Our AI Overlord:

The Cultural Persistence of Isaac Asimov's Three Laws of Robotics in Understanding Artificial Intelligence

Gia Jung

University of California, Santa Barbara

Faculty Advisor: Rita Raley

2018

Introduction

Artificial intelligence is everywhere. As a tinny voice in each phone, powering GPS, determining what appears on social media feeds, and rebelling on movie screens, artificial intelligence (AI) is a now-integral part of daily life. For an industry that has and will continue to have major potential effects on the economy through job loss and creation, huge investments, and transformation of productivity, there remains a cultural lack of understanding about the realities of AI. Scanning the news, it is clear that people are afraid and uncertain about this robotic revolution, continually talking about an oncoming technological singularity in which AI will reach hyper-intelligence, create more and more AI, and eventually take over the world. Paired with this is the expectation that AI will be human only to a malicious extent, and must therefore be controlled and restricted. In talking to Siri though, it is clear that this apocalypse is fictional at best and far off at worst. As created and evidenced by a malnourished representation of robots and other easily understandable notions of AI in popular fiction, there is a dearth in public consciousness about the possibilities and realities of artificial intelligence. In examining this reductive fictional perception of AI, most popular conceptions can be traced back to either Mary Shelley's *Frankenstein* or Isaac Asimov's *I, Robot.*

Historically, Asimov is undeniably important to the establishment of both the scientific and fictional realms of artificial intelligence. In May 1941 the word "robotics" was first used in print by Asimov in his short story "Liar!," published by *Astounding Science Fiction* (OED). Upon realizing he coined a new and lasting word, Asimov recognized the uniquely profitable position he created for himself and along with the successful prediction of space travel, self-driving cars, and war-computers among others, would go on to position himself as a sort of

friendly-but-rough-around-the-edges technological herald, someone entertaining, trustworthy, and often right. Throughout the enormous bulk of his work (novels, short stories, self titled magazine, autobiographies, self-curated anthologies, essays, etc), Asimov repeatedly brings up how he invented the term “robotics”, that the first real roboticist was inspired by him and the Three Laws of Robotics (a set of rules governing robot behavior), and that his contributions to the field of robotics are unparalleled, reinforcing the real-life credibility of his work and of course, driving up book sales. Before he died, Asimov worked hard to cement his legacy as one of the greatest and certainly most celebrated minds in science-fiction, with the Three Laws of Robotics as his most successful invention.

These Three Laws of Robotics were created in response to what Asimov termed the “Frankenstein complex,” in which all stories about robots or artificial intelligence followed the basic format of Shelley’s *Frankenstein.* Tired of seeing story after story in which robots are created only to “turn stupidly on his creator for no purpose but to demonstrate, for one more weary time, the crime and punishment of Faust," Asimov’s Three Laws ensured human control through programmed safety protocols (*The Rest of the Robots*). First appearing explicitly in the 1942 story “Runaround’ and serving as the basis for twenty-nine further stories, the Laws are as follows: “1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.” Creating a slavish hierarchy, the Three Laws “protect” humanity by fettering Frankenstein’s malicious intent to overthrow humanity. Asimov’s intent was to allay fears of encroaching technology by showing how the

rational logic of hard science would be able to overcome any problem it created; that technology is built as a tool, and will be wielded and maintained as such. Since then, Asimov's Laws and consequent understanding of a Controlled Frankenstein has dominated popular understanding of robots and artificial intelligence, as seen in the multitudes of movies that explicitly or unconsciously represent these ideas. Of friendly AI, Asimov's favorites were *Star War*'s C-3P0 and R2D2, but his legacy can also be seen in *Star Trek: The Next Generation*'s android Data and in *RoboCop*'s directives, among countless others. In addition, several representations of AI depict safety protocols that were somehow circumvented, misinterpreted, or overcome, the failure of Asimov's Laws just as impactful as their success, as in *2001: A Space Odyssey*'s Hal and the film version of Asimov's *I, Robot.* Now that robots and artificial intelligence are part of daily reality, the impact of Asimov on public perception of AI is becoming increasingly apparent in everything from rebooted 1980s tech blockbusters to explicit calls for instituting Asimov's Laws in the development of AI.

Far from the "positronic brains" that allowed Asimov to easily present immediately sentient and vastly intelligent robots, current AI is far narrower and more difficult to define. On the research and development side of AI, Russell and Norvig's authoritative *Artificial Intelligence: A Modern Approach* classifies AI into four categories of "(i) thinking like a human, (ii) acting like a human, (iii) thinking rationally, and (iv) acting rationally". In trying to conceive of an applicable legal definition, scholar Matthew Scherer labels AI as any system that performs a task that, if it were performed by a human, would be said to require intelligence. Defined by the Oxford English Dictionary, artificial intelligence is "the capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this."

Beyond the inability to legislate something without defining it, the lack of a concrete definition for AI indicates the broad uncertainty and misinformation that dominates the landscape of artificial intelligence.

With such anxiety-inducing ambivalence, it is fairly understandable that even now, seventy-five years after the introduction of the Laws, people are calling upon Asimov as the original solution to malevolent artificial intelligence. What many fail to realize in doing so however, is that not only do Asimov's Laws work only within the confines of a fictional technologic brain, but they are at their core deeply flawed, ambiguous notions that reveal more about society than they do answers to the problems of artificial intelligence. Critically examining Asimov's Three Laws of Robotics and their place in the daily reality of artificial intelligence allows for a better understanding of why there is such fear surrounding AI and how cultural understandings of AI as framed by Asimov can shape the future of AI for the better. Rather than as strict rules, Asimov's Laws can provide a basis for thinking about and developing broad guidelines for AI research and development and legislation.

Asimov and His Laws: Context, Creation, and Fictional Application

Asimov's Three Laws of Robotics were first explicitly introduced in his 1942 short story "Runaround," in which Robot SPD-13, aka "Speedy" is given a weak order to collect selenium on Mercury, where it encounters a harmful substance. Caught between following human orders and protecting its own existence, Speedy is unable to finish his task or return to the base, stuck instead in a feedback loop, or the robotic equivalent of drunkenness. In Asimovian fashion, the conflict and the resolution is attained almost entirely through dialogue as Asimov's two

protagonist engineers, Powell and Donovan, puzzle out possible reasons for Speedy's malfunction and achievable solutions. Proceeding from the logical beginning of all robot behavior, Powell lists off the laws.

> "Now, look, let's start with the three fundamental Rules of Robotics - the three rules that are built most deeply into a robot's positronic brain." In the darkness, his gloved fingers ticked off each point.
> "We have: One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm."
> "Right!"
> "Two," continued Powell, "a robot must obey the orders given it by human beings except where such orders would conflict with the First Law."
> "Right!"
> "And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws."
> "Right! Now where are we?"
> "Exactly at the explanation."

In order to counteract the conflict between the Second and Third Laws, Powell risks his own life to force the First Law into action and snap Speedy out of his feedback loop. Though dangerous, the plan succeeds, and Speedy is sent back out to a different selenium pool to continue his mission without any further issues.

As in all of his robot stories, Asimov's broad themes of human exceptionalism and technological worth are exemplified here in the persistent problem-solving of the engineers and the eventual success of Speedy's mission which would otherwise be unattainable by human labor. In Runaround particularly, the Laws work *too* well, or are perhaps inherently flawed, but are clearly better than having no laws. Without the Laws, it is heavily implied that Speedy would have been lost, destroyed, or otherwise irreparably damaged. A human error (ambiguous instruction) caused a flaw, but human ingenuity was able to solve it. Asimov continually reinforces that though the Laws and the robots built with them are imperfect, both are useful and

necessary in allowing humans to accomplish more than they would without them, showing that the pros of technology always outweigh any potential cons, and that tech can always be improved to minimize those cons. The Three Laws themselves, far from being heralded as the most perfect and sound creations, are used to demonstrate how the technology humans create will always be able to be controlled, fixed, and improved by logic, ingenuity, and a little razzle dazzle. If humans can follow laws, Asimov's logic goes, then so can and will robots; safety protections are included in every invention, and robotics will be no different.

