What can science fiction tell us about the future of artificial intelligence policy?

Andrew Dana Hudson¹ · Ed Finn¹ · Ruth Wylie²

Received: 25 February 2021 / Accepted: 24 August 2021 / Published online: 20 September 2021
© The Author(s) 2021

Abstract

This paper addresses the gap between familiar popular narratives describing Artificial Intelligence (AI), such as the trope of the killer robot, and the realistic near-future implications of machine intelligence and automation for technology policy and society. The authors conducted a series of interviews with technologists, science fiction writers, and other experts, as well as a workshop, to identify a set of key themes relevant to the near future of AI. In parallel, they led the analysis of almost 100 recent works of science fiction stories with AI themes to develop a preliminary taxonomy of AI in science fiction. These activities informed the commissioning of six original works of science fiction and non-fiction response essays on the themes of “intelligence” and “justice” that were published as part of the *Slate* Future Tense Fiction series in 2019 and 2020. Our findings indicate that artificial intelligence remains deeply ambiguous both in the policy and cultural contexts: we struggle to define the boundaries and the agency of machine intelligence, and consequently find it difficult to govern or interact with such systems. However, our findings also suggest more productive avenues of inquiry and framing that could foster both better policy and better narratives around AI.

Keywords Artificial intelligence · Technology policy · Science fiction · Futures · Policy futures

1 Introduction

When we think of artificial intelligence, often we think first not of a product or app we can download, but of a character in a book or film. Science fiction has been imagining our current technological moment—or one near at hand—for at least a century, when Karel Čapek coined the term “robot” from a Czech word meaning “forced labor” in the 1921 play *Rossum’s Universal Robots*. The foundations of this AI imaginary stretch back another century to Mary Shelley’s founding of science fiction with *Frankenstein*, a

story about the perils of creating artificial life. As classicist Adriene Mayor argues (2018), we could even consider the Greek myths of Hephaestus’ golden servant automatons, the golems of Jewish folklore, and the automatons of Hindu mythology as proto-science fictions about technological life. Perhaps more than any other technology, artificial intelligence has been entangled with science fiction and mythologies of technology from the beginning.

Already we have mentioned many angles from which to look at this imaginary, including artificial intelligence, artificial life, technological life, robots, and automatons. We could add more, from both science and fiction: algorithms, agents, neural networks, machine learning, and Turing tests; narrow, general, and super AI; a host of technical acronyms. Which of these apply to, say, Arnold Schwarzenegger in *The Terminator*, and which apply to the Siri function on your iPhone? For AI in science fiction and the AI of our present-day technology industry, definitions prove ambiguous, both in terms of how these systems are technically bound through their design and implementation, and in terms of how everyday users interpret them and interact with them (Finn 2018). Because we invest these technological entities with their own agency, the boundaries of their existence

✉ Ed Finn
edfinn@asu.edu
Andrew Dana Hudson
andrewdhudson@gmail.com
Ruth Wylie
ruth.wylie@asu.edu

¹ Center for Science and the Imagination, Arizona State University, Tempe, AZ, USA

² Center for Science and the Imagination, Mary Lou Fulton Teachers College, Arizona State University, Tempe, AZ, USA

inevitably reflect our own assumptions about and definitions of personhood, will, and intelligence back to us.

The stories we tell about AI have foreshadowed and heralded the emergence of these technologies by years, sometimes by decades. Alan Turing's thought experiments of the Turing Machine and the Turing Test, the ethical robots of Isaac Asimov's imagination, and the early robotic prototypes and rhetoric of the cybernetics movement in the 1940s and 1950s could all be framed as technically grounded, speculative stories about AI. This history, and the attendant ambiguity and entanglement about what AI actually is, puts science fiction creatives in a special relationship with AI technology policy. The stories they tell about AI exert significant influence on how AI actually develops and is understood, which in turn plays a major role in determining how AI is governed and regulated.

Despite, or perhaps because of this rich mythic history, we tend to rely on just a few pieces of narrative shorthand in popular discourse about AI. There are the killer robots, remorseless and powerful machines like those in the *Terminator* stories. Equally destructive but far more deceptive, there are the mimic machines that “pass” as human, like the seductive androids of *Battlestar Galactica*, *Ex Machina*, and *Westworld*. There are childlike, Pinocchio-esque characters trying to learn how to develop an inner humanity, as in *Wall-E*, *Chappie*, or *Short Circuit*. And there are the inscrutable, oracular god-computers, like Deep Thought from Douglas Adams' *The Hitchhiker's Guide to the Galaxy* (1979), charged with finding the answer to life, the universe, and everything. When we turn to these stories to understand what AI means, we gaze at distant horizons and ignore all of the humble and mundane ways in which machine intelligence is already transforming our lives, our economies, and our brains, from aircraft autopilot systems and credit scores to social media filter algorithms.

This state of affairs spurred us to launch AI Policy Futures, a research and public engagement project addressing the challenge of how storytelling can enhance policy deliberations and public dialog about how we define, regulate, and assess artificial intelligence technologies. Working in collaboration between the Center for Science and the Imagination (CSI) at Arizona State University and the Open Technology Institute at New America, we spent 18 months studying the intersection of science fiction and AI policy through a combination of quantitative, qualitative, and creative methods.¹ Our research goal was to explore the full

spectrum of AI narratives in science fiction, to identify ideas that might have been overlooked or underestimated with respect to the near-future emergence of AI technologies, and to create a taxonomy of different configurations of possible AI futures. These themes and the taxonomy served as a guide for commissioning new works of science fiction and fostering a grounded dialog between the technology policy and science fiction communities, using compelling stories of the near future as a form of speculative anticipatory governance (Guston 2013).

We were not the first researchers to consider these questions. On the Sci-fi Interfaces blog, Chris Noessel, a member of our project's advisory board, pursued a similar line of inquiry in his “Untold AI” project, which focused on AI as depicted on screens (2019a). Noessel analyzed the messages and moral frames that dozens of television series and movies implied about AI. He compared these findings with guidelines drawn from a variety of white papers and manifestos about AI promulgated by technology-industry advocacy groups and think tanks. He found that on-screen science fiction and policy recommendations from tech share many ideas about AI. In other words, the science fiction that Noessel looked at was somewhat aligned with the concerns of nonfictional AI's various stakeholders. However, Noessel also found stories being told about AI in sci-fi that had no parallel in the messages coming from tech. These he labeled “pure fantasy”—ideas that made for good stories, but did not seem useful in analyzing the real-world present and future of AI. This left a final group: imperatives and messages from the tech industry's manifestos that science fiction was not engaging with. Noessel called these “untold AI” stories, and he recommended that science fiction creatives could help us better understand real-world artificial intelligence by telling stories about these particular ideas.

The sci-fi AI imaginary has also been elucidated by research at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge. In a paper published by *Nature Machine Intelligence*, Kanta Dihal (another member of our advisory board) and her colleagues laid out four pairs of linked hopes and fears for the future of AI (Cave and Dihal 2019). The four dichotomies were: immortality and inhumanity, ease and obsolescence, gratification and alienation, dominance and uprising. The Leverhulme Centre has also been advancing a set of Global AI initiatives, identifying a multitude of cultural frames through which different populations interpret and govern AI.

