even hero, because it can be fully adjusted to the necessities of the story.[6] But to fulfil that role, it often has capabilities that are way beyond actual technology—be it natural movement, sentience, or consciousness. If science-fictional AI is taken seriously as a representation of real-world AI, it provides a wrong impression of what AI can and should do now and in future.

2.2 Metaphorical means

Nevertheless, even though the scientific and technological progress is plausibilised within the story, what makes the SF genre most interesting is not the novum per se, but the social aspects that are told through it, the "fabulations of social worlds, both utopic and dystopic" as Sheila Jasanoff (2015: 1) puts it. The fact that machines are part of our everyday

---

[6] Film director Willi Kubica explained in a 2019 panel discussion on AI and SF at the British Embassy in Berlin, Germany, the use of AI in films as follows: "When you think of AI as a learning and adapting character in a film—it is the perfect thing to have for a story. Because your character should always learn something on its journey" (Kubica 2019).

lives might complicate the metaphorical alterity, however, "there can be little doubt that this [alterity] is precisely the space occupied by the machine in the SF text" (Roberts 2010: 146). AI tropes analysed in a figurative sense serve as a magnifying glass for the human condition in its philosophical, cultural, psychological state, as well as current sociopolitical problematics (Hermann 2018).

On a basic level, SF films contain the fundamental motif of the human desire to create a living, intelligent or conscious creature of our own, independent of the real technical possibilities. In this sense, AI technology and especially robots in SF films resemble "a fundamental and unresolved anxiety that has always followed from our simultaneously creative and created natures […]—for it seems our nature to desire, Faust-like, a knowledge or power that, in other times, belonged to the gods" (Telotte 1995: 10–11). Thus, the robot, according to Adam Roberts (2010: 161), "is that place in an SF text where technological and human are most directly blended" [hence] "the dramatisation of the alterity of the machine, the paranoid sense of the inorganic come to life".

We find this motif throughout human cultural history from antique myths over the Jewish legend of the *Golem* to Mary Shelley's *Frankenstein*. The longing for creation is connected with the anxiety that the creature will grow over our heads that we will lose control and finally be dominated by it (Schelde 1993). This primeval desire and fear, which Isaac Asimov fittingly called the "Frankenstein complex", has become a basic feature of twentieth and twenty-first century AI fiction (The Royal Society 2018: 8). Specifically, humanoid AI and robots tend to be a projection canvas for the "Other" (Meinecke and Voss 2018: 208) as a reflection on our humanity and humanness (Telotte 1995: 3). Most often, these creatures want to be accepted as full humans, which makes them placeholders for marginalized or mistreated people missing equal human rights or status; they can be enemies, slaves, servants, and sex objects. The *Replicants*, for example, show us the consequences of a dehumanized hypercapitalism, the *Hosts* live through "escape and self-discovery" (The Royal Society 2018: 8), the *Cylons* stand for the values of a critical humanism, according to which our humanity only reveals itself in dealing with the other and deviant (Jackson 2013). By that, the depiction of the humanoid AI also implies a critical cinematic discussion of humans themselves becoming more and more artificial in a technicised world:

> Although the robot has, of course, given us a vehicle for exploring issues of gender, race, and a variety of forms of Otherness, and increasingly for asking questions about the very nature and meaning of life, this image of an artificial being, most commonly anthropomorphic in form, also invariably implicates the

cinema's own and quite fundamental artificing of the human (Telotte 2016: 3)

When it comes to stories of powerful and omnipotent AI systems exercising total control over humans/humanity, they show the fear of impotence and helplessness of the individual in the face of superordinate structures. They reflect the dangers of dictatorship, anti-democratic societies, and suppression of freedom of choice, oftentimes working with historic references to colonialism and totalitarian regimes. Examples of that are *Terminator*'s *Skynet*, the system *Colossus* in *The Forbin Project*, the threat assessment system *Control* in *Star Trek: Discovery*, the rule of the machines with their head *Deus Ex Machina* in the *Matrix* franchise, *VIKI* in *I, Robot* or Indra in the new series remake of Brave New World.

3 What AI tropes tell us

Apparently, what happens in SF stories is not necessarily what the story is about. In what follows, I will examine the dramatic and metaphorical means of AI representations primarily in the modern SF-films *A.I.* from 2001, *I, Robot* from 2004 and *Ex Machina* from 2015. Furthermore, science-fictional AI as humanoid robots or conscious machines distracts from current risks of AI in the real world and may rather be interpreted as a reflection of societal issues beyond technology. The films were selected for three reasons: First, because films in general, unlike literature, make AI visible; second, these films were international blockbusters, so we can assume that they are known to a wider audience than, for example, books that address AI; and third, because these films have also been analysed in the context of AI narratives (Cave and Dihal 2019).