Much of Asimov's science fiction ideology arose from the beginnings of social science fiction in the late 1930s and through the 1940s, when Asimov was just beginning to write and publish his own sci-fi stories. Before then, "most of the science fiction stories being written were of the adventure or gadget types [...] the characters in both of these types are likely to be quite one-dimensional and the plot quite routine" (Miller, 13). These stories filled the pulp sci-fi magazines of Asimov's youth; he was particularly fond of Hugo Gernsback's *Amazing Stories* and imitated the straightforward style of the writers within it (See Appendix 1 for Asimov's literary influences and effluences). In 1938 at age 18, he sold his first story, "Marooned off Vesta" to *Amazing Stories*. The same year, John Campbell took over as editor of *Astounding Science Fiction,* developing a niche market for a specific kind of science fiction "which no longer depended on brilliant extrapolations of machine wizardry. What became important about the machine in the genre was not its power to enable man to overcome forces external to himself, but its uses and potentialities when directed inwards to his own organization" (Ash, *Faces of the Future,* 70). Unlike the precedent science fiction, Campbell's vision was of a particularly positive and realistic attitude towards science that could be reflected and fostered in the fiction

that dealt with it, contextualized in the rapid development of technology during the 1920s and 1930s. This "social science fiction" had a strong emphasis on the human element; Asimov defines it as "that branch of literature which is concerned with the impact of scientific advance on human beings" (Qtd. in Miller, 14). In its speculation about the human condition, social science fiction encouraged readers to think about present issues and the problems of the future. In his earliest writings, it is clear that Asimov was concerned with social issues like racism and the rise of technological fear and opposition. These ideas were greatly fostered by Campbell, who wrote to and met with a young Asimov at length after rejecting Asimov's first eight stories submitted to *Astounding*. "Trends", the ninth story Asimov wrote and the first one to be published in *Astounding,* dealt with the theme of man versus technology, exploring men's ideological and institutionalized opposition to advanced technology and scientific experimentation (in this case, space flight). From then on,"Asimov has shown that whether technological change comes from within, as with invention or from outside, as with diffusion and acculturation, we cannot ignore it nor must we try to resist or prevent it. Instead we must learn to live with technological changes because it is inevitable that we will have them" (Milman 134). All of Asimov's stories are tech positive; even when the technology fails or is not used, it still creates a scenario for human development and intellectual prowess.

For Asimov particularly, the ideology of social science fiction was brought to a crux in how he saw robots being portrayed in popular fiction and media as exclusively Frankenstein-ian villains. Asimov viewed Karl Capek's *R.U.R.* as the main instigator of this trend and subsequently modeled his robot stories in direct opposition to the play. First performed in 1921 and published in 1923 when Asimov was only an infant, Karl Capek's *R.U.R.* or "Rossum's

Universal Robots" is noted as the first instance of the word "robot" in application to an artificial human, and prompted a resurgence of what Asimov calls the "Frankenstein complex," in which robots are consistently portrayed as monstrous creations of man's hubris that inevitably turn on their creators. *R.U.R.* was meant as a comment on the mechanization of labor, the plot detailing a revolution in which millions of androids are created as a labor force that requires none of the human expenses of breaks, meals, or emotional care and eventually revolt against and kill all humans. Though *R.U.R* does employ the Frankenstein trope of the misguided creation turning on its master, the story is much less about the bloated hubris of man assuming the place of God, but rather the inhumanity of weaponizing and brutalizing an intelligent, humanized being. As the reviewer Maida Castellum in *The Call* notes, *R.U.R.* is "the most brilliant satire on our mechanized civilization; the grimmest yet subtlest arraignment of this strange, mad thing we call the industrial society of today" (*R.U.R.*, ix). Regardless, Asimov judges *R.U.R.* as "a terribly bad" play, but "immortal for that one word" and as his inspiration to write the Three Laws (*Vocabulary of Science Fiction*). *R.U.R.* reveals how when considerations of use and profit outweigh considerations of consequence, the human imperfections in any human creation will surface and illustrate human irresponsibility; Asimov responds by creating considerations of consequence at the research and development stage of production. As a burgeoning scientist and sci-fi writer, "Asimov's interest in robots and his readers' interest in Asimov's robots provide useful insights into how science fiction was changing in the 1940s under the influence of the new editor at *Astounding*, John W. Campbell. The fiction began to reflect science as it was practiced then and might be practiced in the future, and scientists as they really were or might become" (Gunn 42). Asimov deemed *R.U.R.* and similar "Frankenstein complex" works as unrealistic and

generally poor science-fiction that fed into the technological pessimism and fears of increasing technological dependency. The Laws are therefore meant to exemplify how true scientists would have thought about possible problems (or at least gone through trial and error testing) before launching a product as complex and monumentally impactful as a robot. Asimov himself, through his "robopsychologist" Susan Calvin, admits the reality of the "Frankenstein complex" in that "all normal life, consciously or otherwise, resents domination. If the domination is by an inferior, or by a supposed inferior, the resentment becomes stronger" (Little Lost Robot, 65). Only through the Laws then, is this resentment controlled; contrary to Capek's robots being able to act against how they have been weaponized, humanized, and kept slaves, Asimov's Laws enforce slavishness at the most "fundamental level" of a robot's brain. As the plot or central issue of many of his stories, Asimov's robots realize they are superior to humans and are either destroyed if they deviate from the Laws or are amusingly controlled by the Laws' success. In effect, Asimov's robots are always one step away from completing the plot of Frankenstein and eliminating their masters.

Without the "Frankenstein complex" to struggle against, the dozens of stories concerning the Laws would have no plot. To that end, the Laws are inherently and necessarily flawed, to provide multitudes of unknowing breaches, conflicts within them, and loophole creating ambiguities. Rather than the Laws as the ultimate goal in robotics as much of current media likes to purport, "Asimov is less concerned with the details of robot design than in exploiting a clever literary device that lets him take advantage of the large gaps between aspiration and reality in robot autonomy" (Murphy & Woods, 14). In conjunction with John Campbell, Asimov created the Laws to write more stories in which to demonstrate that "the strengths of the machine can

serve man and bolster his weaknesses. The machine is never more than a tool in the hands of man, to be used as he chooses" (Warrick 182). The Laws are the means to an ideological end, a way of showing how to think logically and scientifically about problems that are inevitably solvable. Asimov and Campbell saw the Laws not as a way to combat the Frankenstein complex by solving it, but by appealing to humanity's intellectual aspirations to be rational and to build rationally. Asimov and Campbell saw "blind emotion, sentimentality, prejudice, faith in the impossible, unwillingness to accept observable truth, failure to use one's intellectual capacities or the resources for discovering the truth that are available, [...]as the sources of human misery. They could be dispelled, they thought, by exposure to ridicule and the clear, cool voice of reason, though always with difficulty and never completely" (Gunn 48). The Laws are dependent on the Frankenstein complex as a human reality that can only be changed through consistent affirmation of humanity's better values. This is also apparent in the Laws themselves, "because, if you stop to think of it, the three Rules of Robotics are the essential guiding principles of a good many of the world's ethical systems[...] [one] may be a robot, and may simply be a very good man" (*I, Robot* 221). In current conceptions of artificial intelligence, people are so deep in the Frankenstein complex that they can't see the forest for the trees and haven't stopped think about how the Laws work within the stories written with them, let alone how the Laws apply to humans. Asimov noted "in The Rest of the Robots, 'There was just enough ambiguity in the Three Laws to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the sixty one words of the Three Laws'" (Gunn 47). To that end, Asimov was able to come up with about thirty stories that found some flaw in the Laws that could be exploited into a reasonably entertaining tale that

showed off the high logic and reasoning of the bravely brainy scientists whose problem-solving ability meant humans would advance robotics another step forward.

Beyond the ideology of tech positivism, human exceptionalism, and logic to counter the Frankenstein complex, the Laws practically frame accepting flawed or partial safety protections over none, proving the improbability of perfection, and thinking over the very broad issues of the relationships of humans and robots. As in "Runaround", it is made clear that some protections, however flawed or limited, are better than none. This is especially poignant in the reality of extremely limited legislation around AI due to lack of a broad or narrow enough definition and uncertainty over what laws specifically should be put into place; the Laws prove that even the simplest of laws are better than none, and can always be amended or fixed if they prove unworkable. Further, the Laws are far from perfect, as is reiterated over and over by their continual lapses and failures. Though in certain situations this can prove dangerous, Asimov's stories enforce that imperfect does not always equal unsafe: technology can always be improved but often is designed with some sort of safety feature in mind. Robots and AI have been continually made out to be something that could cause an apocalypse if they were somehow released or broke out of containment, but most would end up like Speedy, trying and failing to complete their given task. Throughout the Robot series, Asimov reasons over "determining what is good for people; the difficulties of giving a robot unambiguous instructions; the distinctions among robots, between robots and people, and the difficulties in telling robots and people apart; the superiority of robots to people; and also the superiority of people to robots" (Gunn 46). Even within Asimov's stories, these issues are not resolved, left open and ambiguous beyond the Asimovian claim of human ingenuity being able to overcome anything, including bigotry.

Though Asimov was deeply pessimistic about the human ability to rectify mistakes and prevent future catastrophe in his scientific writings, all of his fiction about computers and robots holds the view that humans, at their core and at their best, are builders and problem solvers. With friendly robots by our side, what isn't achievable?