Both of these projects identified the challenges of mapping the AI imaginary and provided valuable framing for our own taxonomical approach. This prior work also highlighted our collective tendency to focus on the long-term destiny of the technological project of machine intelligence, and thereby lose track of the social, economic, and political impacts of AI as it exists already. Our own effort attempted

¹ In 2019, roughly a year into the project timeline, project co-leader Kevin Bankston left his position at the Open Technology Institute for a new role at Facebook. He continues to be involved in this project and as a fellow at the Center for Science and the Imagination, but has not contributed directly to the research since this transition, though he provided some editorial input on this publication.

to focus squarely on this near-term future. We engaged technologists, policy experts, science fiction writers, and researchers in an effort to see and think critically about the AI imaginary.

2 Methodology

We used a mixed-methods approach to address our research goals of exploring the full spectrum of AI narratives in science fiction, to identify ideas that might have been overlooked or underestimated with respect to the near-future emergence of AI technologies, and to create a taxonomy of different configurations of possible AI futures. We began with a series of interviews with experts (technologists, policy thinkers, science fiction writers, and researchers), followed by a workshop with further experts and a public gathering in May 2019. For the purposes of this paper, we use the data collected via the semi-structured interviews ($n = 13$) conducted between November 2018 and March 2019. Interviews averaged 60 min ($SD = 20$ min) and were conducted in-person or via video conference or phone call. Interviews were recorded and transcribed. Semi-structured interviews allow researchers to ask predetermined questions around a topic while also allowing for follow-up questions and prompts that may vary between participants, to more fully explore a topic (Leech 2002). For our work, we asked all participants the same high-level questions about AI and pursued individual themes and observations in a free-flowing way.

Following the convening, a team of three researchers coded 96 science fiction short stories to better understand the spectrum of ways in which AI is depicted in science fiction. Details regarding our iterative process are described below (See “[Developing the taxonomies](#)”). These conversations and structured readings of science fiction helped to inform a set of themes that are under-explored in both science fiction and technology policy and are likely to become pressing in the near future of actual technology implementation, which we discuss in the closing section of this paper.

3 Defining AI

Many of our experts agreed that the protean nature of artificial intelligence is a barrier to effective policy conversations and broader cultural understanding of how to anticipate the potential impacts of these technologies. “Defining AI is a moving target,” argues Damien Williams, an AI ethics scholar at Virginia Tech. There has never been a consensus definition of what qualifies and what doesn’t, and the goalposts have moved in one direction or another with a constant push and pull. Suren Jayasuriya, an AI researcher

at Arizona State University, argues that this is a sign of a healthy diversity in the field. “It’s good that we can’t come up with one definition of AI,” Jayasuriya told us. “You don’t want a policy that applies uniformly over a wide swath of technology.”

Others argued that definitions and sub-definitions are more about hyping up particular products in a competitive investment space. As science fiction writer and information technologist Brenda Cooper put it, “machine learning is just really good computation.” Where and how computation becomes artificial intelligence remains a blurry rhetorical line. The impulse that drives countless startups to gravitate towards the term AI as a descriptor for their work, whether or not any form of machine learning or intelligence is involved, has its inverse in academia, where careers are built on drawing distinctions and rejecting prior nomenclature. “People keep renaming things to get grant money,” said cyberpunk guru Bruce Sterling. “But this interferes with their ability to do actual research.”

“Intelligence” is also definitionally slippery. Several of the thinkers we engaged—such as Cooper, science fiction writer and physicist Vandana Singh, and science fiction author, futurist, and AI Policy Futures advisory board member Madeline Ashby—pointed out that moral and political questions of AI “personhood” are complicated by the existence of nonhuman intelligences that we already share the world with: animals. Scientists studying animal behavior tell us that many species have memories, thoughts, opinions, emotions, preferences, problem-solving skills, and other cognitive capacities presumed to be unique to humans, while also sensing and engaging with the world in deeply unhuman ways (Wasserman 2006; Reznikova 2007). Other research suggests that some plants and trees also sense and behave in ways that could be called “intelligent” (Simard et al. 1997; Wohlleben 2017). But our legal systems do not accord personhood to plants or nonhuman animals. If we expect to create advanced AIs that have some human cognitive characteristics but are in fundamental ways not human, and we also expect debates about whether those AIs are legal persons, will this lead to expanded political consideration for nonhuman “natural intelligences”? When does the moral standing of computational systems bump up against the social and intellectual convention that treats other living beings as property, products, or pests? Thinking about plant and animal cognition can give us tools to better recognize and evaluate intelligence in our machines.

Another frame raised several times in our interviews was that of neurological othering. AI entities in science fiction are often othered by being depicted as incapable of understanding some aspect of the human psyche: love, for instance, or the importance of family. In *Star Trek: The Next Generation*, the android Data struggles to make sense of human emotions and humor. “Think of every replicant as

autistic,” as interviewee Damien Williams said. Thinkers in the field of Disability Studies can help us see that such narratives of othered minds derive from experiences and perceptions of neurodivergent people, those who have suffered brain injuries, and so on. As with animal and plant cognition, expanding our definitions through disability and neurodiversity lenses allows us to question whether “general artificial intelligence” (i.e., AI that is equivalent to a human being) will ever be attainable or even conceptually useful. Instead, we should use these frames to theorize about the multiplicity and diversity of intelligence in a future filled with many varied kinds of “narrow artificial intelligence.” Perhaps they will even push us to recognize the ambiguity and arbitrary normativism we apply to cognition in society.

These questions underscore how “intelligence” is itself the wrong way to think of and imagine complicated information technology. Sterling argues in a talk at the 2014 IMPAKT Festival that cognition and computation are different, and that definitions of AI that mix the two constitute “a category error”: “It’s like thinking that birds have wings and drones have wings, and therefore one day drones will lay eggs and birds will have radar.” The Turing Test, according to Sterling, has long been used in both science fiction and the technology industry as an end-run around this ontological problem. Rather than define intelligence, Alan Turing cleverly invented an emulation test, suggesting that any computer sophisticated enough to pass as a human should be considered intelligent. This conception of intelligence as both a performance and a deception has long troubled our conceptions of AI. In both cases, the nuances of modern-day technology policy require new, more honest discussions and metaphors about what exactly technologists are creating. But of course, these conversations continue to take place in the cultural context of a mythos whose deeply entrenched notions of anthropomorphism, analogy, and familiarity shade all discourse around AI and machine learning.

Intelligence is ambiguous in large part because we connect it so often to different constructions of human identity: technical mastery of a skill or medium; interaction and empathy; anticipation and foresight; idiosyncrasy and the auto-poetic performance of the self. From this perspective, it is tempting to dismiss the “personhood” question as a philosophical fantasy that does not tell us anything useful about the actual techno-culture AI companies are creating. However, Ashby argues that all policy regimes are built on philosophical foundations and supported or maintained in part on philosophical grounds. The abolition of slavery was a personhood argument, for instance. Debates about the personhood of fetuses and zygotes are at the heart of some of the most contentious rifts in American political culture. So whether or not the sci-fi and tech industry imaginaries about AI use what Sterling would call “bad metaphors,” they are nevertheless inescapable, and we should be prepared to

contend with abstract questions about artificial intelligence as a central aspect of policy discussions.

4 Developing the taxonomies

Defining AI is not just hard in the real world—it is hard in sci-fi as well. We found similar ambiguity in our attempts to categorize AI systems depicted in prominent science fiction. Our initial goal in our qualitative content analysis (Bengtsson 2016) was to analyze a broad swath of narratives, but we eventually decided to narrow our focus to more rigorously approach the question of just how science fiction is imagining artificial intelligence. We chose to focus on short fiction, rather than include movies, television shows, novels, video games, and other media. We had three reasons for this decision. First, we believed that the short fiction form was one with few hard rules or material production constraints (unlike film budgets or the computational limits of game engines), and thus empowers creatives to approach topics in an incredibly diverse range of ways. Second, another of the outputs of our project would be a set of original commissioned short stories to run in the Future Tense Fiction series in *Slate*, so we hoped a consistent focus on the short story would improve our outputs there. Third, we felt that we would not be able to rigorously examine the enormous body of work about AI that has been produced across all media, and even an attempt at random sampling would be more a demonstration of our own interests and preoccupations in genre and media terms than a dispassionate survey. Adopting this constraint increases the representative power of our sample.

