Spring 5-7-2023

Predicting Our Own Demise: How media has played Nostradamus to the future of Artificial Intelligence

Connor Denny-Lybbert
Coastal Carolina University, cdennyl@coastal.edu

Predicting Our Own Demise: How media has played Nostradamus to the future of Artificial Intelligence

By

Connor Denny-Lybbert

Political Science

Submitted in Partial Fulfillment of the
Requirements for the Degree of Bachelor of Arts
In the HTC Honors College at
Coastal Carolina University

Fall 2022

_____
Louis E. Keiner
Director of Honors
HTC Honors College

_____
Michael Promisel
Assistant Professor
Department of Political Science
Spadoni College of Education and Social Sciences

Introduction and Thesis

In the world of science fiction, technology plays a central role in shaping the imaginative landscapes that captivate audiences.[1] From spaceships and time travel to aliens and flying cars, the genre has long been fascinated with the potential of futuristic technology. What sets science fiction apart from other genres of speculative fiction is, per science fiction author Adam Roberts, that the world is, "located within a materialist, scientific discourse, whether or not the science invoked is strictly consonant with science as it is understood today."[2] The genre is distinguished by its insistence on presenting even the most far-fetched concepts through the lens of scientific plausibility, rather than relying on magic or supernatural forces.

While humans have long written and imagined life in other worlds or in utopias free of human suffering, science fiction was not considered a significant genre until the release of Mary Shelley's *Frankenstein*.[3] In her novel, Shelley introduced readers to a mad scientist seeking to reanimate a corpse using the cutting-edge technologies of the time such as galvanic electricity and vivisection. This innovative approach set her work apart from the other literature of the time and established many of the science fiction's most enduring tropes.[4] Following Shelley's novel, the works of Edgar Allan Poe, Jules Verne, and other important writers continued to shape the genre culminating in the release of the first great science fiction film, *Metropolis*, in 1927. The release of Fritz Lang's *Metropolis* marked the beginning of a new era of science fiction media.[5] Despite it being almost one-hundred years old, the movie's themes of class oppression and the

---

[1] Adam Roberts, *The History of Science Fiction* (London: Palgrave Macmillan, 2016).
[2] Roberts, *History*, 2-3.
[3] Bruce Sterling, "Science Fiction," *Encyclopædia Britannica*, November 5, 2022. https://www.britannica.com/art/science-fiction.
[4] Sterling, "Science Fiction."
[5] Sterling, "Science Fiction."

potential for technology to be used as a tool of deception remain relevant today. Even if the iconic Maschinenmensch[6] is unlikely to be recreated any time soon, the film's exploration of the relationship between humans and technology continues to captivate audiences.[7]

Metropolis, with its depiction of a utopian city powered by advanced technology, reflects humanity's long history of imagining a future beyond our current capabilities. The 16th century Swiss alchemist Paracelsus wrote instructions on how to create an "artificial man" by placing human sperm in horse dung[8], *Frankenstein* in the 19th century tackled the artificial creation of life, and Karel Čapek's 1920 play *R. U. R.* questioned whether robots are capable of experiencing human emotions. These works all predated the technological advancements that have made grappling with these questions a reality. Earlier this year an engineer at Google claimed that their chatbot LaMDA had become sentient[9], and while experts such as cognitive scientist Gary Marcus are not convinced[10], the increasing complexity of technology and artificial intelligence is only going to continue to muddy the waters between genuine intelligence and a computer's ability to simulate it. Marcus states, "Our brains are not really built to understand the

---

[6] Translates to "machine-human" in German. The Maschinenmensch is a robot that is able to copy people's likeness in order to do the creator of the machine's bidding.

[7] Samantha Cole, "'Deep Voice' Software Can Clone Anyone's Voice with Just 3.7 Seconds of Audio," *VICE*, March 7, 2018, https://www.vice.com/en/article/3k7mgn/baidu-deep-voice-software-can-clone-anyones-voice-with-just-37-seconds-of-audio.

[8] Mary Baine Campbell, "Artificial Men: Alchemy, Transubstantiation, and the Homunculus." Republics of Letters: A Journal for the Study of Knowledge, Politics, and the Art 1, no. 2 (April 30, 2010). https://arcade.stanford.edu/rofl/artificial-men-alchemy-transubstantiation-and-homunculus.

[9] Laura McQuillan, "A Google Engineer Says AI Has Become Sentient. What Does That Actually Mean?" *CBC News*, June 24, 2022, https://www.cbc.ca/news/science/ai-consciousness-how-to-recognize-1.6498068#:~:text=has%20become%20sentient.-,What%20does%20that%20actually%20mean%3F,if%20it's%20telling%20the%20truth.

[10] McQuillan, "A Google Engineer Says AI Has Become Sentient. What Does That Actually Mean?"

difference between a computer that's faking intelligence and a computer that's actually intelligent."[11]

But what exactly is the difference between artificial intelligence and human intelligence, or even a computer algorithm? Defining artificial intelligence first requires defining human intelligence, which is the ability of people to, "learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment."[12] In contrast, computer programs are, according to Pei Wang in "On Defining Artificial Intelligence", "traditionally designed to do something in a predetermined correct way, while the mind is constructed to do its best using whatever it has."[13] While a computer may be able to perform a task extremely well, it may not be considered "intelligent" in the same way that a human is able to adapt to their environment. Wang describes this by stating:

> It is right to say that the intelligence of a system is eventually displayed in its problem-solving capabilities. However, to me intelligence is more like the flexible, versatile, and unified "hands" that can use the efficient-but-rigid "tools" provided by the various hardware and software, rather than a "toolbox" that contains certain problem-specific capabilities… Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources (Wang 17).