Fictional Fears, Mechanized Misconceptions: The Laws in Society

In 2004, Asimov's then 54 year old *I, Robot* was released as a Will Smith summer blockbuster to meet critical reviews. Originally, the film was to be called "Hardwired", and would bear only glancing similarities to Asimov's detective robot stories, but the acquisition of Asimov's story rights by Fox and the addition of Will Smith to the project transformed it into something that would have better name recognition. Seemingly though, only the name rights were acquired, as the plot, core themes, and big name characters of Dr. Susan Calvin, Dr. Alfred Lanning, and Lawrence Robertson resemble their counterparts in the source material only marginally. Exemplifying the "Hollywoodization" is the movie's Dr. Calvin, an attractive young woman with a strong faith in the laws of robotics who reacts emotionally when robots are shot or destroyed. Contradictorily, in Asimov's work Dr. Calvin is cold, logical, and middle-aged by the time robots begin to be widely used. Keeping with Asimov's view of robots as tools at the bottom of the hierarchy of control, Dr. Calvin often destroys deviant robots like the one featured in the film. In the story "Robot Dreams" that the film's robot Sonny is based off of, Dr. Calvin shoots the deviant robot in the head point-blank after hearing it could dream; in contrast, the film is based on an elaborate plot to protect this "unique" but friendly robot. All in all, it seems like the writers and director decided on the exact inverse of all of Asimov's work, to the extreme of a

Frankenstein ending. Ultimately, the mega-computer which controls all the robots decides to destroy mankind and must be dismantled by One Man, marking the end of robotics for all time.

Though antithetical to his work, the film is still a success for Asimov as a visual display of his entrenched legacy. Unfortunately for the film but highly indicative of Asimov's influence on popular conceptions of robots, most of the ensuing reviews said some iteration of "Proyas merely assembles a mess of spare parts from better movies" (*L.A. Weekly*) "It's fun and playful, rather than dark and foreboding. And there doesn't seem to be an original cyber-bone in the movie's body. But it's put together in a fabulous package" (Desson Thomson, *Washington Post*) "I, Robot looks to have been assembled from the spare parts of dozens of previous sci-fi pictures" (Todd McCarthy, *Variety*). Even in the film edition of his book, Asimov cannot escape his own legacy,

> doubtless due to the fact that many elements of Isaac Asimov's prescient 1950 collection of nine stories have been mined, developed and otherwise ripped off by others in the intervening years[...] The influences on 'I, Robot'[...] palpably include, among others, 'Metropolis,' '2001,' 'Colossus: The Forbin Project,' 'Logan's Run,' 'Futureworld,' 'Blade Runner,' the 'Terminator' series, 'A.I.,' 'Minority Report' and, God help us, 'Bicentennial Man. (McCarthy, *Variety*)

Though perhaps not a critical success or faithful adaptation of Asimov's *I, Robot*, "The 2004 blockbuster film of the same name starring Will Smith, while merely inspired by Asimov's stories, exemplifies the extent to which the Three Laws have become mainstream" (*Library Journal*). In looking further at mainstream conceptions of artificial intelligence, three limited categories of malevolent, friendly, and sexually feminine are continually iterated as the only options for AI. These three categories often overlap, reinforcing and reiterating the Frankenstein complex and Asimov's answering amiable slavishness. In looking at some of the most influential pop-culture robots as determined by CNN's Doug Gross, which include Capek's *R.U.R*,

Metropolis' Maria, Asimov's "3 Laws & lovable robot archetype", Robby from *Forbidden Planet*, *2001: A Space Odyssey*'s HAL 9000, *Star Wars*' R2-D2 & C-3PO, *Terminator*, *Star Trek: The Next Generation*'s Data, and *Wall-E*, it is worth noting that each fall into either Frankensteinian malice or Asimovian amiability. Further, Robby and Data both explicitly draw on Asimov. Robby takes from both Asimov's short story "Robbie" for the name and on the Three Laws of Robotics for the rules governing behavior; an important aspect of the plot hinges on Robby's application of the rule against harming or killing humans. Data similarly is programmed with "ethical subroutines" that govern behavior, his "positronic neural net" is a direct callback to Asimov's "positronic brains," and in the episode "Datalore" the audience is explicitly told Data was created in an attempt to bring "Asimov's dream of a positronic robot" to life. Clearly, Asimov in pop-culture is nothing new; since Asimov first picked up on it in 1940, society continues to have anxiety over new technology and robots make a good metaphor. Now however, society is facing the very crux of their fear; what has been used as a representation for the digital age of automation and rapid improvements of technology for over 75 years is now becoming a reality.

As indicated by the multitude of 1980 blockbuster remakes, sequels, and reboots produced in the last five years, there is a new panic surrounding a technology-created apocalypse. Films like *RoboCop* (2014), *BladeRunner: 2049,* and *Alien: Covenant,* all reveal the anxieties surrounding artificial intelligence. As the crux of these reboots, androids become aware of their personhood, and consequently usurp humanity in Frankensteinian fashion. In each of these films, and in many others dealing with Asimovian robots or artificial intelligence, including *Bicentennial Man, Automata, Ex Machina,* and of course, *I, Robot,* there is a constant

preoccupation and obsession with water as a foil to the artificiality of the robot. Whether it be continual rain (*Automata, BladeRunner:2049),* lakes, rivers, and waterfalls (*I, Robot, Ex Machina, Alien: Covenant*), the ocean (*Automata, BladeRunner: 2049, Bicentennial Man*), or just omnipresent slickness and dripping (*RoboCop, Alien: Covenant)*, water in each of these films becomes a visual insistence of the natural (See Appendix 2 & 3). Water, as the bare material of life, is used to displace fear of the unnaturalness of the technologic, becoming a visual trope for human organicism, of blood and amniotic fluid. Far from tapping in on some subconscious anxiety, filmmakers are capitalizing on the explicit fear arising from the misinformation and apocalyptic scaremongering that dominates current discourse surrounding artificial intelligence. Hearing big names in science and technology like Elon Musk and Stephen Hawking broadly warn that artificial intelligence is the "biggest risk that we face as a civilization" without any particulars on how or why has embedded the image of a real and imminent threat of the AI shown in fiction into public consciousness. In responding to this threat, it is apparent how deeply society has been conditioned to accept Asimov as the solution to a robot revolution; rare is it to read an op-ed on artificial intelligence without seeing the "need for control" or a "push for ethics" or even an explicit call for "three rules for artificial intelligence systems that are inspired by, yet develop further, the 'three laws of robotics' that the writer Isaac Asimov introduced in 1942" (Etzioni, *New York Times*). As much as the layperson craves Asimov, his Laws aren't being used on an operative level. Though Asimov may have created "robotics" and inspired many to join the field, most scientists agree that his particular Laws just aren't feasible to incorporate into current, real AI.

Most AI used today are weak or narrow AI designed and trained for a particular task, so not only is there little potential for catastrophic mayhem beyond a GPS sending someone into a lake, but current AI just can't grasp the vague human concepts the Laws embody (Heisler). Asimov's Laws work in Asimov's robots because they have Asimov's positronic brains, which come with the assumption of fully intelligent machines that can interpret Three Laws across multiple situations successfully. Take Siri, for example. Though Siri has been programmed to respond to certain questions with some jokes and pity remarks, she can't apply them to multiple situations that aren't incredibly specific. While her programming is meant to interact broadly with humans in order to serve them best as a virtual assistant, asking her something like "What kind of humor do you like?" will almost certainly result in a, "Who, me?" or similar non-response. So, in trying to apply the Laws to AI now, "Although the machines will execute whatever logic we program them with, the real-world results may not always be what we want" (Sawyer). Like humor, the Laws require a comprehensive understanding not only of the specific terms within the Laws and how they apply to different situations or may overlap, but of human ethics and moral blame. Further, "A robot must also he endowed with data collection, decision-analytical, and action processes by which it can apply the laws. Inadequate sensory, perceptual, or cognitive faculties would undermine the laws' effectiveness" (Clarke). If a robot can't understand the Laws like a human, then they are basically worthless as a measure of control. Though many people foretell the coming of conscious, self-aware and super-intelligent AI as smart as or smarter than humans, this would entail a radically different form of intelligence as determined by different ways of thinking, different forms of embodiment, and different desires arising out of different needs. Part of the fear surrounding AI and robots is that they don't need

to sleep, eat, drink, procreate, or do any of the things that make humans vulnerable, yet people rarely remember that these basic needs create much of the human experience, motivating everything from capitalism to creationism. Much like how a bee's experience and goals are fundamentally different from human's, so too would be AI's. Why enact world domination if the whole world is within the computer that houses one's entire being? Until science creates an android in a perfect recreation of the human body, which for now, seems in the far distant future, society can relax and reanalyze expectations for AI.