We also chose to limit our analysis to short stories published in the twenty-first century. As we have discussed in this paper, artificial intelligence has been a subject of science fiction for decades, even centuries. This was again an effort to limit the amount of content we would need to examine, but also to focus on fiction produced during a time when most of the population was engaging regularly with computers and algorithms. Were our familiar ideas about killer robots, et cetera, an artifact of science fiction from the mid-twentieth century, or had they found fresh purchase in our contemporary imaginary? How did the proliferation of digital technologies shape science fiction notions and expectations about AI? We believe these questions are central to an examination of the near future of AI policy, and is best addressed by keeping our focus on contemporary, as opposed to classic, stories.

To compile our corpus, we looked at the set of short stories ($n=975$) that had received some form of recognition in the genre, either by being nominated for one of the three major science fiction awards (the Hugos, the Nebulas, or the Locus Awards) or that had been included in the annual

Year's Best Science Fiction anthologies edited by the late Gardner Dozois. After removing duplicates of stories earning more than one of these honors, we further limited our pool to stories that were freely available online. Sourcing texts in this way, rather than simply listing the AI science fiction works we had already encountered and read, allowed us to limit the degree to which our own preconceptions and interests shaped our analysis. However, it was also a significant task to pore over almost 20 years of award nominations and 19 volumes of *Year's Best* collections to determine which stories actually include and discuss AI. The presence of robots or talking computers is only occasionally mentioned in a short story's title, and so we had to look deeper. Some stories we could include based on a keyword search or quick scan of the first few pages. Others required detailed scans or even full, close reading to determine whether the story featured AI. To complete this step, one researcher reviewed all of the stories and identified ones that featured AI to be included in the next stage of coding. This resulted in a corpus of 96 stories.

We found ambiguity within this first step of the process. We quickly realized we had a natural bias toward recognizing AI in *characters* but not in *systems*. If a story features a robot or talking computer, especially one with a name or a prominent emotional role in the story, it was easy to include it in our corpus. If a story instead features computational systems that make complex decisions but do not talk, it was harder to decide whether or not that story should be included. The same was true with stories that imply technology running in the background that we think of as AI products in our contemporary reality (facial recognition, autonomous vehicles, etc.). Indeed, when such stories made it into our corpus, in step 2, our research assistants would sometimes flag them as being erroneously included, or ask if we had reason to think particular characters in the story are secretly robots. What those of us compiling the corpus interpreted as AI, those of us reading and categorizing the stories did not always see. In part, this is the nature of short stories, which have less room for extended explanations and detailed descriptions of technological minutiae (and narrative in general—technical explanations can interrupt and distract readers following a plot). In part, however, we believe this is due to the ambiguous AI imaginary.

Rather than adopting a purely emergent coding strategy, we first identified a number of themes based on advice from our advisory board and experts we interviewed, extended discussion among our project team, and trial and error. These themes reflect our hypotheses and analysis goals. For example, we hoped to categorize stories based on the sophistication and purpose of the AIs they depict; their engagement with policy questions; the positive, negative, or ambiguous tone of the AI imaginaries they propose; and other questions (see final codebook in [Appendix 2](#)). An

early debate on the project team was around whether we should attempt to categorize AI in stories as protagonists or antagonists. While this might help determine just how prevalent movie-based tropes are, we ultimately decided that this framework was too reductive and would mean losing a great deal of literary nuance in many stories. We also discussed trying to categorize AI around embodiment—whether they walk in an android body, are projected as a hologram, speak as a disembodied voice on a computer, control swarms of nanobots, and so on. The range of possibilities here, many of which overlap, and the disconnect between these portrayals and the current policy issues we were interested in, led us to omit this question from the final analysis as well (cf. Noessel 2019b). Similarly, we ultimately eliminated questions about who or what creates and controls the AI, which proved difficult to determine in many of the short-form narratives we were analyzing.

Once we developed a first codebook, we went through several more iterations by having four readers (including members of the project team) test the frameworks on small batches of 5–10 stories. Afterwards we would debrief with our coders to determine where questions made sense, where discussion was needed to establish shared definitions, and where the framework failed to make sense or produced contradictory interpretations. We eliminated some questions that did not seem clear or useful, and adjusted the wording. In one case, we split a question that seemed to be asking too much into two separate parts. After these initial trials, we also instituted a confidence metric, which we asked coders to include in their explanation boxes for each question. With each rating, our coders would mark whether they felt “highly confident,” “moderately confident,” or “not confident” in their rating, as well as indicate a second-choice answer. We also had coders briefly explain their answer to each question.

After finalizing a coding structure and set of questions, each story in our corpus was read and analyzed by two coders who had not been involved in developing the previous iterations of the codebook. Where there was disagreement about a rating, a third researcher would read the story and break the tie. We found a great deal of disagreement in these ratings by all three coders. We were applying difficult questions to literary works that are deliberately nuanced, even ambiguous, in their themes, their descriptions, and their imagined futures. These disagreements would be problematic if our goal was to create a generalized taxonomy using a small set of sample data; in this case, however, we were attempting to understand the characteristics of a data set (contemporary, widely read, short-form science fiction) using an extensive sample of that data. As we discuss below, the ambiguity was an important feature of the data set, revealed through our efforts at taxonomizing it. This highlighted how challenging it is to meaningfully discuss AI in a nontechnical way, both because the definitions are slippery,

as discussed above, and because AI encompasses so many different kinds of technologies, both in science fiction and in the real world. Despite these challenges, we believe that our analysis offers useful insights into the current state of the policy/sci-fi intersection that suggest a path forward for the AI imaginary.

5 Results and implications for technology policy

Throughout the project, our most significant finding has been the slipperiness of AI as a concept itself, both in fiction and policy. Because we struggle to articulate what AI *is*, in terms of its functional boundaries, its status as a person or entity, and its position of agency and responsibility within broader social systems, it is difficult to determine what AI should or should not do.

The study resulted in three main findings. First, that narrow AI will have a greater social impact in the near term, but sci-fi short stories mostly concern general AI. Second, most stories do concern themselves with policy, governance, or constraints, but mostly bias. Third, most of the stories depict problematic AI rather than AI that works well. The paper discusses each of these in more detail below.

Of the stories we analyzed, a majority (52%) feature AI that we categorized as having “general” (human-equivalent) intelligence. Examples include “Articles of Faith” by Mike Resnick, in which a church’s robot janitor decides to embrace religion, and “Fandom for Robots” by Vina Jie-Min Prasad, about a museum-dwelling AI obsessed with an anime cartoon show. Another significant portion of stories—such as “Computer Virus” by Nancy Kress (2001)—describe “super” intelligences with cognitive capacities vastly greater than human beings, sometimes likened to gods. Few stories feature the “narrow” artificial intelligences that are extremely good at some tasks (such as identifying skin cancer from photos) but unable to reason through any other kind of problem—in other words, the kind of AI that many of the experts we talked to predicted will dominate the field for a long time. Examples of those that do depict narrow AI include “Elephant on Table” by Bruce Sterling (2017), featuring complex but inhuman surveillance algorithms, and “Henry James, This One’s For You” by Jack McDevitt (2005), about a computer program that can generate the great American novel.