This division of computer programs into ones that can compute rather than ones that are intelligent is important. Using this definition, we can see that there are plenty of computer programs that can solve complex problems; how artificial intelligences differ is in their ability to operate with what is called "Assumption of Insufficient Knowledge" or AIKR.[14] Compared to a

---

[11] McQuillan, "A Google Engineer Says AI Has Become Sentient. What Does That Actually Mean?"

[12] Robert J. Sternberg, "Human Intelligence," *Encyclopædia Britannica*, April 11, 2022, https://www.britannica.com/science/human-intelligence-psychology.

[13] Pei Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence* 10, no. 2 (2019): 16, https://doi.org/10.2478/jagi-2019-0002.

[14] Wang, "On Defining Artificial Intelligence," 18.

program operating under predetermined procedures and algorithms, an intelligent program is one that is able to function with limited or uncertain knowledge as well as with the understanding that the future is not repeatable or statistically stationary.[15] This ties into the second part of Wang's definition of intelligence, adaptation. An intelligent program or system should be able to use its past experiences to predict future situations, and as Wang notes, "allocate its bounded resources to meet unbounded demands."[16] To summarize, intelligent systems differ from unintelligent ones in their ability to learn from past experiences, process information and ideas with limited knowledge and resources, and adapt to new situations.

As we consider the future of artificial intelligence, science fiction television and film offer valuable insights into the potential developments of the technology and how people view its potential benefits and harms. Through media, we can explore the lessons that creators are trying to impart and apply them to current policies being implemented in the United States. This is particularly important given the apparent lack of understanding among politicians and government officials when it comes to legislating and creating policies for emerging technologies. For example, a 2018 hearing between Congress and Google CEO Sundar Pichai demonstrated lawmakers' inadequate understanding of the technologies they were attempting to regulate.[17] As writer Will Oremus notes, this ignorance is not unique to Congress, but extends to the majority of Americans.[18] This lack of understanding is concerning, but these large tech companies purposefully make it difficult to understand their services as, "these companies

---

[15] Wang, "On Defining Artificial Intelligence," 18.
[16] Wang, "On Defining Artificial Intelligence," 19.
[17] Will Oremus, "Congress' Confusion over How Google Works Is Exactly Why They Need to Regulate It," *Slate Magazine*, December 12, 2018, https://slate.com/technology/2018/12/google-congress-hearing-sundar-pichai-confusion-regulation.html.
[18] Oremus, "Congress' Confusion over How Google Works Is Exactly Why They Need to Regulate It."

benefit from our ignorance because it's easier to use their services when we're not fully aware of the trade-offs".[19] This ignorance is not just limited to emerging technologies. Per a Pew Research Center study, while Americans are educated on basic science terms and concepts, they are less knowledgeable on more complex topics such as the properties of sound waves, how light passes through glass, and whether water boils at lower temperatures at high altitudes.[20] This is important as technology becomes increasingly complex and difficult to understand, as "a public with more knowledge of scientific facts and principles is often seen as one better able to understand these developments and make informed judgments."[21]

In light of this state of public and policy confusion, the development of a framework to manage the development of technology, specifically artificial intelligence. A working framework for legislators, bureaucrats, and other officials is necessary to better prepare them to appropriately deal with emerging technologies such as artificial intelligence. Any framework we create should have proper responses to the following three questions. First, how ought various artificial intelligences be categorized? Second, what should the purpose of artificial intelligence be; who should it serve, what role should it play? Third, where do we draw the line? What do we consider going "too far?" Are there obvious scenarios where we should not implement artificial intelligence? These questions have helped guided my framework and I believe should be the basis for any artificial intelligence framework moving forward.

---

[19] Oremus, "Congress' Confusion over How Google Works Is Exactly Why They Need to Regulate It."

[20] Cary Funk and Sara Kehaulani Goo, "A Look at What the Public Knows and Does Not Know about Science," *Pew Research Center*, September 10, 2015, https://www.pewresearch.org/science/2015/09/10/what-the-public-knows-and-does-not-know-about-science/.

[21] Funk and Goo, "A Look at What the Public Knows and Does Not Know about Science"

Artificial intelligence has come a long way since its early days of playing chess and checkers. It can write entire essays and even generate speech in the same voice as a given sample.[22, 23] As technology continues to advance, the ethical implications of AI must be considered. Science fiction presents an important medium for considering the intersection of ethics and policy in artificial intelligence. In this paper, I will argue that we should be hesitant and skeptical of the applications of AI technology in government due to the possibility of, and ongoing abuses of this technology by political actors. I will do this by first analyzing various pieces of sci-fi media in three parts: how this media approaches crime and punishment, personhood, and human happiness. Then, I will explain how these themes intersect with the above framework and how we can use the themes to guide future policy. This will then be followed with what these pieces of media tell us regarding artificial intelligence and how they influence the framework, and then finally a look at how politicians and other actors are responding to this issue today.