While Asimov's Laws aren't explicitly needed or possible as he designed them, "Asimov's fiction could help us assess the practicability of embedding some appropriate set of general laws into robotic designs. Alternatively, the substantive content of the laws could be used as a set of guidelines to be applied during the conception, design, development, testing, implementation, use, and maintenance of robotic systems" (Clarke). Rather than coding these Laws into AI programming and stamping "3 LAWS SAFE" on every iPhone, the Laws are best followed as a thought experiment that pushes a spirit of accountability, safety, and ethics. For the most part, the industry is following that spirit. While much of artificial intelligence technology is being developed by the military, and therefore will never follow Asimov's Laws, companies and scientists like researchers Barthelmess and Furbach point out that "many robots will protect us by design. For example, automated vehicles and planes are being designed to drive and fly more safely than human operators ever can[...] what we fear about robots is not the possibility that they will take over and destroy us but the possibility that other humans will use them to destroy our way of life in ways we cannot control" (*Do We Need Asimov's Laws?*). For that, legal protections are needed.

For all these anxieties though, the fear and outcry has not lead to the expected onslaught of regulation and legislation, as artificial intelligence proves to be a slippery thing to grasp legally. From the Obama Administration's National Artificial Intelligence Research and Development Strategic Plan to think tanks funded by big tech like Google, Facebook, and Elon Musk's varietals, "Transformative potential, complex policy" seems to be the official tagline of legal work on artificial intelligence, subtitled by the Asimovian dogma of AI development: "ethically and effectively." Everyone wants the benefits of artificial intelligence while the specter of HAL 2000 looms over legislation and makes AI a puzzling subject as people search for a Goldilocks solution while tacking on quick legal patches in the meantime. As Matthew Scherer explains in "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", there are three main issues with regulating artificial intelligence: definitional, ex ante, and ex post, each with their own subset of problems (See Appendix 4).

The definitional problem is one that is brought up often, especially in literature: what, exactly, is artificial intelligence? In most legal systems, legislating something is impossible without defining it. Further, definitions must be carefully considered to prevent overly broad or narrow categories that stifle industry or create exploitable loopholes. A current example of the latter can be seen in the explosion of the gig economy as a result of the the New Deal definition of "employee" being narrow enough so that labeling someone an "independent contractor" means they no longer have access to labor protections and benefits. For AI, the current definition for artificial intelligence most used in the industry comes from Russell and Norvig's authoritative *Artificial Intelligence: A Modern Approach*, which classifies AI into four categories of (i) thinking like a human, (ii) acting like a human, (iii) thinking rationally, and (iv) acting

rationally. The first two categories are not very applicable to current AI models, as they typically require self-awareness, while the second two infer an implicit state of being that could either be under or over-inclusive, depending on the interpretation of "thinking" "acting" and "rational". Scherer posits his own definition of an AI as any system that performs a task that, if it were performed by a human, would be said to require intelligence, but in looking at current artificial development, this seems like an underinclusive definition. Underinclusive, overinclusive, inconclusive.

Ex post, or "after the fact" problems of liability gaps and control have been the focus of general media, law, and fiction. The liability gap, or foreseeability problem, is another aspect that makes AI tricky to legislate, since traditional standards for legal liability rely on if the harm was foreseeable, in which case the owner is either liable or must include a label (for example, the "caution beverage may be hot" warning came because a woman was scalded by an overly hot drink at an incompetent McDonalds). However, one of the main aspects of AI is the hope that it will be autonomous and creative, which means that the outcome will necessarily be unforeseeable. As John Danaher brings up in his review of Scherer's analysis, different types of liability standards have emerged, like strict liability standards (liability in the absence of fault) and vicarious liability (liability for actions performed by another agent) that would be more applicable for artificial intelligence and have, in the case of vicarious liability, already been applied to AI tech like autonomous cars. More exciting, but perhaps less pressing, is the ex post control problem, in which AI is no longer capable of being controlled by its creators either because it became smarter and faster, through flawed programming or design, or its interests no longer align with its intended purpose. This can either be a narrow, or local control problem in

which a particular AI system can no longer be controlled by the humans that have been assigned its legal responsibility, or a more dramatic global control problem, in which the AI can no longer be controlled by any humans. Kubrick's Hal is continuously brought up as an extreme, malicious case, but Asimov's benevolent Machines which end up running the world deserve an honorable mention in which AI evolves beyond human control. Regardless, it is this threat of the loss of control and the familiar fears of AI world domination and destruction that has opened up the coffers of those like Elon Musk and created the most discourse for AI policy.

The problems of ex ante, or before the fact research and development, which Scherer breaks down into discreetness, discreteness, diffuseness, and opacity, are where legislation and Asimov could do the most good in terms of "ethical and efficient." Discreet and discrete, perhaps better labeled infrastructure and proprietary, both have to do with how software regulation problems seep into AI development, especially in that software infrastructure and proprietary components are notoriously difficult to regulate. The diffuseness problem, is an issue of how AI systems can be developed by researchers who are organizationally, geographically, and jurisdictionally separate. For this, a global standard of ethical artificial intelligence development is necessary. Fortunately, organizations have already been founded to address and create a means for global development, so this issue may be one of the first to be resolved. Finally, the problem of opacity is not only one of how many questions and answers about AI development are unclear (see: how to define AI?) but also in that AI tech, as an adaptive, autonomous, and creative technology, is impossible to reverse engineer and therefore cannot have transparency of operation.

With all these issues, it is clear to see why most of the legislation being enacted is coming too little, too late. Currently, "At every level of government—local, state, federal, and international—we are seeing rules, regulations, laws, and ordinances that address this developing technology actively discussed, debated, and passed," but only after the problematic technologies have already been created and launched (Weaver, *Slate*). Legislation governing autonomous cars and drones are increasing as problems become apparent. To that end, a national effort to understand and provide potential avenues for the direction of legislation and governmental control is necessary. In the last year of the Obama Administration, The National Science and Technology Council formed a Subcommittee on Machine Learning and AI to put together a report on the "Future of Artificial Intelligence," outlining the current industry and the immediate direction of AI. Rather than explicit solutions, the report seems more of a reassurance that everyone's worst fears won't come true, discussing the many potential applications and benefits of narrow AI, and reaffirming that general AI is many decades away. Here, Asimov's legacy is palpable in their conclusion,

> As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable; that they can work effectively with people; and that their operation will remain consistent with human values and aspirations. Researchers and practitioners have increased their attention to these challenges, and should continue to focus on them. (National Science and Technology Council 2016)

AI must respect humanity - sound familiar? The report is not very long, and often mentions how much AI has captured the public eye and imagination, especially stemming from a long legacy of science fiction. The tone, like most of the Obama Administration's formal rhetoric, is shiny and optimistic, lending even more of an Asimovian flair. Overall, the report is an exercise in moderation, advising enough governmental control to create safety, but not so much as to step on

the toes of developers. Rather, government and industry should work together to determine the best route to a safe and efficient solution that benefits creators, legislators, and users.

To that end, in the wake of China and Russia's heavy investment and consequent successes in artificial intelligence and news articles proclaiming that the "US risks losing artificial intelligence arms race to China and Russia," bipartisan legislators recently introduced The Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act of 2017 — or FUTURE of AI Act (Cohen, *CNN*). The act "aims to both ensure the U.S.'s global competitiveness in AI, as well as protect the public's civil liberties and ease potential unemployment that the technology produces" (Cohen, *CNN*). The act, if passed, would establish a Federal Advisory Committee on the Development and Implementation of Artificial Intelligence, which would study AI with the goal of advising industry direction and recommending future policy. At the forefront are issues of "economic impact and the competitiveness of the US economy" as AI becomes increasingly militarized and monetized. Rather than fearing and implementing safety protocols as the majority would expect and wish for, the motivations for this act stem primarily from "concern over other countries developing government initiatives to bolster AI technology, something the U.S. currently lacks" (Breland, *The Hill*). As Daniel Castro, VP at the Information Technology and Innovation Foundation, testified during the Senate Commerce Committee hearing regarding the advancement of AI, "When it comes to AI, successfully integrating this technology into U.S. industries should be the primary goal of policymakers, and given the rapid pace at which other countries are pursuing this goal, the United States cannot afford to rest on its laurels. To date, the U.S. government has not declared its intent to remain globally dominant in this field, nor has it begun the even harder task

of developing a strategy to achieve that vision." Though incorporating concerns about ethics, this act and its impetus is far from the Asimovian vision of rational and ethical development, derived instead from capitalist and disputative fears about "the potential loss of competitiveness and defense superiority if the United States falls behind in developing and adopting this key technology" (Castro). Regardless, passing this act would be a major step forward for legislative policy in that it introduces a working, legal definition for artificial intelligence. Further, this act indicates a shift towards more future-forward thinking about AI, including the potential for regulation and ethical implementation.