As most of the science fiction that we encountered focused on AI that were of human or beyond-human intelligence, most of the AI in those stories are imagined as beings with personalities, motives, and agency—or, at least two out of those three. In other words, these are AI *characters* with technological origins, not technologies featuring AI systems. The sci-fi AI imaginary is still preoccupied with the basic

conundrums raised by Shelley in *Frankenstein*: the possibilities, responsibilities, and hazards of creating a fully formed being. As science fiction author Lee Konstantinaou put it during a panel at our May 2019 convening, “most SF about AI isn’t about AI at all, not in a forecasting way—it’s about us, our needs and insecurities.”

The vast majority (89%) of stories we looked at discuss AI policy, governance, or constraints—a gratifying finding, given the framing of our project. In “Mika Model” by Paolo Bacigalupi (2016), for instance, ambiguity around governance of AI is the fulcrum on which the plot turns: is a sex robot that kills its abusive owner a murderer or a malfunctioning consumer product? Several stories, such as “I, Row-Boat” by Cory Doctorow (2006) and “Cat Pictures Please” by Naomi Kritzer (2015), make reference to Asimov’s Three Laws—the classic example of AI governance. Others, such as Hannu Rajaniemi’s “His Master’s Voice” (2008), imagine governance as mainly concerned with controlling the ability of AIs to copy and reproduce themselves. Yet others, such as Charles Stross’s “Lobsters” (2001), discuss the question of AI citizenship and rights. However, we quickly found that, just as defining AI had proved tricky, so too did our raters struggle to apply a clear and rigorous definition of *policy*. The concept easily sprawled. In AI, it is difficult to distinguish between questions of governance and questions of technology design, and in the constrained creative space of a short story, that ambiguity is magnified by the narrative lens of the author’s interests and intentions.

For instance, we might speak of governing AI technologies with government institutions that set and enforce laws, such as standards for how autonomous vehicles should be programmed or bans on facial recognition technology, as some activists are now calling for. Perhaps governance might even be represented in science fiction via the Turing Police in William Gibson’s *Neuromancer* (1984) or the Krishna Cops from Ian McDonald’s *River of Gods* (2004)—agents that actively hunt down wayward AIs. On the other hand, we might think of governance as taking place primarily in the code of AI products. This might be explicit, such as the “Three Laws” that govern Asimov’s robots. It might also be more subtle: an AI character that speaks of how it thinks or processes information, the kinds of problems it can solve and those it cannot. Is governance in the code of the algorithms that Facebook uses to feed users content, or in the act of Congress calling Mark Zuckerberg to testify about his company’s effects on American democracy? What portion of that spectrum should be thought of as “policy”? This question complicated our ability to identify policy insights in science fiction stories.

We also found that 72% of the stories we examined depict AI technology either as hazardous or as having unintended consequences. This may stem from stories’ need for drama and conflict; stories where everything works perfectly are

less likely to capture our interest, get published, or win awards. But partly this is the nature of the technology itself, whose power magnifies flaws in data and human institutions. As Williams pointed out in an interview, wish-granting djinn are good analogies for AI in many stories: we frame our requests poorly and algorithms give us too literally what we wish for. “It’s not the genie’s fault you didn’t understand the nature of your desire,” Williams said.

However, our interviews and discussions with both science fiction thinkers and AI policy experts revealed a wide range of vital policy concerns sometimes overlooked in the more fantastical imaginary of robot uprisings.

Perhaps the most commonly discussed policy challenge posed by AI technology is bias. AI systems, such as facial recognition or predictive policing, are often built on datasets developed through already-biased practices. If racist policing practices lead to more arrests in low-income and non-white neighborhoods, an algorithm predicting crime or criminality in that neighborhood will make that racial bias a part of how it sees the world. Similar, less egregious examples proliferate throughout the AI tech space, from algorithms used for hiring and processing insurance claims to biased language used in natural language processing (NLP) algorithms. Aylin Caliskan, Joanna Bryson, and Arvind Narayanan conducted a study that found that NLP systems were likely to assume that a “doctor” was male and a “nurse” was female. Systems connected masculine names to concepts like “career,” “professional,” and “salary,” and feminine names to words like “wedding” and “parents” (2017). These are sexist human assumptions that have crept into our machines, and preventing or eliminating such bias is a tricky and time-intensive task. Williams, along with many other experts, argues that a key way to address these problems is to promote greater diversity at technology companies and on teams that develop AI systems.

Complicated systems entrench and amplify existing biases, make them harder to address, and fuel the tendency for using AI to avoid human responsibility and culpability. As Ashby argued, the transfer of decision-making from humans to machines absolves governments and corporations of liability and accountability for mistakes, injustices, and even war crimes. They make it easier for institutions to say, “Sorry you got hit by a self-driving vehicle, but there’s no driver to charge with reckless endangerment.” “Sorry, we can’t approve your insurance claim. The algo makes those decisions.” “Sorry the automated predator drone blew up your village. It was a glitch that no human will be held accountable for.” From redlining to policing to war, AI systems, even ones that only nominally make decisions, make it harder for humans to get justice against institutions and the individuals that run them. As these examples suggest, the problems are not technical but rather sociotechnical—the intersection of technologies, policies, and people.

AI systems also pose difficult policy questions around transparency and consent. Do convincingly human automated systems—for instance, a voice agent that might call the salon for you to book a haircut—need to make humans aware that they are interacting with a machine? Do we need to establish consent before our AI systems reach into people’s lives? In an interview, ASU theatre scholar Michael Rohd articulated a rubric for thinking about how we interact with AI either knowingly or unknowingly, and willingly or unwillingly. An AI therapy app we deliberately download would be an example of a system we interact with both knowingly and willingly. A streaming service that non-transparently feeds us algorithmically generated content might be considered AI we interact with unknowingly but willingly. An AI prison warden would interact with prisoners knowingly but unwillingly. And a top-secret AI system used for government surveillance would interact with citizens unknowingly and unwillingly. Because we struggle to define what AI “is”—the capabilities and boundaries of a particular system in sociotechnical space—it is difficult to provide full and knowledgeable consent in every case.

The final area of policy concern that emerged in our discussions is the question of assessment. As ACLU technology policy director Kade Crockford explained to us, AI (and digital technologies in general) are often rolled out to the public or into institutions without a full assessment of their social impacts. Once they are in use, these genies are hard to get back into the bottle. Police procedures get tied into databases, employees get trained on new systems, human workers are laid off and replaced by automated tools. Particularly in the criminal justice system, invasive technologies can be in use for years before a case makes it to the Supreme Court for final judgment as to whether they violate Constitutional rights. Crockford argues that we should have discussions about whether AI systems are ethical and socially beneficial *before* they are developed and brought to market or put in use. The protean nature and fluidity of these systems only exacerbate the ambiguity and assessment problems: even experts who believe that they understand the risks and capabilities of a problematic system like Amazon’s Rekognition facial recognition technology might discover that the tools, and their use cases in implementation, have changed overnight.

As pressing as these areas of concern are, they are often too mundane to be grist for popular science fiction stories. These policy questions were largely nonexistent in the dozens of stories we analyzed—with the caveat that unintended consequences, broadly conceived, are a regular theme of many stories about AI.