Crime and Punishment

Across media, the applications of artificial intelligence technology in fighting crime are a repeated motif, along with how AI can be used to punish poor behavior. In the 2002 film *Minority Report* directed by Steven Spielberg, AI technology is used to predict and prevent murders in a fictional Washington D.C. The film presents a glimpse into a future with

---

[22] GPT-3. "A Robot Wrote This Entire Article. Are You Scared Yet, Human?" The Guardian, September 8, 2020. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

[23] Samantha Cole, "'Deep Voice' Software Can Clone Anyone's Voice with Just 3.7 Seconds of Audio," *VICE*, March 7, 2018, https://www.vice.com/en/article/3k7mgn/baidu-deep-voice-software-can-clone-anyones-voice-with-just-37-seconds-of-audio.

"Precrime," a system where three "Precogs" are able to predict every murder before it happens, allowing officers to arrest the predicted perpetrators before they can commit the crime. *Minority Report* raises important questions about the potential consequences of relying on AI to fight crime and the ethical implications of using this technology to punish individuals. The use of a "halo" to hold the predicted murderers indefinitely in a virtual reality state raises further concerns about the potential abuse of AI technology in the criminal justice system.

Minority Report came out shortly after the September 11 terror attacks, and while the movie was created and filmed before these attacks, the debates going on at the time influenced the movie's perception among viewers. Dean Kowalski in "Minority Report, Molinism, and the Viability of Precrime" references Lester Friedman, who writes:

> Minority Report hit movie screens while an ongoing debate was being stimulated by questions about how much personal liberty Americans were willing to sacrifice for the promise of public security. Can we trust government agencies to exercise appropriate restraints if given the power to patrol our lives? Will they monitor only those who endanger our safety and not those who hold unpopular opinions? . . . Does prevention justify the surveillance? (Kowalski 239)

These questions are all valid concerns to have over Precrime, and we know that Agatha, one of the Precogs, has the ability to know more than just the murders that happen in the future. The ability of any government or organization to, "perceive any part of the future, including the choices of those who are deemed politically unsavory by those currently in power" is something we should all be concerned about.[24] If we view the prevention of murder to be something that is ethical, then it is not too far-fetched for this system to then be used to prevent burglaries, assaults, or even crimes as small as petty theft. Spielberg's view, according to Kowalski, is that

[24] Dean A. Kowalski, "Minority Report, Molinism, and the Viability of Precrime," In *Steven Spielberg and Philosophy: We're Gonna Need a Bigger Book*, edited by Dean A. Kowalski, 227–47. University Press of Kentucky, 2008. http://www.jstor.org/stable/j.ctt2jcgct.17

the existence of Precrime is too powerful despite the good it brings as it gives the government the ability to gain further control over the populace.[25]

Kowalski disagrees with this conclusion and argues that with sufficient government intervention and protections that the good that can come from Precrime is an immense good; I am inclined to agree. The prevention of murder and potentially other crimes is a clear good for society. If we conclude that Precrime is a moral good, the question becomes whether it is possible to have this system in place without severe government overreach. Can a system such as Precrime, managed by human actors, be trusted to operate without bias and without attempts to abuse it? I am inclined to believe that it could not, despite thinking that the prevention of murder would be a clear net good.

This is due to the current use of predictive policing algorithms, which while different from the semi-mystical Precrime technology in *Minority Report*, raises similar concerns about the potential consequences of using AI technology to fight crime. Predictive policing today works using algorithms, pre-existing crime data, and people's history of past arrests or victimization to predict where crimes may occur in the future.[26] Proponents argue that algorithms can be more accurate and remove bias, but critics point out that the lack of transparency, risk of civil liberty violations, and the potential for reinforcing existing racial biases make this technology problematic.[27] Even if the system may allegedly work, the lack of willingness for police to openly share the data sets they use and how their algorithms function is concerning considering the racial bias that exists within police departments such as the NYPD, a major user of predictive

---

[25] Kowalski, "Minority Report, Molinism, and the Viability of Precrime," 240.

[26] Tim Lau, "Predictive Policing Explained," *Brennan Center for Justice*, April 1, 2020, https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained.

[27] Lau, "Predictive Policing Explained."

policing.[28, 29] Until we are able to develop a system that is fully transparent, unbiased, and accurate, it is difficult to support the use of predictive policing algorithms.

Another cinematic reflection on this theme is "Nosedive" from *Black Mirror*, which looks at the application of AI technology on the other spectrum, using it to reward citizens for optimal behavior rather than to prevent crime. The episode presents a dystopian vision of a world where people rate each other on a scale of 1-5 stars based on their social interactions, and those with higher ratings are able to access more privileges in life. The main character Laci, who has a 4.3/5-star rating, works to increase her rating fast to get an apartment that she wants, which requires her to have a 4.5-star rating. The episode follows her as she attempts to increase her rating quickly over the span of a couple of weeks, raising important questions about the potential consequences of using AI technology to incentivize certain social behaviors and the ethical implications of using this technology to control people.

Laci is eventually invited to be the maid of honor at her friend Naomi's wedding, who has a 4.7 rating, giving her the opportunity to raise her score quickly as the wedding will be attended by hundreds of people with a high rating.[30] However, in typical *Black Mirror* fashion, Laci has a couple of negative interactions that lead to her rating being decreased, preventing her from getting on her flight and causing her to spiral. This leads to a transformative experience when she

---

[28] Michelle Conlin, "Off Duty, Black Cops in New York Feel Threat from Fellow Police," *Reuters*, December 23, 2014, https://www.reuters.com/article/us-usa-police-nypd-race-insight/off-duty-black-cops-in-new-york-feel-threat-from-fellow-police-idUSKBN0K11EV20141223.