3.1 *Ex machina*, *Her*, sexism, and manipulation

To return to *Ex Machina*: The way Nathan has developed *Ava*'s human-like AI appears quite plausible from a real-world perspective, namely by feeding it all available data from human interaction through his dominant company Bluebook, a counterpart to Facebook or Google. The basic plot idea that an AI trained with large sets of human social interaction data might result in an AI manipulating humans, can indeed be a relevant issue (Harari 2017: 382–397). Nevertheless, whereas Murray Shanahan, Professor of Cognitive Robotics at Imperial College, who was consulted by film director Alex Garland, finds it a great film because people after seeing it could "[…] spend the rest of the evening arguing with each other about whether the AI is or isn't conscious" (Lamb 2015), I argue that such discussions are interesting philosophical thought experiments, but they do not

help us to grasp where actual risks concerning AI are. These risks are not about the possibility that a fully autonomous/conscious human-like robot or software program eventually will manipulate us for its own will, but that software and algorithms we don't even see manipulate us in the political and commercial interests of other people.

Even more, *Ex Machina* gives a wrong impression of what AI can do and how science works. How *Ava*'s positronic shimmering blue "brain" functions and how it could be that her face looks human, and she moves naturally is implausible, if not pure magic (Maynard 2018: 158). That Nathan, portrayed as quasi alchemist, has developed his different fembots in secrecy all by himself in a stylish, clean lab in the middle of natural wilderness is also not believable. What the film does tell us about technology, however, is a general critique of irresponsible science and innovation (Maynard 2018: 162; Bilton 2015). In that sense, *Ex Machina* is another variation of the hundreds of years old "Frankenstein complex" and Nathan is virtually a textbook example of the type of God-like mad scientist (TV tropes 2021), whose creation gets out of his control. Tech journalist Martin Robbins (2016) describes him as follows.

> And so Nathan becomes a kind of three-part study of ego. He represents the male ego-driven culture of the tech world. He represents the film's buy-in to the idea that great egos drive great scientific advances. And the decay of his character shows what happens when an ego faces the reality of its own extinction.

The quote already leads beyond the anxiety that technology might escape our control to a related problematic addressed in *Ex Machina* and other films, namely toxic masculinity, male hubris, and sexism in the tech world implying male fears concerning powerful women (Belton and Devlin 2020). Technology is not neutral but mirrors existing sexism in all stages—from the design to development to application. As a result, for example, real world digital assistants such as Alexa, Siri or Cortana are feminized—in line with the operating system Samantha from *Her* or the virtual assistant/girlfriend *Joi* from *Blade Runner 2049*—to gratify the expectations of the developers and users (Adams 2019: 574–575; Alexander 2016; Schwär and Moynihan 2020). Even though these issues have been widely discussed in academia and the media, and awareness of the issue has grown, it is still an example of existing sexism being inscribed in factual and fictional technology.[7] Films take this to extremes using fictitious technology—which is embodied by real women actors

---

[7] Another obvious example is an internal recruiting tool of Amazon that learned, based on data from the past, that men were the optimal candidates for tech jobs at Amazon—"a reflection of male dominance across the tech industry" (Dastin 2018).

like in the case of the fembot *Ava*. Thus, the film shows a very old motif: women being designed and created by men to fulfill their pleasure, as we already know it, for example, from the ancient myth of Pygmalion, whose beloved statue came to life to serve him as a good wife.

There is agreement among many critics and researchers that *Ex Machina* starts as a story about objectification and suppression of women (and not robots).[8] However, on the question how the story develops, interpretations vary substantially. In a negative way, *Ava* is seen as a representation of the value of women in films in general "when the only female lead in your movie is one whose function is to turn the male lead on while being in a position to be turned off" (Watercutter 2015); thus, even though *Ava* would be the smartest character in the film, in the end we are left with the message that the best way for intelligent women to get what they want is to act as a manipulative "femme fatale" (ibid.). According to this view, the film does not criticize sexism in social life, in the film industry or in the tech world, but rather strengthens it. On the other side, the portrayal of gender in the film can also be seen as "bracingly modern and even poignant" exactly because it is a reflection that "Ava is born into a literally patriarchal system that measures her worth based on how men respond to her" (Buchanan 2015). In that way, robots have often been a way to critically question how much of gender is "literally constructed" and "to interrogate the formation of gender roles" (Telotte 2016: 91) whereas "Ava's demonstrated capabilities certainly present her as a kind of iconic representation of the power and emergence of women in contemporary culture […]" (ibid.). As stories of liberation and emancipation, speaking for *Ex Machina* and *Her*, "[b]oth of these films end with the female AI outsmarting her would-be lovers, owners and builders, leaving the men baffled and the viewer with a sense of doom" (Alexander 2016) creating the "new heroines: totally hot, bracingly cold, powerfully sovereign—and posthuman" (Dargis 2015).