6 Commissioning original fiction

Our findings suggest that there is a disconnect between the problems of imaginary AI as envisioned by science fiction and the policy problems posed by AI technology products. Much pop science fiction about AI features killer robots, robot uprisings, and computational creatures rebelling against their creators. Interestingly, our own study of AI in short fiction found few stories that use these tropes (though a few did, such as Paolo Bacigalupi's story "Mika Model"). Only a few stories we read feature AI in violent or military roles, or AI designed to provide humans with companionship. Most of these fictional AIs manage complex systems, like cities, corporations, or starships—or pursue their own ends. This may be the difference between AI on the screen and AI on the page; a chrome robot that blows things up à la *The Terminator* or a sexy robot that seduces à la *Ex Machina* are more visually interesting than a disembodied AI making stock trades.

Yet those pop narratives have shaped public and industry discourse around AI. Prominent commentators on the future, such as Stephen Hawking and Elon Musk, have warned of an AI takeover as a consequence of tech advancement. Such narratives come with a sense of inevitability that does not leave room for public choices about technology. An imaginary built around time-traveling killer robots occludes the messier quandaries of real machine intelligence as it is being implemented today, such as privacy, consent, and bias. Our collective fixation on the anthropomorphic destiny of AI also makes it hard for the public to recognize the real promise of AI technologies to do good. We need a more nuanced conversation that explores a broader range of possible AI morphologies and cultural roles, and how such systems might positively impact society.

Our project sought to address this challenge by commissioning original fiction to fill in some of the gaps in our cultural representations of AI. Between the lines of inquiry pursued by our advisory board members and our discussions with experts at our May 2019 convening, we developed a sense of what kinds of stories about AI would be most useful for better engaging the public with contemporary tech policy questions. This informed our approach to commissioning original short stories that were published as part of *Slate's* Future Tense Fiction series. First, we asked that stories be grounded in an understanding of AI as a technology product likely created within a capitalist economic framework. We were less interested in AIs that spontaneously dragged themselves out of the primordial ooze of the Internet or that found consciousness through some accident or twist of fate. We did not want killer robots, all-powerful Skynet analogs, or AI

girlfriends. We did want stories that extrapolate from already-contentious policy debates and explore the nuance and ambiguity of intelligent machines as they are now, or might soon become.

The first three stories published in *Slate* offer compelling examples of this way forward for the AI imaginary. In "Affordances" (2019), bestselling science fiction author, activist and AI Policy Futures advisory board member Cory Doctorow wrote about how predictive policing can reinforce and be used to justify already existing societal biases, and also depicted the massive amounts of hidden human labor required to make technologies like facial recognition function seamlessly. In "A Priest, a Rabbi, and a Robot Walk Into a Bar" (2019), speculative fiction writer and AI Policy Futures graduate researcher Andrew Dana Hudson explored a future tech industry building advanced AI voice agents that can navigate diverse cultural and religious sensibilities, and imagined how AI might be used by faith groups to advance their evangelical agendas. Finally, advisory board member and Hugo Award finalist Malka Older offered "Actually, Naneen" (2019), a story about robots used for childcare work and the complications that come with treating an integral part of the family like a product with planned obsolescence.

In these glimpses of AI policy futures, AI is sometimes not a thing that thinks, but a thing that *sees*. Sterling has made the point that the most popular uses of AI are photo and video editing effects, such as giving yourself cat ears in Snapchat or TikTok. We see this lens in Doctorow's story, in which a popular facial recognition service is used to let people into their locked homes—and sometimes witnesses doorstep crimes. The computational labor involved is sourced in part from climate refugees, who work to categorize images the facial recognition algorithm does not understand. We see in this detail another way forward for the AI imaginary: acknowledgement that AI systems require a vast amount of human labor to function. Another example is Hudson's story about the debates that might take place in the tech company offices where the sentences spoken by future-Siris are written, curated, edited, and approved.

And this imaginary sees AI policy questions replicating, amplifying, and reigniting values disputes across almost every sector. Hudson's story imagines the cultural work of AI as reaffirming our beliefs and mores as well as answering our questions, and in affirming those beliefs, contributing to the imagined communities from which they spring. This also includes understanding that AI will play a role in areas of life that may seem, at first glance, less technological—such as childcare and childhood, as Older explores and Hudson and Doctorow touch on. Taken together, the stories describe not just how AIs might "see" cultural phenomena, from childrearing to politeness, but also how they are seen, as servants, family members, or agents of oppression.

Our second trio of AI stories expanded on this notion, focusing on the question of justice. Holli Mintzer's story "Legal Salvage" (2020) addresses the question of legal personhood for AIs as an intersection of law, ethics, and aesthetics, with a sentient robot that asserts its personhood in part by demonstrating taste and curatorial competence. Tochi Onyebuchi's story "How to Pay Reparations: a Documentary" (2020) takes on a different dimension of justice, asking if algorithms could be employed to counteract systemic racism and right historical wrongs through economic interventions. In "The State Machine" by Yudhanjaya Wijeratne (2020), machines have taken surveillance and personalization to their logical extreme: an autonomous AI governance system that continually updates the constitution and legal framework, while also micromanaging the lives of its citizens.

All three of these justice stories address the challenge of instrumentalizing moral and ethical rules in code. As we invest computational intelligence with real political, legal, and economic power, we effectively endorse structures of value and ethical practice embedded in those systems. Sometimes the machines are a convenient moral dumping ground for the difficult decisions that humans no longer wish to be blamed for. Sometimes the emergence of AI forces us to reconsider broader questions of personhood and justice. But each of the stories asserts a bedrock faith that justice itself cannot be automated, and that individual actions and moral positions matter. Mintzer's story of an AI who achieves legal personhood dwells on the central importance of individual actions and identity, as well as the deep relationship between aesthetic and moral value. The central role of political economy in each of these three stories invites possible readings of AI and moral value, in particular the question of how work, identity, and civic participation intertwine. These authors' sketches of AI and personhood echo Adam Smith's *Theory of Moral Sentiments* and the concept of society as a construct built on responsibility and care towards strangers.

One of the stories from our coding exercise that we found most representative of the broader AI imaginary implied by both the corpus we analyzed and the broader pop culture narratives about AI is "Zima Blue" by Alastair Reynolds (2005), which was adapted into an animated short for the Netflix show *Love, Death & Robots*. The story centers on Zima, a futuristic artist working on planet-sized canvases who is obsessed with a particular shade of blue. The story describes his process of artistic transcendence, growing beyond the limitations of his body to commune with the cosmos in extreme environments. In the end, Zima is revealed to be an AI, an android who evolved to his present form upgrade-by-upgrade from a simple, Roomba-like robot that tirelessly scrubbed the blue tile of a swimming pool.

"Zima Blue" elegantly articulates perhaps the most mesmerizing aspect of today's AI imaginary: that our

still-primitive machines might one day evolve into galaxy brains that make us look primitive in turn. Our analysis found that most stories take place in the far future and focus on general and super-intelligent AIs. There is a great interest in science fiction around the eventual destiny of technology, and such grand visions certainly make for fascinating stories, because they redirect our thoughts to the eternal questions of human experience and the purpose of existence. But sometimes this focus can amount to missing the trees for the forest.

It is time for science fiction to leave behind the tropes of both the killer robot and the computer god. These concepts stretched our brains to vast possibilities throughout the twentieth century, but now in the twenty-first they confuse our attempts to grapple with the real technological sea change already underway around us. Focusing on the far-off destiny of AI distracts us from the political and policy issues posed by these new technologies today. It is time to grow past the fixation on artificial *people*, and think more deeply about what it means to make *systems* more intelligent.