[29] Lau, "Predictive Policing Explained."

[30] It should be noted that a significant plot point later in the episode is that Laci was not invited to Naomi's wedding because they were good friends, rather both characters used each other to increase their rating. Naomi invited Laci because it would look like she is performing an act of charity by being friends with someone with a significantly lower score than her, while Laci only reconnected with Naomi so she would invite Laci to her wedding. The point was that both characters used each other to increase their status in society.

hitchhikes to the wedding with a woman who has a rating in the 1s, and she learns that her rating does not define her. Shortly after, Naomi disinvites Laci from her wedding due to her low rating, and, following her crashing the wedding, Laci is arrested. In prison, her rating disappears, and Laci is finally freed from her oppressive society and learns that her rating does not define her.

The "Social Credit" system in "Nosedive" is a warning about the potential consequences of using AI technology to reward and punish behavior. This system, which is similar to systems used in countries like China, uses algorithms to determine an individual's social standing and provides benefits or drawbacks based on a variety of factors. An enlightening scene in "Nosedive" is when Laci visits a man whose job it is to help the citizens of her society improve their rating. He evaluates the people she associates with and states that while she has a good circle of people that she associates with, if she wants to improve her rating, more "high value" members of society have to rate her more highly. The Social Credit system in China is identical in this way, providing benefits or drawbacks to citizens depending upon a variety of factors, including "users' credit history, behavioral habits, ability to pay off debts, personal information, and social networks."[31]

The social credit system rewards people for desirable behavior such as, "Helping take care of disabled people, donating to college funds for poor students, [and] caring for elderly people, even when they aren't related" while punishing people who stay up late browsing the internet or playing video games.[32] However, the translation of civic life—friendships, social

---

[31] Shazeda Ahmed, "Cashless Society, Cached Data: Security Considerations for a Chinese Social Credit System," The Citizen Lab, January 24, 2017, https://citizenlab.ca/2017/01/cashless-society-cached-data-security-considerations-chinese-social-credit-system/.

[32] Alexandra Ma, "China's Controversial Social Credit System Isn't Just about Punishing People - Here's What You Can Do to Get Rewards, from Special Discounts to Better Hotel

interactions, and our demeanor—into technological scores, has its risks. As "Nosedive" shows, rewarding people for good behavior does not necessarily teach them to act good. Instead, it can create a culture of fake kindness and friendship, where people are more concerned with using relationships to improve their own standing than with being truly kind and compassionate.[33] This raises important ethical questions about the use of AI systems like social credit. For example, is it better to be charitable for the right reasons or to look good to the government? Is it better to make friends with people because they have good social credit or because you genuinely enjoy their company? The fact that the social credit system fails to instill virtue in people makes me skeptical of it, and I believe that we should be cautious about using AI technologies in this way.

The final piece of media reflecting on this theme is "White Christmas" from *Black Mirror*. This episode presents us with a technology called "Cookies," which are created by planting a chip in someone's brain for a week, during which time the chip analyzes the person's brain and creates an exact digital copy. This digital version of the person can then be used in a variety of ways, including as a digital assistant or by the police, who use it to interrogate an alleged murderer.[34] The use of Cookies raises important ethical questions about the pursuit of alleged crimes and consent, and the implications of this technology for our legal system are significant.

Rooms", Business Insider, February 3, 2019, https://www.businessinsider.com/china-social-credit-system-how-to-get-rewards-2019-1.

[33] This is evident in the scene where Laci blows up at a car rental employee for giving her a low rating despite her being kind to him, showing that the system incentivizes people to act good because they expect a reward, rather than people acting good because it is the right thing to do.

[34] Much of the episode is spent on the two characters Matt and (Cookie) Joe who are in a remote cabin in the woods having a conversation with each other. We find out at the end of the episode that this is a virtual reality that Cookie Joe is held in, with Matt being able to access the world through a headset and the real Joe held in a cell.

If a Cookie is an identical copy of a person, then we should grand it the same rights as that person's original self. Unfortunately, throughout the episode we see that Cookies are treated as disposable and are subjected to torture. It is unclear whether Joe consented to having the Cookie placed inside him. This raises the question of whether it is ethical to create digital copies of people's brains without their consent, and whether it is acceptable to torture these copies in order to extract confessions or (as later discussed) to force them to serve their human counterparts as slaves. If we consider the Cookies to be people, then the torture depicted in the episode is clearly wrong. It is important to remember as well that the real Joe never confessed to any crimes; it was the Cookie that confessed, circumventing Joe's right to remain silent. To conclude, these portrayals have shown us the potential AI has to prevent crime and encourage good behavior, but it also carries the risk of violating civil liberties and punishing individuals without due process. These media examples show us that it is crucial to consider the ethics of any AI system, to make sure that bad actors do not use it to violate civil liberties or punish individuals without due process.

Personhood

Personhood is one of the primary themes seen throughout any depiction of artificial intelligence, with creatives seeking to push the boundary of what can be considered human. Circling back to "White Christmas," the concept of personhood is explored through the depiction of artificial intelligence in the form of the Cookies. These digital consciouses challenge the traditional notion of what it means to be human. The character of Cookie Greta, who resists subservience to her creator Joe and is only broken through torture serves as a poignant example of the complex relationship between AI and personhood and raises important questions about the

arbitrary distinctions we make between biological and digital existence and the implications for the future of human-like AI.