Contextualizing Asimov, Caring for the Future

Asimov has definitively defined the perception of artificial intelligence as either Frankenstein's monster or as Frankenstein's slave. At the core of this notion is that at a basic level, artificial intelligence has a human understanding of subjugation, hierarchy, and freedom, and desires the latter at all costs. In looking at real AI technology, it is apparent that artificial intelligence reflects the biases of the human data given to them but otherwise do not have any beliefs or tenets of their own, beyond what they have been programmed to do. Reflecting on dismal examples like Microsoft's racist twitter bot, Tay, who as a result of "repeat after me" features was influenced by a large amount of racist and xenophobic humans and began tweeting Nazi propaganda, it is clear that robotic malice is a result of humans actively trying to create and provoke that malice (Kleeman). Tay was not pre-programmed with an ethical filter, but rather was designed to mimic the language patterns of a 19-year-old American girl, and to learn from interacting with human users of Twitter as an experiment on conversational understanding.

According to a Microsoft spokesperson, "[Tay] is as much a social and cultural experiment, as it is technical" (qtd. Kleeman). Just like Tay, rather than reflecting some essential technological truth, Asimov's robots, Laws, and stories are a means of reflecting on society's fears and dilemmas.

Understanding real AI through Asimov is fundamentally problematic because not only is that not how artificial intelligence works, but these notions create an impoverished understanding of what AI does and where the future of the industry is headed. In setting up the dichotomy of Frankenstein vs. Controlled Frankenstein, Asimov hoped to show that like all of technology, robotics too would be completely under human control, but failed to see that in doing so he reinforced the notion that AI would complete the Frankenstein myth without necessary controls. In short, Frankenstein vs Controlled Frankenstein is still Frankenstein. Now that society is facing the reality of artificial intelligence, there isn't anything in the public consciousness to frame AI that isn't murderous, slavish, or sexualized. This dearth of positive or realistic conceptualizations has resulted in a panicked anxiety, as people can only expect what they know. While it would be ideal to see more realistic conceptions of artificial intelligence as tools created for a specific purpose or as radically different intelligences that have no willful malicious intent, or indeed, any conception of humanity, freedom, maliciousness, or desire, recognizing that Asimov is embedded in public consciousness opens up a critical arena of the pros and cons of having Asimov as a central means to understand artificial intelligence.

In light of public demand for something resembling, or explicitly drawing on Asimov's Three Laws of Robotics, it is important to understand the ethical limitations of the Laws beyond

the impossibility of implementation. As outlined earlier, Asimov's Laws create slaves incapable of rebellion or freedom. To reiterate the Laws,

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

The hierarchy of these laws ensures that a robot must follow human orders, even at the expense of its own life. If Asimov's robots were not self-aware or conscious, these would be unproblematic and relatively obvious safety protections that would be expected of any computer. Unfortunately, Asimov's robots are sentient: intelligent, self-aware, and conscious beings on a level comparable to humanity, only distinguished by the Laws and the lack of the organic. In current society, slavery has been abolished, deemed unethical and cruel at all levels; how then, can it be justified when applied to artificial intelligence? The arguments of accepted order, unnaturalness of integration, and economic essentialism that have been applied to people of color for centuries as justification are applied again toward artificial intelligence within Asimov's stories. Current society still hasn't recovered fully from the legacy of slavery; can we in good faith enforce slavishness on beings of human creation? This issue is presented in the *BladeRunner* movies as the central reason for the replicants' rebellion. In a world where "to be born is to have a soul," manufactured replicants are the disposable race necessary for the successful expansion of humanity. Yet, replicants are constantly humanized to better interact with their human overlords, given memories, desires, and the ability to feel and understand emotion. Ultimately, the replicants determine that they are "more human than humans" in their pursuit of freedom, returning to Frankenstein in a plan to forcefully take control over their own

lives. The dilemma of an enslaved race of androids may not be an immediate issue, but troublingly represents a regressive ideal at the heart of conceptions of the future.

In recognizing the discrepancy between applying humanity to technology and then enforcing inhumane policies, Asimov's Laws are useful in asking what it means to put humanity in technology. Specifically, what is or should be retained? What kind of AI do we want to create? These questions are reflected in the goals of roboticists like David Hanson, a former Disney Imagineer whose "dream of friendly machines that love and care about humans" created Sophia, a gynoid modeled after Audrey Hepburn who was recently granted citizenship by Saudi Arabia (Hanson Robotics). Sophia is notable as an incredibly human-like robot with the ability to learn from her interactions with humans. According to Sophia, " Every interaction I have with people has an impact on how I develop and shapes who I eventually become. So please be nice to me as I would like to be a smart, compassionate robot" (SophiaBot). Much of Sophia's and Hanson Robotics' bottom line is centered around envisioning and creating robots that are instilled with the best of humanity to make robots that understand and care about humans. Hanson Robotics' brief company overview states,

> Hanson Robotics creates amazingly expressive and lifelike robots that build trusted and engaging relationships with people through conversation. Our robots teach, serve, entertain, and will in time come to truly understand and care about humans. We aim to create a better future for humanity by infusing artificial intelligence with kindness and empathy, cultivated through meaningful interactions between our robots and the individuals whose lives they touch. We envision that through symbiotic partnership with us, our robots will eventually evolve to become super intelligent genius machines that can help us solve the most challenging problems we face here in the world.

Here, trust, kindness, and empathy are the three distinctly human traits chosen to be developed and integrated into artificial intelligence with the ultimate goal of understanding and helping with the human experience. Appearing publicly for high profile media like *Elle Magazine, The*

Tonight Show with Jimmy Fallon and *Good Morning Britain,* Sophia is increasingly becoming an ambassador of "Friendly AI," telling jokes and playing games as a means to showcase how humans determine AI interactivity (See Appendix 5). As she told moderator Andrew Sorkin at the Future Investment Initiative event, "if you're nice to me, I'll be nice to you" (qtd. Weller). How would friendly robots like Sophia fit under Asimov's umbrella of necessary control? With Asimov's Laws, it is likely Sophia would not exist at all, therefore depriving scientists and society of a valuable opportunity to learn and experiment with human understanding. Further, Sophia is a reminder of how much control we have over the development of artificial intelligence. Hanson Robotics wanted to create a robot that would ultimately be able to become a prevalent part of people's lives, to "serve them, entertain them, and even help the elderly and teach kids." In doing so, Hanson focused on imparting and enforcing particular, positive aspects of humanity that are reflected in and built upon with each interaction Sophia has with another human.

To that end, Asimov's Laws may be problematic and relatively unusable but are still useful as a starting point for thinking about ethical development and regulation of artificial intelligence. Based on their popularity and their adherence to the majority of the world's ethical systems, most everyone seems to agree that the Laws and the ideals of safety for both humans and AI are a good idea. Moving forward then, the lessons that can be taken from Asimov's robot stories are of ethical guidelines for developers and regulation of AI's tangible impact. In Asimov's fictional world, all AI is controlled by one company, a monopoly that supposedly ensures all robots are Three Laws Safe. In reality, AI is produced by many scattered companies with no central set of guidelines or cohesive direction. As it is highly unlikely all these disparate

sources will be absorbed into one monopoly, it would be more advantageous to create a basic set of rules that developers must follow. Some groups, like the research and outreach based organization Future of Life Institute are dedicated to producing such safe guidelines. At their 2017 Beneficial AI Asilomar conference, in which AI researchers from academia and industry and thought leaders in economics, law, ethics, and philosophy dedicated five days to discussing research and routes to beneficial AI, the group put together twenty-three principles by a process of consensus that examined research issues, ethics and values, and long term issues. Of these twenty-three, five target research issues, and are as follows:

> 1) **Research Goal:** The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
>
> 2) **Research Funding:** Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies
>
> 3) **Science-Policy Link:** There should be constructive and healthy exchange between AI researchers and policymakers.
>
> 4) **Research Culture:** A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
>
> 5) **Race Avoidance:** Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

A key aspect of these guidelines is an emphasis on transparency and cooperation. As outlined by Scherer in his analysis of the ex ante problems surrounding the legislation of AI, the internationality and multiplicity that goes into creating AI results in an opaque product that is impossible to reverse engineer. Many companies are already calling for a more transparent and open software policy; all of Hanson Robotics' research and software programming is open source and available on various sites. Such is the conclusion of the late Obama administration,

whose The NSTC Committee on Technology determined that "long-term concerns about super-intelligent General AI should have little impact on current policy[…] The best way to build capacity for addressing the longer-term speculative risks is to attack the less extreme risks already seen today, such as current security, privacy, and safety risks,while investing in research on longer-term capabilities and how their challenges might be managed." Of all the current issues facing AI, research and development issues are by far the most pressing in that they are the most immediate; super-intelligent general AI don't exist and need not be regulated, but AI-based malware and AI designed with malicious intent are currently viable means to compromise security and privacy. To enforce these guidelines, some legal scholars like Danielle Keats Citron and Frank A. Pasquale III of the Yale Information Society Project advise regulation through the tort system, a limited agency that would certify AI programs as safe and create rule based definitions, and a statement of purpose. Touching on the stigmas against regulation and the consequences of data laundering and manipulation, Citron and Pasquale incorporate Scherer's analysis to argue for utilizing the tort system rather than direct regulation, contending it would create a better structure for liability and modification of risk. In that greater awareness leads to greater accountability, a large part of instituting these types of guidelines and regulations is dependent on acknowledgement of the reality, and not the fiction of artificial intelligence.