7 Futures

The project has provided promising initial findings to validate our approach: our taxonomy revealed that the imaginary of AI in science fiction is more complex and nuanced than the stories that tend to be most widely disseminated through major films and bestsellers. Our conversations with experts demonstrated that there is a pressing need for new conversations and narratives about AI that focus on the near term and on the complex, amorphous sociotechnical dynamics of machine learning and artificial intelligence as it is already being deployed. Finally, the fiction and response essays we commissioned have been well received, demonstrating that it is both feasible and worthwhile to treat this imaginary as a design space, or better yet an ongoing dialog, where policy experts, scholars, writers, and others can grapple with the stakes of technical and ethical possibilities. The impact of this fiction is difficult to measure in meaningful quantitative terms, but it is real: within weeks of publication, these stories have led to speaking invitations, new assignments in academic reading groups and syllabi, and new conversations about technology policy. Based on these reactions, we speculate that there is both an opportunity and a need for what we might term "policy fiction" that responds to Sheila Jasanoff's call for works that "situate technologies within the integrated material, moral, and social landscapes that science fiction offers up in such abundance" (2015).

There have long been bridges between the imaginaries of science fiction and technology policy, but deliberately combining perspectives and methods from both of these worlds could lead to richer and more nuanced policy deliberations

that are also more accessible and engaging to the public. Looking forward, we hope both to advance this work in the field of AI and to continue developing the methodologies piloted here to span science fiction and technology policy in other arenas. Other rapidly evolving fields like synthetic biology are haunted by their own short lists of ghost stories and nightmare scenarios, and would benefit from grounded explorations of the near future.

Such extensions would also allow us to better understand what is unique about the protean nature of AI in human culture, and what is true of any potentially transformative technology. Humanity's relationship with AI is usually a variation of the Mechanical Turk: a black box that seems to function as an independent, thinking machine, but which in fact obscures the labor and agency of many human beings. The ambiguity of AI as a bundle of technologies, cultural roles, and accumulated mythological baggage underlines the fact that we will not be disentangling the humans from the machines anytime soon. But this presents us with a paradox of form: code and policy are both relatively rigid systems for expressing human intentions. They are arduous and slow to create, difficult to maintain, and often fail to keep up with reality. One approach to this problem is to see code and policy as structures of knowledge that are then exposed to shifting cultures of interpretation. Code and law are like boats we launch into the river of culture: the water continues to shift and flow, reconfiguring these objects and changing their functions as the waterways evolve. Some of these tools survive almost unchanged over long periods, like the doctrine of *habeas corpus* or elements of the UNIX operating system. Others remain functional but accrue new interfaces and analogies like so many barnacles, such as MOCAS, the COBOL-based contract management system the Pentagon has been running since 1958. Still other policy and code structures collapse in time, of course, disappearing with nary a bubble after a few short years—an occurrence so frequent no example is needed. In the final category are protean, ship of Theseus-style entities, still afloat but hardly resembling their original form, like the leviathan expansions of Facebook or Google since their inceptions.

In every case, however, the cultural frame in which these systems operate is fluid. Employees might peer “at” the Pentagon’s MOCAS now through sleek HTML interfaces instead of chunky lines of cathode-illuminated text. The cultural frame of what military contracting means has shifted significantly since 1958. But the code is still there, meaning different things to us today even as it completes the same operations it always has. This fluidity of cultural perception offers an important insight for how we manage the policy and technical development of AI. Like the Mechanical Turk, all intelligent systems *perform* their intelligence and their agency, whether they are not really intelligent at all, like Microsoft’s Clippy assistant, or if they represent the

current state of the art. We are still developing interpretive frameworks to competently manage, collaborate with, and integrate intelligent machines within broader sociotechnical systems.

We launched this project in part as a response to the poverty of imagination reflected in dominant AI narratives: the killer robot, the omniscient oracle, the android who becomes human. The mythological and intellectual taproots of AI, from *Pygmalion* and *Frankenstein* to the Turing Test, all imagine AI as another self, an anthropomorphic Pinocchio figure we can use as a way to reflect on ourselves. We keep looking for the humanity in our machines, but the results of our work here suggest that this may be blinding us to many different kinds of agency. Because of this anthropocentric bias, we also risk miscategorizing or underestimating the power and agency of the intelligent systems that already influence us, from autopilots and algorithmic news filters to financial systems sifting resumes and loan applications.

There are other interpretive frames we can use for nonhuman intelligence and agency, as several of our experts pointed out. Animal and biological metaphors can be particularly useful in modeling the intention and purpose of intelligent machines without presupposing their humanity. Perhaps one day we will understand autonomous vehicles to be less like KITT from *Knight Rider* and more like horses: creatures that have their own needs and desires but limited cognitive capacity for higher-level thought or emotion. “The Voluntary State,” by Christopher Rowe (2004), features just this sort of future car. One might object that this kind of analogy may lead back to the “birds as drones” category error Sterling articulates about intelligence, but in this case biological metaphors can provide a richer contextual frame. Understanding some computational systems as agents or entities with particular capacities and goals is more culturally tractable when we draw on our long cognitive relationships with other species to frame those capacities and goals.

This represents an inversion of the history of computational research, which maintains our ancient fascination with remaking the natural through human artifice. From Vaucanson’s duck and Norbert Wiener’s cybernetic moth to the McCulloch-Pitts neuron and the brain metaphors prevalent in machine learning today, computer science borrows from nature to make creatures of silicon that generally seek to impersonate the organic, the human. Reversing the flow of those analogies, we might create new cultural roles and grammars of action for AI systems that perform their work as familiar but nonhuman agents.

In other instances, our best analogy to a complex intelligent system might be ecological rather than biological. Indigenous peoples around the world maintain knowledge and observational systems based on humanity’s long history of working in awareness of and in concert with local ecologies. While we have grown used to interacting with a

computational avatar, voice, or personal interface, those surface appearances are usually deceptive, obscuring systems that are as extensive, intricate, and nonhuman as a forest or a river delta. Analogizing the behaviors of complex, interdependent algorithmic systems to a waterfall or a flock of birds might give us new ways to visualize the agency of these systems and mentally model our interactions with them.

Adopting better metaphors for AI might ultimately shape not just cultural interpretations but also the creation of new systems. AI researchers have in recent years begun confronting the “hard problems” of social and cultural data, which are deeply steeped in the messiness, contradictions, and ambiguities of humans. If we move away from designing AI to model or impersonate human intelligence or agency, we might be able to create meaningful new constructs. Several of our Future Tense Fiction stories explore the ways in which AI systems can embody or animate ideas. Tochi Onyebuchi’s story imagines an algorithm to enact reparations. Malka Older’s robo-nannies perform not just the service labor but the affect and love inherent in childcare. In another recent Future Tense story on the theme of AI (not commissioned for this project), author Karl Schroeder (2020) imagines AIs that literally assume the roles of biological and ecological entities, like rivers and protected territories. Today we often build AIs to imitate us in particular tasks or roles. But we can go beyond AIs that do human work, or even AIs that do the work of pretending to be human. There is an opportunity to build AIs that enact our most humane ideas, speaking for the voiceless, the forgotten, and the oppressed, inverting the power structure of futures like Doctorow’s “Affordances.” Hudson’s story “A Rabbi, a Priest and a Robot Walk into a Bar” runs with the fact that AIs embed cultural and religious values, not as tacit by-products of their development but as core features. While Hudson’s narrative imagines a future where AI becomes another battleground in culture wars, we could also design systems to remind us of those people and values we are quick to ignore. Every AI system already embeds philosophies, models of action, or structures of belief, but if we intentionally designed them to embody and speak those truths, they could occupy different cultural roles.