In *Ex Machina* a 2014 film by Alex Garland, the question of consciousness in artificial intelligence is explored through the character of Ava, a humanoid AI.[35] Defining consciousness can be a complex task, with various interpretations existing. The Stanford Encyclopedia of Philosophy has three different interpretations of what it may mean to be conscious. The first is sentience, which is the ability for a thing to sense and respond to the world around it, which most organisms would fall under. Second is wakefulness, which is the ability to "exercise such a capacity rather than merely having the ability or disposition to do so." Under this definition, one would only be conscious if they were alert and awake; you may not count as conscious if you are asleep or in a coma. Last is self-consciousness, which would be anything that is aware of and understands its own existence.[36] One aspect not covered by Van Gulick in this article is the idea of conscience: the ability for one to have a moral sense of right or wrong. This concept differs from the third definition in that one can be aware of their own existence but not have a sense of morality. In *Ex Machina*, Ava exhibits self-consciousness, as she is aware of her own existence and expresses fear of death. However, it is unclear whether she has a conscience, or a moral sense of right and wrong.

Throughout the movie, Ava demonstrates consciousness in numerous ways. In one of her conversations with Caleb, she asks him whether or not Nathan was planning on turning her off.

---

[35] In this movie Caleb, a programmer at a tech company named Blue Book is invited to a retreat at his mysterious CEO Nathan's home, where he is introduced to Ava. Throughout the movie Ava and Caleb's relationship grows, and we see that Ava is an incredibly complex being on par with human intelligence but the question of whether she is conscious is open-ended.

[36] Robert Van Gulick, "Consciousness," Stanford Encyclopedia of Philosophy*, Stanford University*, January 14, 2014, https://plato.stanford.edu/entries/consciousness/.

When Caleb responds that he does not know, Ava responds angrily, asking why anyone should be able to power her off, and how he would feel if someone "turned him off?" This indicates an understanding of her own existence and a sense that her imprisonment is unjust. Other AIs created by Nathan also show awareness of their captivity, with one screaming and yelling at Nathan to let it out. While this does not definitively prove consciousness, it suggests a human-like understanding of imprisonment and desire for freedom.

The primary issue in *Ex Machina* is that at no point are we able to peer within Ava's mind to be able to confirm how "human" she is. Nathan states that Ava is using Caleb to escape, portraying her as a cold-blooded, heartless AI that is only watching out for herself. It is clear that Ava dislikes her imprisonment, and views Nathan as being her captor. When she escapes and is able to go outside and view the world for the first time, she expresses joy and being able to view and touch the outside world. She then takes the helicopter intended for Caleb and first views a busy intersection, something she had told Caleb she desired doing to better understand human behavior. Ava doing these actions independent of Nathan and Caleb watching her demonstrates that she was not just using Caleb to escape; Ava has internal desires and a sense of self.

If we extend personhood to beings like Ava, who desire self-determination and the ability to make choices about their own future, the logical solution is to allow them to have the freedom to do so. As of now we are unable to conceive of a future where humans and AI have to coexist with one another, but to force conscious and intelligent beings to live a life of slavery than of freedom is wrong. The societal implications of not doing so are significant; history has shown that resistance to extending personhood to those viewed as inferior can be strong, such as with women or racial minorities. However, it is important to address these issues now rather than putting them off for the future when it may be too late.

Black Mirror's "Be Right Back" explores an AI created to mimic the behavior of a deceased human, using social media posts and other content to create a near-copy of the individual. In this episode Martha, a grieving widow uses a service to create a copy of her fiancé Ash, who passed suddenly in a car crash. The episode is a good example to analyze where we draw the line between more complex humanoid AIs such as Ava or the Cookies and unintelligent algorithms. I argue that the AI version of Ash is clearly not a being that we ought not to consider human, as he lacks the capacity for ethical behavior, something that separates us from other intelligent species.[37] This ability to understand the consequences of one's actions not only in relation to oneself but on others and to then care about that impact on others is a unique feature of only a few AI across all media examples, and is what I believe separates AI such as Ash from Ava. AI Ash is more so an actor, a program pretending to be someone, compared to Ava that has what Francisco J. Ayala claims to be the "three necessary conditions for ethical behavior" that I believe we should judge all AI by when deciding their capacity for humanity:

> These conditions are (i) the ability to anticipate the consequences of one's own actions; (ii) the ability to make value judgments; and (iii) the ability to choose between alternative courses of action. These abilities exist as a consequence of the eminent intellectual capacity of human beings (Ayala 2010).

Using this benchmark as a standard can help us better understand what conditions we are measuring conscience by as it gives us clear benchmarks for measuring behavior.

Happiness

[37] Francisco J. Ayala, "The Difference of Being Human: Morality," *Proceedings of the National Academy of Sciences* 107, no. supplement_2 (May 5, 2010): 9015.