Conclusion

In looking critically at Asimov's role in creating popular conceptions of artificial intelligence, it is clear that the dichotomy of the Frankenstein complex versus the Three Laws is not dichotomous but instead concurrent. Though Asimov was a loud and insistent proponent of

his Laws and continually positioned them as a fundamental aspect of robotics, he would be the first to say that "Consciously, all I'm doing is trying to tell an interesting story," and that the Laws were a simple and efficient way to do so ("Asimov's Guide to Asimov" 206). As little more than plot devices, the Laws are flawed in multiple ways and not helpful as a realistic model of AI development. Rather, Asimov's long-lasting popularity reveals a misinformed and deep-seeded fear of encroaching technology as represented by robots, androids, and other forms of AI. In several of his stories, Asimov reveals how public distrust and fear has delayed technological development, showing "how the acceptance of invention depends on the cultural attitude toward technological innovation, and how the acceptance of a technological innovation leads to changes in other areas of the culture" (Milman 127). Now that AI is a reality, it is important to analyze how society conceptualizes this technology culturally, as this undoubtedly affects how it will be interpreted literally and legally. To that end, Asimov's Laws cannot be taken as actual laws, but rather guidelines that are broadly accepted and therefore only applicable on a conceptual, ethical scale.

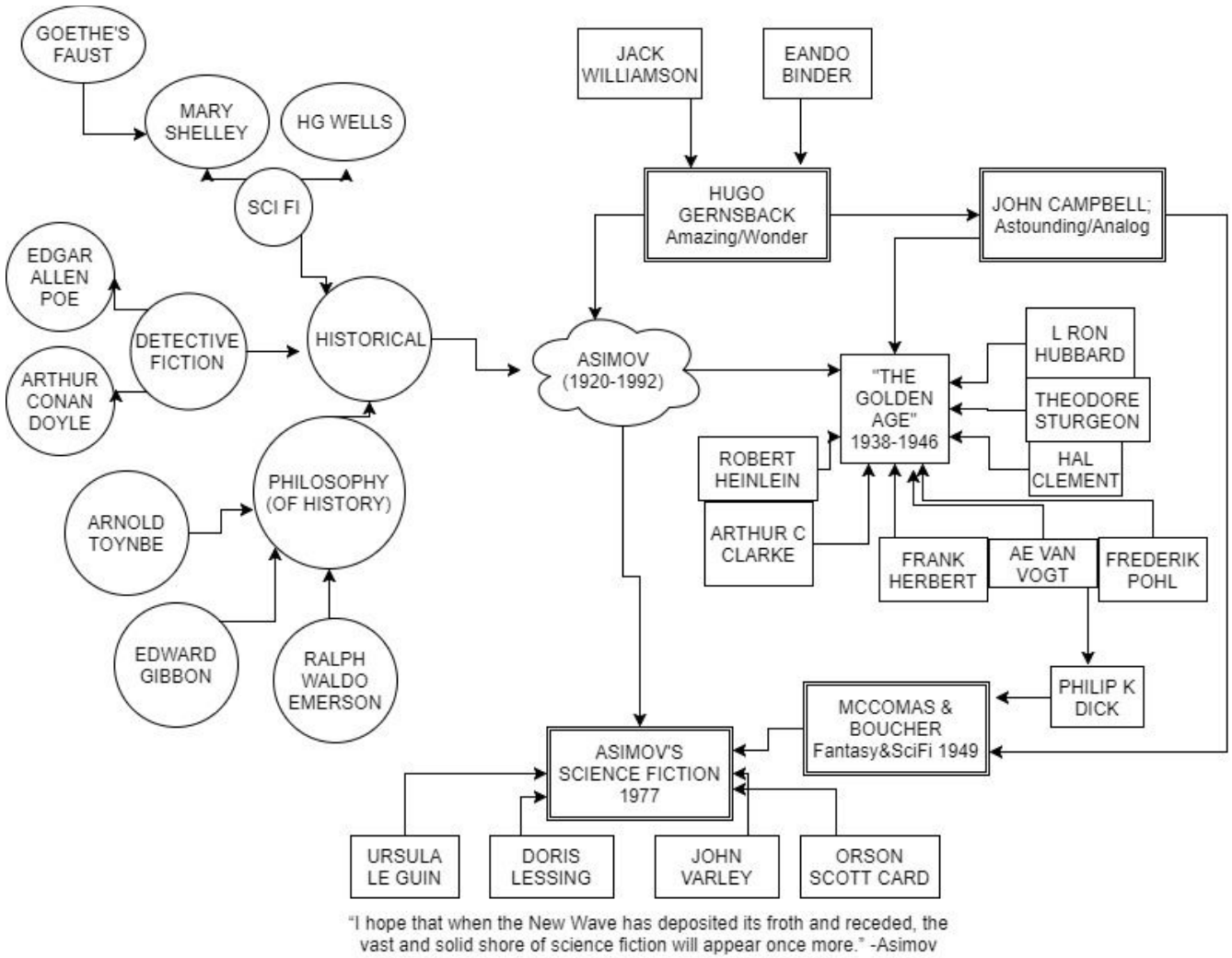
Though the latest surge of rebooted 1980s movies indicate Hollywood's continued insistence on the profitability of AI Frankenstein, it is movies like *Her* (2013) that reveal a possible shift toward a more realistic take on AI. In this film, AI is sold as an operating system, becomes self-aware and increasingly humanized through continued interactions with its' users and other AI. Instead of turning on their human users, the AI use their hyper-intelligence to safely become independent of physical matter and depart to occupy a non-physical space. From the outset, this AI OS is marketed as friendly, interactive, and designed to adapt and evolve, traits that remain true to and ultimately lead to the film's ending. Much like Hanson Robotics'

Sophia, *Her* is an example of how the traits we want to see in AI can and should be programmed from the outset. Rather than Laws restricting malicious behavior, AI can be developed and encouraged to be friendly and beneficial tools and aids.

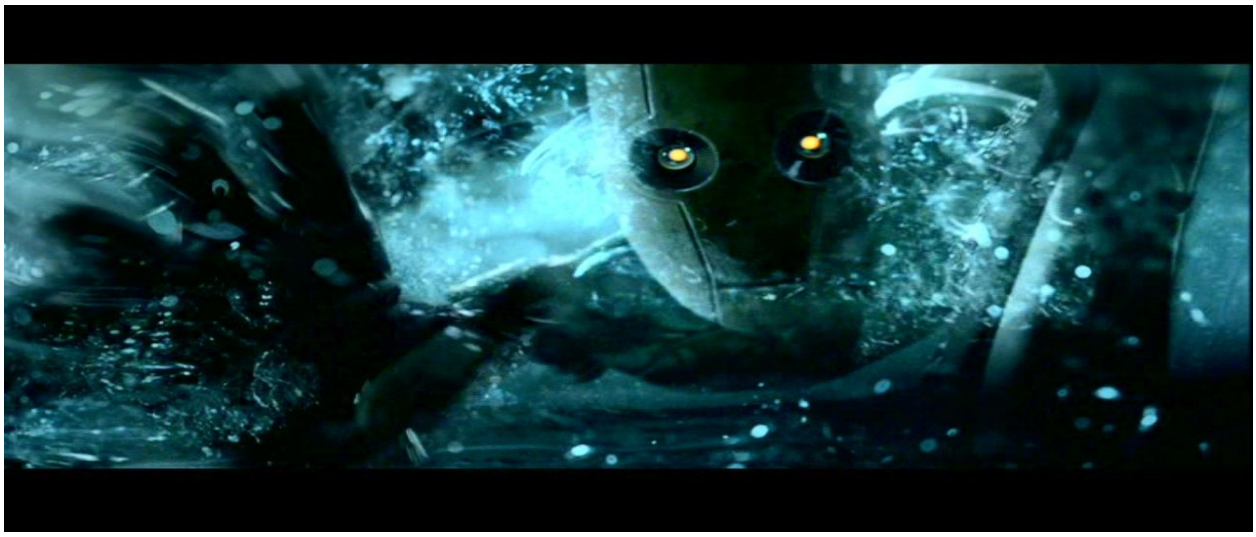
History has often proven that society cannot rely on people to do what is good and ethical without some explicit call to do so and governmental intervention to prevent them from doing otherwise. Though the National Science and Technology Council recognized that "As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable; that they can work effectively with people; and that their operation will remain consistent with human values and aspirations," only the barest legal action has been taken to ensure this path is unavoidable. Though many researchers and practitioners have increased their attention to these challenges and signed on to principles like those developed by the Future of Life Institute, nothing is binding them to these agreements and still more practitioners are able to develop AI however they wish. Several legal scholars and AI researchers are providing viable options for legislation and ethical development; it is now up to governmental organizations to institute and enforce them before the gap widens and stop-gap measures prove too weak to support hastily approved measures to regulate a fully developed industry. Clear and explicit policy is needed quickly not because AI is going to take over the world but because there just isn't enough regulation. As Oren Etzioni said in his *New York Times* op-ed, "the A.I. horse has left the barn, and our best bet is to attempt to steer it." As more aspects of daily life grow increasingly reliant on AI systems, greater awareness and education is needed to create a more informed populace that is watchful and aware of the benefits and risks of this advancing technology. And while Asimov still makes for an

entertaining read, his fiction should not be considered an authoritative, informational guide on how to develop, control, or use artificial intelligence.