These observations are contingent on the preliminary nature of our work. The cultural space of AI is vast and rapidly increasing, as we embed AI technologies into more objects, structures, and organizations. The imagination space of AI in science fiction is also large, and our taxonomy barely scratches the surface of contemporary AI narratives in English across genres from film to video games, much less those in other languages and time periods. Nevertheless, the challenges confronting policy-makers, technologists, and indeed everyone who interacts with intelligent machines are all live issues right now. The next few years will be critical to the regulatory and cultural codes and laws we adopt for AI.

While there are many open questions here ripe for continued research, it is clear that there is still room for better stories about what AI is, and what it should become.

Appendix 1: Questions used in expert interviews

- What science fiction works would you consider the most important for understanding sci-fi narratives about AI? What’s the canon and why? Any that are less well known?
- How should we define AI? What’s the operating definition we should use in the next 10–20 years in the context of tech policy and social change?
- What do you see as the most realistic or important short to mid-term social impact of AI?
- What are the most important things policymakers need to understand about AI? What questions should they be asking about AI?
- We are crafting a taxonomy of depictions of AI in science fiction. What categories or questions do you think are the most useful lenses for examining AI in this fictional context?
- AI narratives often engage with metaphysical questions about the nature of consciousness, the definition of intelligence, the consequences of immortality, etc. How important are these questions to consider when thinking about real-world technology policy on AI?
- What are we missing here? What questions about AI do you think have yet to be explored, by sci-fi or other research? What potential narratives would you like to see?
- Who else should we talk to about this?

Appendix 2: Questions used to code short fiction works

- Title of the work
- Author
- Link/Collection Title
- Short Description of the AI System or Ecosystem (including specific name if applicable)
- WHEN in the future does the story take place? (select one)
 - o Now or the Near Future (a world recognizably similar to ours with a few changes, e.g., *Ex Machina*)
 - o The Distant Future (a world where many things have changed e.g., *Star Trek*)
 - o The Past/an Alternate Timeline
 - o Unknown/Indeterminable

- o Other...
- Explanation
- Who made the AI? (select one)
 - o Humans
 - o Machines/AIs
 - o Self-realizing/emergence OR accidental awakening (of something built by humans)
 - o Aliens
 - o Unknown/Ambiguous
 - o Other...
- Explanation
- What does the AI do in the world? What is its work or primary function? (select one)
 - o Military/warfighting
 - o Specialized physical workforce function (e.g., janitor, nanny)
 - o Love, companionship, friendship (e.g., robot girlfriend, Stepford wives)
 - o Logistics or information management (e.g., managing a city, analyzing a surveillance system, driving a car, piloting a spaceship)
 - o Solving an intellectual problem (e.g., writing a novel, curing a disease, calculating the fate of the universe, picking the president, pondering the meaning of life)
 - o Unknown/Pursues its own goals
 - o Other...
- Explanation
- What kind of intelligence does the AI have? (select one)
 - o Narrow. It might be very good at solving a particular kind of problem, but can't think through novel types of problem and doesn't have broad consciousness like a human.
 - o General. Equivalent roughly to a human. It can reason through many different kinds of problems.
 - o Super. Much more intelligent than a human on reasoning through most problems.
 - o Other. Hard to categorize, perhaps displaying very alien ways of thinking.
 - o Unknown/NA
 - o Other...
- Explanation
- Does the story discuss AI policies, governance structures, constraints, or other ways society manages (or fails to manage) the AI? (select one)
 - o Yes
 - o No
 - o Other...
- Explanation
- Does this story depict AI as hazardous or AI policies having unintended consequences? (select one)
 - o Yes
 - o No
 - o Other...
- Explanation
- Does the story show AI having widespread social or economic impacts? (select one)
 - o Yes, major impacts
 - o Yes, minor effects
 - o No
 - o Other...
- Explanation
- Overall does the story portray the impact of the AI as... (select one)
 - o Positive
 - o Negative
 - o Ambiguous
- Explanation
- Quote (optional)
- How long did it take you to read/code this story?
- Do you have any other comments you want to share? (i.e., if this story should be excluded from the study)
- Coder's name

Appendix 3: List of works analyzed

Title	Author
555	Robert Reed
"The Dog Said Bow-Wow"	Michael Swanwick
"Zen and the Art of Starship Maintenance"	Tobias Bucknell
A History of Terraforming	Robert Reed
A Response from EST17	Tom Purdom
A Stopped Clock	Madeline Ashby
Ancient Engines	Michael Swanwick
And the Dish Ran Away with the Spoon	Paul Di Filippo
Another Word for World	Ann Leckie
Article of Faith	Mike Resnick

Title	Author	Title	Author
Assassins	Jack Skillingstead and Burt Courtier	Robot	Helena Bell
Balancing Accounts	James L Cambias	Robots Don't Cry	Mike Resnick
Butterfly Bomb	Dominic Green	Rock of Ages	Jay Lake
Canoe	Nancy Kress	Sanjeev and Robotwallah	Ian McDonald
Capitalism in the 22nd Century or A.I.R	Geoff Ryman	Search Engine	Mary Rosenblum
Cat Pictures Please	Naomi Kritzer	Silence Like Diamonds	John Barnes
Checkerboard Planet	Eleanor Arnason	Silently and Very Fast	Catherynne M. Valente
Computer Virus	Nancy Kress	Sleepover	Alastair Reynolds
Crystal Nights	Greg Egan	STET	Sarah Gailey
Damage	David D. Levine	Steve Fever	Greg Egan
Death on Mars	Madeline Ashby	Technarion	Sean McMullen
Deus Ex Homine	Hannu Rajaniemi	The Best Christmas Ever	James Patrick Kelly
Dolly	Elizabeth Bear	The Big Ice	Jay Lake and Ruth Nestvold
Elephant on Table	Bruce Sterling	The Days Between	Allen M. Steele
Emergence	Gwyneth Jones	The Discovered Country	Ian R. MacLeod
Entanglement	Vandana Singh	The Djinn's Wife	Ian McDonald
Fandom for Robots	Vina Jie-Min Prasad	The Falls: A Luna Story	Ian McDonald
Far as You Can Go	Greg Van Eekhout	The Fulcrum	Gwyneth Jones
Fireborn	Robert Charles Wilson	The Game of Smash and Recovery	Kelly Link
Firstborn, Lastborn	Melissa Scott	The Island	Peter Watts
Flight from the Ages	Derek Künsken	The Last Boat-Builder in Bal-lyvoloon	Finbarr O'Reilly
For Want of a Nail	Mary Robinette Kowal	The Little Goddess	Ian McDonald
From Babel's Fall'n Glory We Fled	Michael Swanwick	The Meeker and the All-Seeing Eye	Matthew Kressel
Ghostweight	Toon Ha Lee	The One Who Isn't	Ted Kosmatka
Going Deep	James Patrick Kelly	The Other Gun	Neal Asher
Grand Jeté (The Great Leap)	Rachel Swirsky	The Residue of Fire	Robert Reed
Hello, Hello; Can You Hear Me, Hello	Seanan McGuire	The Rider	Jerome Cigut
Henry James, This One's for You	Jack McDevitt	The Smell of Orange Groves	Lavie Tidhar
I, Row-Boat	Cory Doctorow	The Voluntary State	Christopher Rowe
In Babelsberg	Alastair Reynolds	The Voyage Out	Gwyneth Jones
Jackie's Boy	Steven Popkes	There Used to be Olive Trees	Rich Larson
Jonas and the Fox	Rich Larson	They Have All One Breath	Karl Bunker
Kit: Some Assembly Required	Kathe Koja and Carter Scholz	Tideline	Elizabeth Bear
Laws of Survival	Nancy Kress	Today I Am Paul	Martin L. Shoemaker
Lighting Out	Ken MacLeod	Turing's Apples	Alien AI, maybe not long mention
Lobsters	Charles Stross	Tyche and the Ants	Hannu Rajaniemi
Machine Learning	Nancy Kress	Uncanny Valley	Greg Egan
Mayfly	Peter Watts and Derryl Murphy	Utopia, LOL?	Jamie Wahls
Men Are Trouble	James Patrick Kelly	Utriusque Cosmi	Robert Charles Wilson
Meshed	Rich Larson	Vladimir Chong Chooses to Die	Lavie Tidhar
Mika Model	Paolo Bacigalupi	Whending my Way Back Home	Bill Johnson
Mines	Eleanor Arnason	Who's Afraid of Wolf 359	Ken MacLeod
Moby Quilt	Eleanor Arnason	Zima Blue	Alastair Reynolds
Mother of Invention	Nnedi Okorafor		
Nightside on Callisto	Linda Nagata		
Number Thirty-Nine Skink	Suzanne Palmer		
Old Paint	Megan Lindholm		
Oracle	Greg Egan		