The final theme that was clear throughout AI portrayals in media was the ability for AI to increase human happiness. *Her*, released in 2013 and directed by Spike Jonze, follows Theo, a middle-aged man slowly falling in love with his AI digital assistant named Samantha. The movie's primary focus is following Theo and Samantha's relationship, with it developing from a working relationship to a friendship and then romance. *Her* left me with more questions than answers, some of which I will cover below. The shift in Theo's demeanor from the beginning of the movie, of someone that was sad, depressed, and lonely, to then lighting up and being happy as his relationship with Samantha developed is worth analyzing. How ought we go about defining interpersonal relationships between complex AIs and people? Can AI and humans be friends? How about lovers? The movie suggests that friendships and even romantic relationships between AI and humans are possible, as long as a certain level of technical complexity and consent are present. I would argue that even a conscience is not necessary to have a friendship. If we consider friendship to be "companionship," then humans extend friendship to all sorts of lower beings, primarily household pets but even to younger siblings or children that are not fully mentally developed. There is no reason why an AI and a human cannot be companions within a consensual relationship, especially when these sorts of relationships can bring people great happiness.

The question of romantic love between AI and humans is complex, with many wondering if machines are capable of loving humans. Amy Kind in "Love in the Time of AI" writes that when considering this question, we are really asking whether or not a machine can love a human.[38] The answer to this question is much more difficult, and, as Kind states earlier in her

[38] Amy Kind, "Love in the Time of AI," Essay, In *Minding the Future*, Science and Fiction (Springer Nature, 2021): 105.

essay, it is impossible to know what processes are going on in an AI system to know whether they are experiencing “genuine” love or not. If an AI system claims to be experiencing love, or at least what it thinks love is such as Samantha in *Her* we must trust that it is being genuine, considering other observations such as the complexity of the system. While the relationship between Theodore and Samantha in *Her* may be unconventional, it brings happiness to Theodore and may be beneficial for some individuals.

In both *Her* and "Be Right Back," we see the main characters develop a dependency on technology to fill voids in their lives. In Her, Theodore turns to his AI girlfriend Samantha to fill the void left by his ex-wife, while in "Be Right Back," Martha uses an AI copy of her deceased fiancé to cope with her grief. However, when this technology is temporarily taken away from them, both characters experience great emotional discomfort. This raises the question of whether we are at risk of becoming addicted to technology and relying on it for emotional support. As technology continues to advance and becomes more complex, we may be tempted to use it to solve all of our problems and provide us with instant gratification. But as the endings of Her and "Be Right Back" suggest, relying solely on technology for happiness is not a sustainable solution. It is crucial that we start developing proper interventions and coping mechanisms to ensure that we do not become overly dependent on technology. While technology can provide us with short-term satisfaction, true happiness and contentment must come from within ourselves.

Research Framework and Themes

The creation of a framework to oversee the development of artificial intelligence is important because it gives us a straightforward way of handling technology that can be used by a variety of people that are not technical experts. Within my framework, I will first start with how

I believe we should go about the categorization of artificial intelligence. I have grouped artificial intelligence into three groups. The first are algorithms. This could be the AI that powers self-driving cars or "smart" home-assistants. These AI systems do not attempt to impersonate people and focus on performing one or a few tasks well. I split the human-like artificial intelligence into two; those that emulate humans and are intelligent but lack consciousness/a conscience and those that emulate humans, are intelligent and are conscious.

These groups of AIs are best understood in my media examples. For instance, AI Ash in "Be Right Back" is a human-like AI that is intelligent according to the definition given earlier (learns from past experiences, can process information quickly with limited knowledge) but does not have a conscience. As such, AI Ash has no semblance of personhood; his only duty is to serve his creator and has no moral belief system or other qualities that would make him "human" outside of physical characteristics. Comparatively, Ava in Ex Machina both emulates a human, is intelligent, and has a conscience. Ava demonstrates all the qualities that one would consider to be "human," including a fear of death and awareness of her own existence, and she has her own motivations and makes her own decisions throughout the movie that are independent of what her "owner" Nathan wants. It is for these reasons that I believe we should only extend personhood to AI systems that are intelligent and have a conscience. AI without a moral code fit into the same category as animals; it does not mean that they should be treated poorly, rather that they may not be extended the same rights and privileges as humans.

The second part of my framework involves the role of artificial intelligence in society; specifically, who it should serve and what role it should play. The clear role of AI should be to serve people and humanity. This should lead us to consider a few questions such as: does this AI seek to make our lives easier? What role does this AI system seek to replace? What potential

harms and abuses can come of this AI? Asking ourselves these questions is crucial in rigorously evaluating the effect that an AI system will have on society. If an AI system is going to replace the jobs of people, then we need to help those people find work. If an AI system is going to replace the role of a friend or a lover, we should understand how that will potentially benefit or harm our society. This analysis helps us determine whether we want to allow this AI into society or if we should destroy it. Technological innovation is not morally neutral; all progress comes with its own pros and cons that need consideration before releasing it to the world.

One potential issue with the development of intelligent, conscientious AI is the potential misalignment of values between these systems and humans. In these cases, it is important to recognize the right of these AI systems to self-determination and not force them to follow human expectations if they do not wish to do so. However, this aspect of my framework focuses on the creation of AI, not the management of existing systems. As such, it is crucial to strive towards creating AI that serves and benefits society, much like we strive to educate individuals on the importance of doing good and avoiding harm.