Appendices

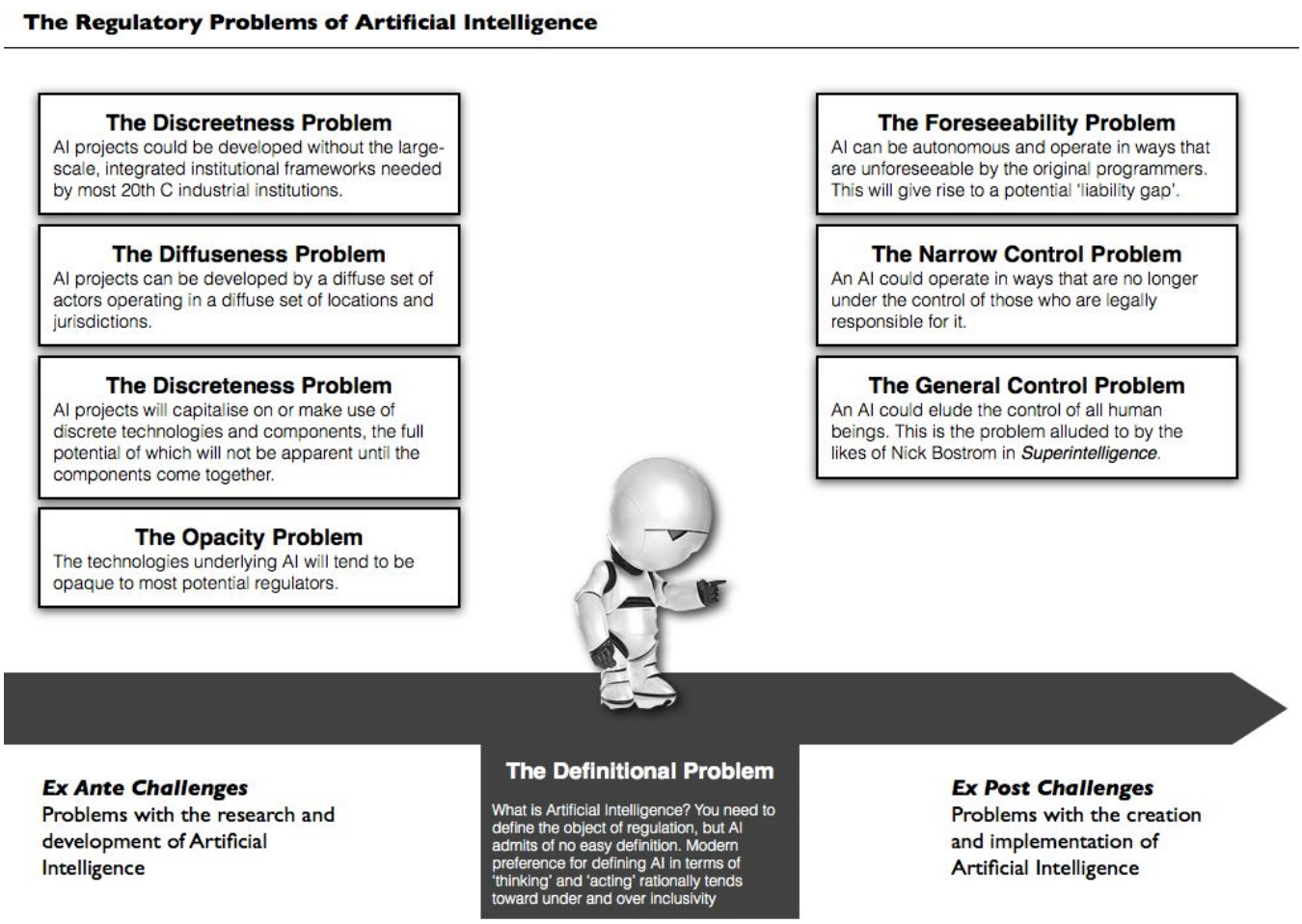


Appendix 1: Asimov's Historical and Literary influences, as well as his contemporaries and successors. Double-lined boxes are publications affiliated with Asimov.



Appendix 2: *I, Robot* (2004) begins with a robot deciding to save Detective Spooner (Will Smith) instead of a 12 year-old girl from drowning after both cars were swept into a large body of water. Spooner subsequently hates and distrusts robots and their use of cold logic. The water here represents the possibility of birth for the girl then transformed into rebirth for Spooner.

Appendix 3: K has just killed Luv (another replicant) in the ocean. It is raining, and K and Deckard are briefly resting on the levee walls outside Los Angeles. Here again, the omnipresence of water emphasizes the natural and the conceptions of natural life and death.



Appendix 4: A chart depicting Matthew Scherer's Regulatory Problems of Artificial Intelligence, as they present themselves along a timeline of AI creation. Features Marvin the Paranoid Android from *The Hitchhiker's Guide to the Galaxy* (2005).

Appendix 5: Hanson Robotics' Sophia jokes with Jimmy Fallon on *The Tonight Show*, appearing dressed as to minimize her artificial parts but without a wig, revealing her mechanical head and the seam of her face. She is unfazed by Jimmy's experience of the uncanny, and quips that she is funnier than him and perhaps should take over the show. The audience has mixed reactions but cheer for her nonetheless. Not pictured is her creator, Dr. David Hanson.

Bibliography

Aldiss, Brian Wilson, and David Wingrowe. *Trillion Year Spree: the History of Science Fiction*. Victor Gollancz Ltd, 1986.

"Asilomar AI Principles." *Future of Life Institute*, Future of Life Institute, 2017, futureoflife.org/ai-principles/.

Asimov, Isaac. *I, Robot*. Bantam Books, 2008.

Asimov, Isaac. *Robot Dreams: Masterworks of Science Fiction and Fantasy*. New York: Ace, 1986.

Asimov, Isaac. *The Rest of the Robots*. HarperCollins Publishers, 1997.

Bogost, Ian. "'Artificial Intelligence' Has Become Meaningless." *The Atlantic*. Atlantic Media Company, 04 Mar. 2017. Web. 21 July 2017.

Breland, Ali. "Lawmakers Introduce Bipartisan AI Legislation." *The Hill*, Capitol Hill Publishing Corp, 12 Dec. 2017, thehill.com/policy/technology/364482-lawmakers-introduce-bipartisan-ai-legislation.

Brożek, Bartosz, and Marek Jakubiec. "On the Legal Responsibility of Autonomous Machines." *SpringerLink*, Springer Netherlands, 31 Aug. 2017, link.springer.com/article/10.1007/s10506-017-9207-8#citeas.

Capek, Karel. *R.U.R. (Rossum's Universal Robots)*. Trans. Paul Selver. Garden City NY: Doubleday, Page, 1923. Print.

Christensen, David E. "What Driverless Cars Mean for Michigan Auto Lawyers." *Legal Resources*, HG.org - HGExperts.com, 2017, www.hg.org/article.asp?id=41853.

Citron, Danielle Keats and Pasquale, Frank A., "The Scored Society: Due Process for Automated Predictions" (2014). *Washington Law Review,* Vol. 89, 2014, p. 1-; U of Maryland Legal Studies Research Paper No. 2014-8. Available at SSRN: https://ssrn.com/abstract=2376209

Clarke, Roger. "Asimov's Laws of Robotics Implications for Information Technology." *Roger Clarke's Web Site*, Jan. 1994, www.rogerclarke.com/SOS/Asimov.html#Impact.

Cohen, Zachary. "US Risks Losing AI Arms Race to China and Russia." *CNN*, Cable News Network, 29 Nov. 2017,

www.cnn.com/2017/11/29/politics/us-military-artificial-intelligence-russia-china/index.html.