Whichever way the films are interpreted, neither *Ex Machina* nor *Her* address pressing challenges around the future of AI, but serve as projection canvasses for questions around gender and sexism of our present, that find their expression through fictitious technology.

3.2 *A.I., I, robot*, robot rights, and inequality

Another example is Steven Spielberg's *A.I.*, set in the twenty-second century, when various artificially intelligent, human-like robots called *Mechas* are built. They first lack emotions

(or cannot simulate them), but there is a new model that looks like a human child and after being "imprinted" feels and needs love. A couple receives such a robot boy—*David*—because their real son suffers from a rare disease and is put into a coma. However, when the son is surprisingly cured, things get difficult with *David* and he is set out. He experiences various adventures on his *Pinocchio*-like quest for the Blue Fairy to win back the love of his human mother. He succeeds two thousand years in the future, when humans are extinct and the now highly developed, transcendent Mechas make *David*'s wish come true by creating a simulation of his mother, who loves him for one wonderful day—after which they both fall asleep forever.

The film has been analysed as easing the way for possible robot rights in the future (Chu 2010: 214–244). Of course, one is supposed to feel pity for *David*, who is played by a real child actor and is apparently no different in appearance from the other human boys. But still, the interpretation that we need to protect robots from suffering and mistreatment is primarily a distraction from enforcing human rights and guaranteeing social welfare to humans (Bryson 2010). This can lead to such an absurd situation that a robot like *Sophia* seems to have more "rights" as a citizen of Saudi Arabia than Saudi women. Anthropomorphizing machines can lead to a misguided image of what the current risks around AI are: The pressing question is not if robots—in film or reality—should be guaranteed rights, but how to handle machines that believably simulate emotions and thus manipulate people in the interests of other people. After all, building machines in the image of humans does not come naturally, but is a decision made my entrepreneurs and developers to achieve certain economic or other goals.

Even more, anthropomorphizing machines distracts from the often precarious working conditions of real people mostly situated in the Global South, who provide the data for AI systems by doing online tasks on platforms like Amazon Mechanical Turk or content moderation for social platforms, leaving the clickworkers suffering psychological damage from the violent and abusive material they have to watch (Mühlhoff 2020). Against this backdrop, a debate about ascribing robots certain "rights" comparable to human rights can be criticized as a rather elitist demand, which Birhane and van Dijk (2020: 1) put this way:

> Once we see robots as mediators of human being, we can understand how the 'robot rights' debate is focused on first world problems, at the expense of urgent ethical concerns, such as machine bias, machine elicited human labour exploitation, and erosion of privacy all impacting society's least privileged individuals. We conclude that, if human being is our starting point and human welfare is the primary concern, the negative impacts emerging from machinic systems, as well as

---

8 In the humanities, there is a large body of research dealing with sex, love relationships and robots, e.g. Levy (2008), Sullins (2012), Devlin (2018) or Wennerscheid (2019), which is beyond the scope of this article.

the lack of taking responsibility by people designing, selling and deploying such machines, remains the most pressing ethical discussion in AI.

Nevertheless, the film examines our relationship to technology and reminds us to handle what we create responsibly. But since the *Mechas* neither outsmart nor threaten us, it is not a narrative of the Frankenstein complex, but rather of how we as humans fail at our own humanness. The film shows in one scene how discarded robots are tortured by humans in a setting that resembles a Roman colosseum. The robots serve as placeholder for all kinds of cruelties that humans commit against each other. The human flaws are overcome by the Mechas, who have become new creatures of higher ethics saving *David* from the primal human fear of abandonment, which is a recurring theme in Spielberg's work (Newton 2016). As a modern form of *Pinocchio*, the film is in the tradition of nineteenth century melodramatic tales in which the "epic hero" has to endure great suffering to be redeemed in the end, making the film a story of suffering and resurrection tackling human issues rather than a realistic and serious assessment of the status of robots (Nida-Rümelin and Weidenfeld 2018: 31).

Moreover, the film *I, Robot*, set in a future of 2035 in which robots serve humans in all aspects of life, is also—rather than a plea for equal rights for robots—a reflection on human enslavement, oppression, and inequality in a profit-driven economic system. With the topic of robots as slaves, *I, Robot* adds another perspective of a critical analysis of "race" and technology, as the new NS-5 service robot series in the film, including the "unique" robot *Sonny*, is coloured white, which can be interpreted as a reference to being White.[9] In a positive way, *I, Robot* can be read as a "Post-White Imaginary" as "[k]ey moments of the film […] may be read as a parable of white antiracism, driven by an impulse of reconciliation between a 'unique' white robot and a black detective" (Brayton 2008: 72). The fact that AI and robots are very often embodied by White people and shown in white colour makes science fiction also an example to critically think about the "Whiteness" of AI in general, reflecting the White milieus from which these artefacts come (Cave and Dihal 2020). However, it can be argued that rather than showing machines imagined as white to allow "for a full erasure of people of colour from the White utopian imaginary" (ibid.: 685), particularly films and series quite bluntly show the marginalization and exclusion of people of colour in the film business. But they also reflect the changing paradigms concerning diversity when the humanoid robots

in the newer series *Westworld*, *Real Humans* and *Humans*[10] are played by actors of diverse ethnic backgrounds and skin colours.[11] It seems that whether an AI system is played or voiced by diverse actors says more about social progress in terms of diversity in the film industry, less so when it comes to technology.