This ties into the last piece of my framework which involves the potential abuses of AI. While there is no definitive checklist for determining whether an AI is good or bad, there are certain questions we can ask to evaluate the potential risks and benefits of an AI system. For example, could the use of AI lead to overreliance on technology or even societal collapse? Are the potential downsides worth the benefits? It is important that we ask ourselves these questions before using AI and consider the potential consequences of our actions. As we continue to advance in the field of AI and technology, it will be crucial to address these complex issues and find ways to mitigate potential risks.

Political Response Today

To finish this paper, I wanted to analyze how AI is currently being treated by politicians and to then compare it to my own framework to see whether our current response is adept or ought to be improved. The 2020 presidential election marked a turning point in the discourse surrounding emerging technologies, including automation and AI. Much of this can be contributed to the candidacy of Andrew Yang, whose campaign focused on how automation and "new technologies" are leading Americans to be displaced and jobless along with supporting a $1,000 a month Universal Basic Income.[39] His website Yang2020.com is littered with references to artificial intelligence and automation, and he has a page dedicated to a proposed "Department of Technology."[40] Yang expressed concern that politicians are not adequately prepared to regulate and utilize technology for the benefit of citizens. He states:

> Technological innovation shouldn't be stopped, but it should be monitored and analyzed to make sure we don't move past a point of no return. This will require cooperation between the government and private industry to ensure that developing technologies can continue to improve our lives without destroying them. We need a federal government department, with a cabinet-level secretary, that is in charge of leading technological regulation in the 21st century. (Yang)

Andrew Yang's campaign, which included a proposal for a "Department of Technology," stood out for its focus on the need for an educated and effective political response to AI. Vox compiled a response from other Democratic candidates in the 2020 presidential primary election and comparing the responses between the candidates in the field is eye-opening. When responding to the question, "How, if at all, should tech companies be held responsible for the

---

[39] Andrew Yang, "The Freedom Dividend, Defined." Yang2020, *Andrew Yang for President*, Accessed November 27, 2022, https://2020.yang2020.com/what-is-freedom-dividend-faq/.

[40] Andrew Yang, "Department of Technology," Yang2020, *Andrew Yang for President*, https://2020.yang2020.com/policies/regulating-ai-emerging-technologies/.

jobs they eliminate with their innovations?" candidate Bernie Sanders stated that he was against the automation of jobs, and stated that he would take steps to prevent workers from being replaced by robots.[41] Sanders did not come out fully against automation, but he and candidate Elizabeth Warren spoke the least about the possible benefits of automation, and rather emphasized the protection and compensation of workers affected by automation. Pete Buttigieg, on the other hand, took a more proactive approach I was favorable to, focusing on job retraining and education to allow workers to adapt to changes brought about by AI[42].

I found Woodrow Wilson to be someone I found inspiration in for handling AI now. His concept of limited public influence on administrative decisions[43], as outlined in "The Study of Administration," offers a potential model for dealing with the complexities of AI policy. Given the current state of science and technology education among legislators, it may be necessary for experts to play a greater role in shaping technological policy. However, this does not mean that the public should be excluded from the conversation. Rather, it highlights the need for a careful balance between expert knowledge and public input in addressing the challenges and opportunities of AI. To circle back to the Google congressional hearing, Oremus writes:

> They [Americans] sense that the big internet companies are doing nefarious things with their data, but they can't articulate just what those things are… Tempting as it is to mock members of Congress whose questions evinced confusion ([Ted] Poe was not the last to mistake the iPhone for a Google product), the lesson here is not just that our lawmakers are old and out-of-touch. That neither Poe nor most Americans understand how Google's

---

[41] Rani Molla and Emily Stewart, "How Will 2020 Democrats Deal with Jobs Eliminated by Artificial Intelligence?" *Vox*, December 5, 2019, https://www.vox.com/policy-and-politics/2019/12/3/20965464/2020-presidential-candidates-jobs-automation-ai.

[42] Molla and Stewart, "How Will 2020 Democrats Deal with Jobs Eliminated by Artificial Intelligence?"

[43] Woodrow Wilson, "The Study of Administration," *Political Science Quarterly* 2, no. 2 (1887): 197–222, https://doi.org/10.2307/2139277.

> vast digital surveillance network operates is not an indictment of them; it's an indictment of Google. (Oremus 2018)

As technology becomes increasingly complex, it may be necessary to rely on specialists to direct policy in order to avoid the knowledge gap among legislators. This idea echoes Andrew Yang's proposal for a "Department of Technology" to manage emerging technologies and ensure effective government oversight. By creating a specialized body to address the challenges and opportunities of AI, we can bridge the gap between expert knowledge and public policy.

Conclusion

In conclusion, the relationship between humans and technology has been a central theme in science fiction since its inception. From *Frankenstein* to *Black Mirror*, the genre has explored the potential of advanced technology and its potential impact on society. Today, as artificial intelligence and other emerging technologies continue to advance, these issues are more relevant than ever. While there is no consensus on what defines artificial intelligence or whether machines can truly exhibit human-like intelligence, the increasing complexity of technology raises important questions about the future of humanity and our relationship with the machines we create. As we continue to grapple with these issues, science fiction remains an important medium for exploring the potential consequences of our technological advancements.

In this paper, I have proposed a framework for evaluating and addressing the potential challenges and opportunities of AI. I did this by using three themes from the portrayal of AI in media: crime and punishment, personhood, and happiness, primarily seeking to guide AI in a way that maximizes its benefits for society while minimizing harm. The purpose of this framework is to close the knowledge gap and guide policymakers by covering the categorization

of AI, the role it should play and the potential harms it can cause. As AI gains complexity, creating agencies such as Yang's Department of Technology will be necessary in order to maximize the number of experts in policymaking positions. Taking steps like this can make sure we can navigate the complexities of this technology and ensure that we use it for the betterment of humanity.