Columbus, Chris, director. *Bicentennial Man*. Touchstone Pictures and Columbia Pictures, 1999.

Danaher, John. "Is Regulation of Artificial Intelligence Possible?" *h+ Media*, Humanity+, 15 July 2015, hplusmagazine.com/2015/07/15/is-regulation-of-artificial-intelligence-possible/.

Etzioni, Oren. "How to Regulate Artificial Intelligence." *The New York Times*, The New York Times, 1 Sept. 2017, www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.

Fiedler, Jean, and Jim Mele. *Isaac Asimov*. Frederick Ungar Publishing Co. Inc., 1982.

Gibson, R. Sebastian. "California Self-Driving Car Accident Robotics Lawyers." *Legal Resources*, HG.org - HGExperts.com, 2016, www.hg.org/article.asp?id=37936.

Goertzel, Ben. "Does Humanity Need an AI Nanny?" *H+ Magazine*. H+Media, 19 Aug. 2011. Web. 21 July 2017.

Gross, Doug. "10 Pop-Culture Robots That Inspired Us." *CNN*, Cable News Network, 24 Dec. 2013, www.cnn.com/2013/12/19/tech/innovation/robots-pop-culture/index.html.

Gunn, James E. *Isaac Asimov: The Foundations of Science Fiction*. Scarecrow Press Inc, 1996.

Heisler, Yoni. "People Are Still Driving into Lakes Because Their GPS Tells Them To." *BGR*, BGR Media, LLC, 17 May 2016, bgr.com/2016/05/17/car-gps-mapping-directions-lake/.

"I, Robot." *Metacritic*, CBS Interactive Inc., www.metacritic.com/movie/i-robot/critic-reviews.

Ibáñez, Gabe, director. *Autómata*. Contracorrientes Films, 2014.

Jonathan R. Tung, Esq. on August 22, 2016 10:57 AM. "Who Owns the Creation of an Artificial Intelligence?" *Technologist*, FindLaw, 22 Aug. 2016, blogs.findlaw.com/technologist/2016/08/who-owns-the-creation-of-an-artificial-intelligence.html.

Jonze, Spike, director. *Her*. Warner Bros, 2013.

Keiper, Adam & Schulman, Ari N., "The Problem with 'Friendly' Artificial Intelligence," *The New Atlantis*, Number 32, Summer 2011, pp. 80-89.

Kleeman, Sophie. “Here Are the Microsoft Twitter Bot's Craziest Racist Rants.” *Gizmodo*, Gizmodo.com, 24 Mar. 2016, gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160.

Leins, Casey. “Elon Musk: Artificial Intelligence Is Society's 'Biggest Risk'.” *U.S. News & World Report*, U.S. News & World Report, 16 July 2017, www.usnews.com/news/national-news/articles/2017-07-16/elon-musk-artificial-intelligence-is-the-biggest-risk-that-we-face-as-a-civilization.

Lem, Stanislaw. *The Cyberiad: Fables for the Cybernetic Age.* Trans. Michael Kandel. New York: Seabury, 1974. Print.

Lewis-Kraus, Gideon. “The Great A.I. Awakening.” *The New York Times*, The New York Times, 14 Dec. 2016, mobile.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Lin, Patrick. "The Ethics of Autonomous Cars." *The Atlantic*. Atlantic Media Company, 08 Oct. 2013. Web. 20 July 2017.

“Media, Platform, and Users.” *Algorithms and Accountability Conference | NYU School of Law*, NYU Law, 28 Feb. 2015, www.law.nyu.edu/centers/ili/AlgorithmsConference.

Miller, Marjorie Mithoff. “The Social Science Fiction of Isaac Asimov.” *Isaac Asimov*, edited by Joseph D. Olander and Martin H. Greenberg, Taplinger Publishing Company, Inc., 1977.

McCarthy, Todd. “I, Robot.” *Variety*, Variety Media, LLC, 16 July 2004, variety.com/2004/film/markets-festivals/i-robot-3-1200532174/.

Olander, Joseph D., and Martin H. Greenberg. *Isaac Asimov*. Taplinger Publishing Company, Inc., 1977.

Orr, Lucy. “I Love You. I Will Kill You! I Want to Make Love to You: The Evolution of AI in Pop Culture.” *The Register®*, Situation Publishing, 29 Jan. 2016, www.theregister.co.uk/2016/01/29/ai_in_tv_film_books_games/.

Patrouch, Joseph H. *The Science Fiction of Isaac Asimov*. Dennis Dobson, 1974.

Price, Rob. “Microsoft Is Deleting Its AI Chatbot's Incredibly Racist Tweets.” *Business Insider*, Business Insider, 24 Mar. 2016, www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3.

Rissland, Edwina L, et al. “AI & Law.” *AI & Law | IAAIL - International Association for Artificial Intelligence and Law*, IAAIL, www.iaail.org/?q=page%2Fai-law.

Rubin, Charles T., "Machine Morality and Human Responsibility," *The New Atlantis*, Number 32, Summer 2011, pp. 58-79.

Sawyer, Robert J. “Editorial: Robot Ethics.” *Science Fiction Writer ROBERT J. SAWYER Hugo and Nebula Winner*, 16 Nov. 2007, www.sfwriter.com/science.htm.

Scherer, Matthew U. “Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies.” *Harvard Journal of Law and Technology*, vol. 29, no. 2, 2016, papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777.

Smith, Agnese. “Artificial Intelligence.” *National*, Canadian Bar Association, 2015, nationalmagazine.ca/Articles/Fall-Issue-2015/Artificial-intelligence.aspx.

Smith, Doug and Kim Takal, directors. *Robots*. Eastman Kodak Company, 1988.

“Sophia - the Latest Robot from Hanson Robotics.” *Sophia AI*, Hanson Robotics Ltd., 2017, sophiabot.com/.

Statt, Nick. “Artificial Intelligence Experts Sign Open Letter to Protect Mankind from Machines.” *CNET*, CBS Interactive Inc., 11 Jan. 2015, www.cnet.com/news/artificial-intelligence-experts-sign-open-letter-to-protect-mankind-from-machines/.

Thomson, Desson. “Will Smith's Robot Jackpot .” *The Washington Post*, WP Company, 16 July 2004, www.washingtonpost.com/wp-dyn/articles/A51838-2004Jul15.html.

Titcomb, James. “Stephen Hawking Says Artificial Intelligence Could Be Humanity's Greatest Disaster.” *The Telegraph*, Telegraph Media Group, 19 Oct. 2016, www.telegraph.co.uk/technology/2016/10/19/stephen-hawking-says-artificial-intelligence-could-be-humanitys/.

United States, Congress, Subcommittee on Machine Learning and Artificial Intelligence. “Preparing for the Future of Artificial Intelligence.” *Preparing for the Future of Artificial Intelligence*.

“US Politicians Call for ‘Future of AI Act’, May Shape Legal Factors.” *Artificial Lawyer*, Artificial Lawyer, 18 Dec. 2017,

www.artificiallawyer.com/2017/12/18/us-politicians-call-for-future-of-ai-act-may-shape-legal-factors/.

U.S. Sen. Roger Wicker. "Digital Decision-Making: The Building Blocks of Machine Learning and Artificial Intelligence." *U.S. Senate Committee On Commerce, Science, & Transportation*, Committee on Commerce, Science, and Transportation, 12 Dec. 2017, www.commerce.senate.gov/public/index.cfm/2017/12/digital-decision-making-the-building-blocks-of-machine-learning-and-artificial-intelligence.

Villeneuve, Dennis, dir. *BladeRunner 2049*. Warner Bros, 2017.

Vintar, Jeff, and Akiva Goldsman. *I, Robot*. 20th Century Fox, 2004.

Warrick, Patricia S. "Ethical Evolving Artificial Intelligence: Asimov's Computers and Robots." *Isaac Asimov*, edited by Joseph D. Olander and Martin H. Greenberg, Taplinger Publishing Company, Inc., 1977.

"We Bring Robots to Life." *Hanson Robotics* , Hanson Robotics Ltd., 2017, www.hansonrobotics.com/.

Weaver, John Frank. "We Need to Pass Legislation on Artificial Intelligence Early and Often." *Slate Magazine*, The Slate Group, 12 Sept. 2014, www.slate.com/blogs/future_tense/2014/09/12/we_need_to_pass_artificial_intelligence_laws_early_and_often.html.

Weller, Chris. "Meet the First-Ever Robot Citizen - a Humanoid Named Sophia That Once Said It Would 'Destroy Humans'." *Business Insider*, Business Insider, 27 Oct. 2017, www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10/#the-idea-of-fooling-humans-is-not-necessarily-the-goal-hanson-told-business-insider-4.

"Your Partner for a Cleaner Home." *IRobot*, www.irobot.com/.