Acknowledgements The authors would like to thank our many collaborators on this project at ASU, *Slate* Magazine, and New America. We are particularly grateful for the hard work of our readers, Adam Clark and Samantha Geiser, and our undergraduate researchers, Sakshi Hedge and Michelle Emmanuelli, as well as our colleagues Joey Eschrich,

Kevin Bankston, Torie Bosch, Anthony Nguyen, Andrés Martínez, the New America events and communications teams, and the *Slate* editorial team. We are indebted to the excellent feedback and suggestions of our advisory board and the participants in our lively discussions at the Future Tense workshop and public event in May 2019. This research was made possible thanks to funding from the William and Flora Hewlett Foundation and Google.

Funding This work was supported by grants from the William and Flora Hewlett Foundation and Google as part of the AI Policy Futures project.

Data availability The data that support the findings of this study are available on request from the corresponding author.

Code availability Not applicable.

Declarations

Conflict of interest The authors does not have any conflict of interest.

Ethical approval The protocol was reviewed by Arizona State University's Institutional Review Board and determined to be exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Participants provided their informed consent regarding publishing their interview data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams D (1979) *The hitchhiker's guide to the galaxy*. Pan Books, London
- Bacigalupi P (2016) Mika model. *Slate future tense*. https://www.slate.com/articles/technology/future_tense/2016/04/mika_model_a_new_short_story_from_paolo_bacigalupi.html. Accessed 18 Nov 2020
- Bengtsson M (2016) How to plan and perform a qualitative study using content analysis. *NursingPlus Open* 2:8–14
- Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186. <https://doi.org/10.1126/science.aal4230>
- Čapek K (1921) *R.U.R.: Rossum's universal robots*. Aventinum, Prague
- Cave S, Dihal K (2019) Hopes and fears for intelligent machines in fiction and reality. *Nat Mach Intell* 1(2):74–78. <https://doi.org/10.1038/s42256-019-0020-9>
- Doctorow C (2006) *I, row-boat*. Overclocked: stories of the future present. Thunder's Mouth, New York
- Doctorow C (2019) Affordances. *Slate future tense*. <https://slate.com/technology/2019/10/affordances-cory-doctorow-sf-story-algorithmic-bias-facial-recognition.html>. Accessed 18 Nov 2020
- Finn E (2018) A smarter way to think about intelligent machines. *The New York Times*. <https://www.nytimes.com/2018/11/15/opinion/killer-robots-ai-humans.html>. Accessed 7 Aug 2020
- Gibson W (1984) *Neuromancer*. Ace, New York
- Guston DH (2013) Understanding 'anticipatory governance.' *Soc Stud Sci* 44(2):218–242. <https://doi.org/10.1177/0306312713508669>
- Hudson AD (2019). A priest, a rabbi, and a robot walk into a bar. *Slate Future Tense*. <https://slate.com/technology/2019/11/future-tense-fiction-priest-rabbi-robot-bar-andrew-hudson.html>. Accessed 18 Nov 2020
- Jasanoff S (2015) *Future imperfect: science, technology, and the imaginations of modernity*. *Dreamscapes of modernity: socio-technical imaginaries and the fabrication of power*. University of Chicago Press
- Kress N (2001) Computer virus. *Asimov's Science Fiction* 25(4):102–135
- Kritzer N (2015) Cat pictures please. *Clarkesworld Magazine* 100:35–42
- Leech BL (2002) Asking questions: techniques for semistructured interviews. *Polit Sci Polit* 35(4):665–668
- Mayor A (2018) *Gods and robots: myths, machines, and ancient dreams of technology*. Princeton University Press, Princeton
- McDevitt J (2005) Henry James, this one's for you. *Subterranean* 2:79–82
- McDonald I (2004) *River of gods*. Pyr, Amherst, NY
- Mintzer H (2020) Legal salvage. *Slate Future Tense*. <https://slate.com/technology/2020/07/legal-salvage-holli-mintzer.html>. Accessed 18 Nov 2020
- Noessel C (2019a) Gendered AI: gender and embodiment. *Sci-fi Interfaces*. <https://scifiinterfaces.com/2019/05/28/gendered-ai-gender-and-embodiment>. Accessed 7 Aug 2020
- Noessel C (2019b) Untold AI. *Sci-fi Interfaces*. <https://scifiinterfaces.com/tag/untold-ai/?order=asc>. Accessed 7 Aug 2020
- Older M (2019) Actually Naneen. *Slate Future Tense*. <https://slate.com/technology/2019/12/actually-naneen-malka-older-robot-nanny.html>. Accessed 18 Nov 2020
- Onyebuchi T (2020) How to pay reparations: a documentary. *Slate Future Tense*. <https://slate.com/technology/2020/08/how-to-pay-reparations-short-story.html>. Accessed 18 Nov 2020
- Rajaniemi H (2008) His master's voice. *Interzone* 218:34–39
- Reynolds A (2005) Zima blue. *Postscripts* 4:5–33
- Reznikova Z (2007) *Animal intelligence: from individual to social cognition*. Cambridge University Press
- Rowe C (2004) The voluntary state. *Sci Fiction*. https://www.lexal.net/scifi/scifiction/originals/originals_archive/rowe/rowe1.html. Accessed 18 Nov 2020
- Schroeder K (2020) The suicide of our troubles. *Slate Future Tense*. <https://slate.com/technology/2020/11/karl-schroeder-suicide-of-our-troubles.html>. Accessed 11 Dec 2020
- Simard S, Perry D, Jones M et al (1997) Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature* 388:579–582. <https://doi.org/10.1038/41557>
- Sterling B (2014) Keynote. *IMPAKT Festival*, Utrecht, The Netherlands. <https://youtu.be/tpoIB-xfnGI>. Accessed 18 Nov 2020
- Sterling B (2017) Elephant on table. In: Brin D, Potts SW (eds) *Chasing shadows: visions of our coming transparent world*. Tor, New York, pp 302–321

- Stross C (2001) Lobsters. *Asimov's science. Fiction* 25(6):70–89
- Wasserman EA (2006) *Comparative cognition: experimental explorations of animal intelligence*. Oxford University Press, Spain
- Wijeratne Y (2020) The state machine. *Slate Future Tense*. <https://slate.com/technology/2020/09/state-machine-yudhanjaya-wijeratne.html>. Accessed 18 Nov 2020
- Wohlleben P (2017) *The hidden life of trees*. William Collins, NY

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.