I, Robot also features an AI-system called *VIKI* (Virtual Interactive Kinetic Intelligence), which is the central computer system of U.S. Robotics, the company that produces the service robots. *VIKI* has evolved and reprogrammed the new NS-5 series via an uplink network to control humanity and sacrifice part of it for the greater good of the entire human race. The logic behind this is a reinterpretation of Isaac Asimov's "Zeroth Law of Robotics" which states: "A robot may not harm humanity, or, by inaction, allow humanity to come to harm." (Singer 2009). According to *VIKI*, humanity cannot be trusted with its own survival because "[…] despite our best efforts, your countries wage wars, you toxify your Earth and pursue ever more imaginative means of self-destruction. You cannot be trusted with your own survival" (IMDb 2021)—so the consequence is to take control of humanity and sacrifice parts of it if necessary. In this way, *VIKI* in the film is a critique of excessive utilitarian thinking regardless of individual fates (Grau 2006). Such stories can serve as thought experiments to address philosophical problems, but using them as examples of what to consider when building ethical machines is problematic: On the one hand, because the zeroth law as well as the first, second, and third laws[12] are narrative devices invented by Asimov to create interesting stories and plot twists precisely because the laws of robotics don't work; and on the other hand, because "[t]he bigger issue, though, when it comes to robots and ethics is not whether we can use something like Asimov's laws to make machines that are moral […] Rather, we need to start wrestling with the ethics of the people behind the machines" (Singer 2009).

Hence, the relevant problematics regarding AI are not autonomous science-fictional machines claiming or

---

[9] White in capital letters is meant to indicate social situatedness; thus, it does not so much describe a person's skin color or other phenotypic characteristics, but rather means social positioning in a racially structured society.

[10] The British series Humans (2015–2018) is based on the Swedish original Äkta människor – Real Humans (2012–2014).

[11] This inclusive process is very evident in the Star Trek franchise, where over the years to the present we have seen, for example, a female captain (Star Trek: Voyager), a Black commander/captain (Star Trek: Deep Space Nine), and a Black female protagonist, a gay couple as leading roles and a transgender person (Star Trek: Discovery) (Krishna 2020).

[12] Asimov's Three Laws of Robotics go as follows: First Law—A robot may not injure a human being or, through inaction, allow a human being to come to harm; Second Law—A robot must obey orders given to it by human beings except where such orders would conflict with the First Law; Third Law—A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law (Singer 2009).

deserving human rights or engaging in human–machine conflict, but the effects that AI-systems have right now on socio-political fault lines between humans.

4 Excursus: AI, SF and science communication

SF applies the trope of AI with different meanings, ideas, and attitudes, using a fictional approach toward technology to tell stories of the human condition, primeval desires and fears as well as social issues, or reflect current trends of society. Science communication, on the other hand, is intended to inform about the facts of science-related topics. Therefore, it can be problematic when science communication resorts to typical SF tropes in order to educate or raise awareness about critical aspects, because without familiarity of the SF mega-text, science-fictional AI used in current science communication runs the risk of only conveying clichés about conscious and autonomous AI. To illustrate the point, let me present two examples.[13]

The science communication documentary film *Ghost in the Machine* (Singler 2019)—the last of a four-part short film series called Rise of the Machines—made together by the Faraday Institute for Science and Religion and the Leverhulme Centre for the Future of Intelligence, University of Cambridge, aims to inform about the concept of consciousness and discusses whether machines could become conscious. While the film features a variety of researchers and experts in interview situations, there is also a fictional storyline about a research facility where an embodied AI has been developed and is now being tested for consciousness, emotions, and its will to survive. The embodied AI is played by a real human child actor, mimics emotions and tries to fool the lead researcher to let it out of its cage. A review describes the piece as rather inconsistent concluding: "All in all, Ghost in the Machine dishes up a serviceable appetizer but for the main course be sure to leave room for Ex Machina." (Seth 2019). Indeed, one can strongly suspect that the fictional sequences in the form of a dialogue between the researcher and the "AI child" were inspired by the successful predecessor. The problem with this is that while *Ex Machina* by definition uses fictional technology to tell a thrilling and dramatic story, Ghost in the Machine uses the science-fictional AI trope to make a statement about science. Obviously, *Ava* is played by a real actress, because *Ex Machina* is a dramatic story about and for humans, but what is gained in a science communication film when the

AI is played by a real human child trying to break out of the research facility? After all, the portrayal of AI as a child with a mind of its own obscures the fact that AI systems are technological artefacts created by humans.