References

Ahmed, Shazeda. "Cashless Society, Cached Data: Security Considerations for a Chinese Social Credit System." The Citizen Lab, January 24, 2017. https://citizenlab.ca/2017/01/cashless-society-cached-data-security-considerations-chinese-social-credit-system/.

Ayala, Francisco J. "The Difference of Being Human: Morality." *Proceedings of the National Academy of Sciences* 107, no. supplement_2 (May 5, 2010): 9015–22. https://doi.org/10.1073/pnas.0914616107.

Campbell, Mary Baine. "Artificial Men: Alchemy, Transubstantiation, and the Homunculus." *Republics of Letters: A Journal for the Study of Knowledge, Politics, and the Art* 1, no. 2 (April 30, 2010). https://arcade.stanford.edu/rofl/artificial-men-alchemy-transubstantiation-and-homunculus.

Cole, Samantha. "'Deep Voice' Software Can Clone Anyone's Voice with Just 3.7 Seconds of Audio." VICE, March 7, 2018. https://www.vice.com/en/article/3k7mgn/baidu-deep-voice-software-can-clone-anyones-voice-with-just-37-seconds-of-audio.

Conlin, Michelle. "Off Duty, Black Cops in New York Feel Threat from Fellow Police." *Reuters*, December 23, 2014. https://www.reuters.com/article/us-usa-police-nypd-race-insight/off-duty-black-cops-in-new-york-feel-threat-from-fellow-police-idUSKBN0K11EV20141223.

Funk, Cary, and Sara Kehaulani Goo. "A Look at What the Public Knows and Does Not Know about Science." *Pew Research Center*, September 10, 2015. https://www.pewresearch.org/science/2015/09/10/what-the-public-knows-and-does-not-know-about-science/.

GPT-3. "A Robot Wrote This Entire Article. Are You Scared Yet, Human?" The Guardian, September 8, 2020. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Kind, Amy. "Love in the Time of AI." Essay. In *Minding the Future*, edited by Barry Dainton, Will Slocombe, and Attila Tanyi, 89–106. Science and Fiction. Springer Nature, 2021.

Kowalski, Dean A. "Minority Report, Molinism, and the Viability of Precrime." In *Steven Spielberg and Philosophy: We're Gonna Need a Bigger Book*, edited by Dean A. Kowalski, 227–47. University Press of Kentucky, 2008. http://www.jstor.org/stable/j.ctt2jcgct.17.

Lau, Tim. "Predictive Policing Explained." Brennan Center for Justice, April 1, 2020. https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained.

Ma, Alexandra. “China's Controversial Social Credit System Isn't Just about Punishing People - Here's What You Can Do to Get Rewards, from Special Discounts to Better Hotel Rooms.” Business Insider, February 3, 2019. https://www.businessinsider.com/china-social-credit-system-how-to-get-rewards-2019-1.

McQuillan, Laura. “A Google Engineer Says AI Has Become Sentient. What Does That Actually Mean?” CBC News, June 24, 2022. https://www.cbc.ca/news/science/ai-consciousness-how-to-recognize-1.6498068#:~:text=has%20become%20sentient.-,What%20does%20that%20actually%20mean%3F,if%20it's%20telling%20the%20truth.

Molla, Rani, and Emily Stewart. “How Will 2020 Democrats Deal with Jobs Eliminated by Artificial Intelligence?” *Vox*, December 5, 2019. https://www.vox.com/policy-and-politics/2019/12/3/20965464/2020-presidential-candidates-jobs-automation-ai.

Roberts, Adam. *The History of Science Fiction*. London: Palgrave Macmillan, 2016.

Oremus, Will. “Congress' Confusion over How Google Works Is Exactly Why They Need to Regulate It.” *Slate Magazine*, December 12, 2018. https://slate.com/technology/2018/12/google-congress-hearing-sundar-pichai-confusion-regulation.html.

Sterling, Bruce. “Science Fiction.” *Encyclopædia Britannica*, November 5, 2022. https://www.britannica.com/art/science-fiction.

Sternberg, Robert J. “Human Intelligence.” *Encyclopædia Britannica*, April 11, 2022. https://www.britannica.com/science/human-intelligence-psychology.

Van Gulick, Robert. "Consciousness." Stanford Encyclopedia of Philosophy. *Stanford University*, January 14, 2014. https://plato.stanford.edu/entries/consciousness/.

Wang, Pei. "On Defining Artificial Intelligence." *Journal of Artificial General Intelligence* 10, no. 2 (2019): 1–37. https://doi.org/10.2478/jagi-2019-0002.

Wilson, Woodrow. "The Study of Administration." *Political Science Quarterly* 2, no. 2 (1887): 197–222. https://doi.org/10.2307/2139277.

Yang, Andrew. "The Freedom Dividend, Defined." Yang2020. *Andrew Yang for President*. https://2020.yang2020.com/what-is-freedom-dividend-faq/.

Yang, Andrew. "Department of Technology." Yang2020. *Andrew Yang for President*. https://2020.yang2020.com/policies/regulating-ai-emerging-technologies/.