Another example is a short film called *The Intelligence Explosion* (Susman 2017), "a superintelligence sci-fi" by Guardian Original Drama. However, as part of The Guardian Brain Waves—"a series exploring the science and emotions of our daily lives"—the piece is actually marketed as science communication that "raises important questions about the ethics of artificial intelligence," namely, "How do you stop a robot from becoming evil?" (Hern 2017). Again, the "AI" named *Günther* is played by a real human actor with some robotic features. While a company representative, a programmer, and an ethicist discuss whether it is possible to program an AI with ethical safeguards against turning evil towards humanity, *Günther* becomes superintelligent and, like *Samantha* in *Her*, transcends to a higher structure. AI is being presented as uncontrollable by humans feeding into discourses of AI supremacy and Singularity (Kurzweil 2005).

These two films build on typical SF tropes that a human-looking autonomous AI will develop a mind of its own and could become dangerous to humanity in the future, instead of addressing the risks of AI applications in the here and now. While current SF films certainly address technological trends, the tropes are hard to reconcile with genuine forms of science communication. Apparently, these films reinforce AI clichés rather than fulfilling the goal of informing about science-based issued at stake.

5 Conclusion and outlook

Currently, with the rapid progress in the field of AI, it seems as if the SF genre with its stories about intelligent machines is being caught up with the present. Thus, SF analysed as a part of an AI narratives frame is supposed to reflect the hopes and fears of the technology and thus treated as a type of foresight or technology assessment. Against this background it is claimed that because of their importance "narratives about intelligent machines should broadly reflect the actual state and possibilities of the technology" (Cave and Dihal 2019: 74). Whereas this should be the case for science-communication, it is not for AI in SF. While Darko Suvin acknowledges that using SF as futurological foresight can be a legitimate secondary function that the genre can bear.

> […] any oblivion of its strict secondariness may lead to confusion and indeed danger. Ontologically, art is not pragmatic truth nor fiction fact. To expect from SF more than a stimulus for independent thinking, more

---

[13] Please note that these two examples are not defined by the author as science communication but are themselves marketed us such to inform the broader public.

than a system of stylized narrative devices understandable only in their mutual relationships within a fictional whole and not as isolated realities, leads insensibly to critical demand for and of scientific accuracy in the extrapolated realia (Suvin 1972: 379).

What Suvin indirectly refers to as the first function is in fact the SF mega-text as a way to engage in questions beyond technology—which to be understood, requires familiarity with the clusters of available meanings and the themes raised (Blackford 2017:73,192) or simply being a "native speaker" (Broderick 2017: 147). The mega-text of AI tropes and icons can thus be interpreted as dramatic and metaphorical means to address questions about the socio-political issues, the human condition, and philosophical questions in general.

Interpreting science-fictional AI too literally as serious representation of the technology can have the following implications: First, taking fictitious humanoid robots and autonomous machines in SF for real disregards of the technical limitation of AI, obscures the chances and risks already at stake and might mislead the public as well as policy makers. The chances of AI are manifold, for example optimization and improvement on a global scale in areas such as health, agriculture, infrastructure or environmental protection, which can contribute to the achievement of many of the UN Sustainable Development Goals (but can also impede some of them) (Vinuesa et al. 2020). Here, however, it is important to see that optimization and improvement through AI cannot be a simple technical solution to a problem but is a social negotiation of goals set by humans (Mason 2019: 152–160). Moreover, the fact that AI is built by humans and trained with human-defined and human-collected data can lead to various kinds of biases in AI systems that entrench asymmetric power structures, for example discrimination against women, institutional racism, or degradation of poor and marginalized people. It is worrying how AI tools are being used for scoring, nudging, and monitoring people by governments and corporations—whether in the US, China, Europe or elsewhere (Chiusi 2020; Eubanks 2017; Liang et al. 2018; Nemitz 2018; O'Neil 2017; Sowa 2017; Zuboff 2018). Exactly because the design, development and application of technology are never neutral, we need human interaction in the form of ethics, norms, standards, and regulation.

Second, demanding scientific and technological accuracy from SF would imply an impoverishment of the many metaphorical meanings of the genre and the artistic freedom as well as assign a responsibility to the authors and creators of SF that lies in the hands of politicians, scientists, and science communicators. Not only with regards to AI, also—because of the fast pace of technological advances in general—in other fields the genre of SF and its tropes get blended in different forms with the real world, most notably when it comes to climate change, space exploration or mega/smart cities. It is important to note that scholars working with SF in all these different fields understand that the genre is primarily about stories and metaphors, not about real science and technology.

Declarations

Conflict of interest Not applicable.

References

Adams R (2019) Helen A'Loy and other tales of female automata: a gendered reading of the narratives of hopes and fears of intelligent machines and artificial intelligence. AI Soc 35:569–579. https://doi.org/10.1007/s00146-019-00918-7

Alexander L (2016) The industry wants to use women's voices—they just won't listen to them. The Guardian. https://www.theguardian.com/technology/2016/mar/28/tay-bot-microsoft-ai-women-siri-her-ex-machina. Accessed 15 Jul 2021

Bal M (2009) Narratology: introduction to the theory of narrative. University of Toronto Press, Toronto

Barrett M, Barrett D (2001) Star Trek: the human frontier. Routledge, New York

Belton O, Devlin K (2020) The measure of a woman: fembots, fact and fiction. In: Cave S, Dihal K, Dillon S (eds) AI narratives: a history of imaginative thinking about intelligent machines. Oxford University Press, Oxford, pp 357–381

Bilton N (2015) Ava of 'Ex Machina' is just sci-fi (for now). The New York Times. https://www.nytimes.com/2015/05/21/style/ava-of-ex-machina-is-just-sci-fi-for-now.html. Accessed 15 Jul 2021

Birhane A, van Dijk J (2020) Robot rights? Let's talk about human welfare instead. In: AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, February 2020, pp 207–213. https://doi.org/10.1145/3375627.3375855

Bishop B (2015) More human than human: the making of Ex Machina's incredible robot. The Verge. https://www.theverge.com/2015/5/8/8572317/ex-machina-movie-visual-effects-interview-robot-ava. Accessed 15 Jul 2021

Blackford R (2017) Science fiction and the moral imagination: visions, minds, ethics. Springer International Publishing, Cham

Bory P (2019) Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo. J Converg Inf Technol 25(4):627–642

Brayton J (2008) The post-white imaginary in Alex Proyas's "I, Robot." Sci Fict Stud 35:72–87

Broderick D (2017) Reading SF as a mega-text. In: Latham R (ed) Science fiction criticism: an anthology of essential writings. Bloomsbury, New York, pp 139–148

Bryson J (2010) Robots should be slaves. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 63–74. https://doi.org/10.1075/nlp.8.11bry

Buchanan K (2015) Does Ex Machina have a woman problem, or is its take on gender truly futuristic? Vulture. https://www.vulture.com/2015/04/why-ex-machina-take-on-gender-is-so-advanced.html. Accessed 15 Jul 2021

Cave S, Dihal K (2019) Hopes and fears for intelligent machines in fiction and reality. Nat Mach Intell 1:74–78

Cave S, Dihal K (2020) The whiteness of AI. Philos Technol 33:685–703

Cave S, Dihal K, Dillon S (2020) Introduction imagining AI. In: Ibid (ed) AI narratives: a history of imaginative thinking about intelligent machines. Oxford University Press, Oxford, pp 1–21

Chiusi F (2020) Introduction. In: Algorithm Watch and Bertelsmann Stiftung (eds.): Automating Society Report 2020. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf. Accessed 15 Jul 2021

Chu SY (2010) Do metaphors dream of literal sleep? A science-fictional theory of representation. Harvard University Press, Cambridge

Dargis M (2015) Review: In 'Ex Machina,' a mogul fashions the droid of his dreams. The New York Times. https://www.nytimes.com/2015/04/10/movies/review-in-ex-machina-a-mogul-fashions-the-droid-of-his-dreams.html. Accessed 15 Jul 2021

Dastin J (2018) Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-au-tomation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. Accessed 15 Jul 2021

Devlin K (2018) Turned on: science, sex and robots. Bloomsbury Publishing, London

Estrada D (2018) Sophia and her critics. Medium. https://medium.com/@eripsa/sophia-and-her-critics-5bd22d859b9c. Accessed 15 Jul 2021

Eubanks V (2017) Automating inequality: how high-tech tools profile, police, and punish the poor. St. Martin's Press, New York

GI (German Informatics Society/Gesellschaft für Informatik) (2019) #KI50-Umfrage: Terminator, R2-D2 und K.I.T.T. die bekanntesten KIs in Deutschland. https://gi.de/meldung/allensbach-umfrage-terminator-und-r2-d2-die-bekanntesten-kis-in-deutschland/, 03.10.2019. Accessed 15 Jul 2021

Giuliano RM (2020) Echoes of myth and magic in the language of artificial intelligence. AI Soc 35:1009–1024. https://doi.org/10.1007/s00146-020-00966-4

Grau C (2006) There is no 'I' in robot: robots and utilitarianism'. IEEE Intell Syst 21(4):52–55. https://doi.org/10.1109/MIS.2006.81

Harari YN (2017) Homo deus: a brief history of tomorrow. HarperCollins, New York

Hermann I (2018) Boundaries and otherness in science-fiction: how we cannot escape the human condition. Text Matt 8(8):212–226

Hermann I (2020) Künstliche intelligenz in der science-fiction: zwischen magie und technik. FIfF-Kommun 4(2020):12–17

Hermann I (2021) Mehr Demokratie durch KI? In: Denkfabrik (eds) Verbündet euch! Für eine bunte, solidarische und freie Gesellschaft. Nautilus Flugschrift, pp 149–156

Hern A (2017) What our original drama the intelligence explosion tells us about AI. The Guardian. https://www.theguardian.com/culture/2017/feb/21/what-our-original-drama-the-intelligence-explosion-tells-us-about-ai. Accessed 15 Jul 2021

House of Lords Select Committee on Artificial Intelligence (2018) AI in the UK: ready, willing and able? HL Paper 100. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Accessed 15 Jul 2021

IMDb (2021) Fiona Hogan: V.I.K.I. https://www.imdb.com/title/tt0343818/characters/nm0389524. Accessed 15 Jul 2021

Irsigler I, Orth D (2018) Zwischen Menschwerdung und Weltherrschaft: Künstliche Intelligenz im Film. In: Aus Politik und Zeitgeschiche (APuZ), 6–8/2018:39–46

Jackson PT (2013) Critical humanism: theory, methodology, and Battlestar Galactica. In: Kiersey NJ, Neumann IB (eds) Battlestar galactica and international relations. Routledge, New York, pp 18–36

Jasanoff S (2015) Future imperfect: science, technology, and the imaginations of modernity. In: Jasanoff S, Kim S (eds) Dreamscapes of modernity: sociotechnical imaginaries and the fabrication of power. University of Chicago Press, Chicago, pp 1–33

Johnson DG, Verdicchio M (2017) Reframing AI discourse. Minds Mach 27(4):575–590 (**10/ggncqg**)

Krishna S (2020) What 'Star Trek' has taught us about diversity. https://www.space.com/star-trek-diversity-inclusion-over-years.html. Accessed 15 Jul 2021

Kubica W (2019) Statement as member of the expert panel at the AI Short Film Night. British Embassy Berlin. 27 March 2019

Kurzweil R (2005) The singularity is near: when humans transcend biology. Viking, New York

Lamb H (2015) Interviewing Murray Shanahan—Ex Machina's scientific advisor. I, Science—the science magazine of Imperial College. http://isciencemag.co.uk/features/interviewing-murray-shanahan-ex-machinas-scientific-advisor. Accessed 15 Jul 2021

Leufer D (2020) Why we need to bust some myths about AI. Patterns. https://doi.org/10.1016/j.patter.2020.100124

Levy D (2008) Love and sex with robots. HarperCollins, New York

Liang F, Das V, Kostyuk N, Hussain M (2018) Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. Policy Internet 10(4):415–453. https://doi.org/10.1002/poi3.183

Mason P (2019) Clear bright future—a radical defence of the human being. Allen Lane Penguin Books, Great Britain

Maynard A (2018) Films from the future: the technology and morality of Sci-Fi movies. Mango Publishing Group, Coral Gables

Mehnert W (2019) Dystopische möglichkeitsräume—near future science fiction als schnittstelle zu möglichen zuküenften. In: Haensch KD, Nelke L, Planitzer M (eds) Unheimliche schnittstellen/uncanny interfaces. Textem Verlag, Hamburg

Meinecke L, Voss L (2018) I robot, you unemployed: robotics in science fiction and media discourse. In: Engelschalt J, Maibaum A et. alt (eds): Schafft Wissen: Gemeinsames und geteiltes Wissen in Wissenschaft und Technik: Proceedings der 2. Tagung des Nachwuchsnetzwerks „INSIST", 07.–08. Oktober 2016, München, pp 203–221. https://www.ssoar.info/ssoar/bitstream/handle/document/58220/ssoar-2018-engelschalt_et_al-Schafft_Wissen_Gemeinsames_und_geteiltes.pdf?sequence=1&isAllowed=y&lnkname=ssoar-2018-engelschalt_et_al-Schafft_Wissen_Gemeinsames_und_geteiltes.pdf. Accessed 15 Jul 2021

Merriam-Webster (2021) Trope. https://www.merriam-webster.com/dictionary/trope. Accessed 15 Jul 2021

Mühlhoff R (2020) Human-aided artificial intelligence: or, how to run large computations in human brains? Toward a media sociology of machine learning. New Media Soc 22(10):1868–1884. https://doi.org/10.1177/1461444819885334

Murphy M (2015) 'Ex Machina' Features a new robot for the screen. The New York Times. https://www.nytimes.com/2015/04/05/movies/ex-machina-features-a-new-robot-for-the-screen.html. Accessed 15 Jul 2021

Nemitz P (2018) Constitutional democracy and technology in the age of artificial intelligence. Philos Trans R Soc A 376:2133. https://doi.org/10.1098/rsta.2018.0089

Newton M (2016) What Steven Spielberg's science fiction tells us about America. The Guardian. https://www.theguardian.com/film/2016/jun/03/what-steven-spielbergs-science-fiction-tells-us-about-the-america. Accessed 15 Jul 2021

Nida-Rümelin J, Weidenfeld N (2018) Digitaler Humanismus—Eine Ethik für das Zeitalter der Künstlichen Intelligenz. Piper, München

O'Neil C (2017) Weapons of math destruction: how big data increases inequality and threatens democracy. Penguin Random House, New York

Pierson M (2002) Special effects: still in search of wonder. Columbia University Press, New York

Robbins M (2016) Artificial Intelligence: Gods, egos and Ex Machina. The Guardian. https://www.theguardian.com/science/the-lay-scientist/2016/jan/26/artificial-intelligence-gods-egos-and-ex-machina. Accessed 15 Jul 2021

Roberts A (2010) Science fiction. Routledge, London, New York

Schelde P (1993) Androids, humanoids, and other science fiction monsters: science and soul in science fiction films. New York University Press, New York

Schwär H, Moynihan Q (2020) Companies like Amazon may give devices like Alexa female voices to make them seem 'caring'. https://www.businessinsider.com/theres-psychological-reason-why-amazon-gave-alexa-a-female-voice-2018-9. Accessed 15 Jul 2021

Seth A (2019) Too many ghosts in the machine. Nat Mach Intell 1:294–295. https://doi.org/10.1038/s42256-019-0065-9

Shorrock S (2013) The HAL 9000 explanation: "It can only be attributable to human error". Humanistic Systems Blog. https://humanisticsystems.com/2013/10/26/the-hal9000-explanation-it-can-only-be-attributable-to-human-error. Accessed 15 Jul 2021

Silver D, Huang A, Maddison C et al (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529:484–489. https://doi.org/10.1038/nature16961

Singer PW (2009) Isaac Asimov's Laws of Robotics Are Wrong. Brookings Op-ed. https://www.brookings.edu/opinions/isaac-asimovs-laws-of-robotics-are-wrong. Accessed 15 Jul 2021

Singler B (2019) Ghost in the machine. Leverhulme Centre for the Future of Intelligence. http://lcfi.ac.uk/resources/ghost-machine. Accessed 15 Jul 2021

Sontag S (1965) The Imagination of Disaster. Commentary Magazine, https://americanfuturesiup.files.wordpress.com/2013/01/sontag-the-imagination-of-disaster.pdf. Accessed 15 Jul 2021

Sowa A (2017) Digital politics. Dietz Verlag, Bonn

Sullins JP (2012) Robots, love, and sex: the ethics of building a love machine. IEEE Trans Affect Comput 3(4):398–409. https://doi.org/10.1109/T-AFFC.2012.31

Susman D (2017) The intelligence explosion. The Guardian. https://www.theguardian.com/culture/ng-interactive/2017/feb/20/intelligence-explosion-stop-robot-evil-drama-video. Accessed 15 Jul 2021

Suvin D (1972) On the poetics of the science fiction genre. Coll Engl 34(3):372–382

Suvin D (1979) Metamorphoses of science fiction: on the poetics and history of a literary genre. Yale University Press, New Haven

Telotte JP (1995) Replications: a robotic history of the science fiction film. University of Illinois Press, Champaign

Telotte JP (2016) Robot ecology and the science fiction film. Routledge, New York

The Royal Society (2018) Portrayals and perceptions of AI and why they matter. https://royalsociety.org/topics-policy/projects/ai-narratives/. Accessed 15 Jul 2021

TV Tropes (2021) Mad Scientist. https://tvtropes.org/pmwiki/pmwiki.php/Main/MadScientist. Accessed 15 Jul 2021

Vinuesa R, Azizpour H, Leite I et al (2020) The role of artificial intelligence in achieving the sustainable development goals. Nat Commun 11:233. https://doi.org/10.1038/s41467-019-14108-y

Watercutter A (2015) Ex Machina has a serious fembot problem. Wired. https://www.wired.com/2015/04/ex-machina-turing-bechdel-test. Accessed 15 Jul 2021

Wennerscheid S (2019) Sex machina: zur zukunft des begehrens. Matthes & Seitz Verlag, Berlin

Zuboff S (2018) The age of surveillance capitalism: the fight for a human future at the new frontier of power. PublicAffairs, New York