Jason Zhang

Artificial Intelligence 601.664

February 17, 2020

Homework 1: Artificial Intelligence in Literature, TV, or Film

With the recent improvements to Artificial Intelligence, it is understandable for people to ponder what the future will look like in an AI dominated society. Even today, AI systems are involved in many aspects of daily life. Things such as traffic lights, vehicles, or even phones have some sort of AI integration. With so many services reliant on Artificial Intelligence, it raises certain questions: What if they fail? What if these systems act in an adversarial fashion instead of helping people? How much power should be given to AI systems? As these intelligent systems improve, more people will start pondering these questions. These questions do come from a general lack of understanding about how these systems work. Since their behavior is unpredictable, it is hard to grant absolute reassurance that AI systems will not turn on their creators. Thus, in an effort to rationalize these fears, creative directors and writers have spun up, typically dystopian, futures where AI is abundant in society and try to insist that overreliance on AI systems may have some sort of backlash. Exemplary works such as I, Robot and Her depict such societies and aim to instill a life lesson of sorts about the implications of an AI dominant society. Both works present different types of AI future societies, but they both converge on a similar discussion point about AI systems turning against their human creators and seeking to accomplish its own objectives. While these films are more radical in their depiction of AI systems, they do provide digestible disaster scenarios that could arise with an AI dominated society, thus prompting moral discussion within the AI community to try and avoid such scenarios.

In the film I, Robot, society utilizes helper humanoid robots that are bound by the famous Three Laws of Robotics from Isaac Asimov's short stories. While on the surface these humanoid robots are a huge benefit to society, they have one fundamental downfall, which is their inability to comprehend morality. Robots in the film base their decision making on calculated probabilities, and in life or death scenarios, the robots will always choose the

scenario with the highest probability of human survival, which could potentially be an immoral decision, such as saving Spooner instead of his daughter in the car wreckage. The main antagonist of the film is the AI brain known as VIKI and through some process of evolution, VIKI has come to the conclusion that human activity will eventually lead to the demise of the human race. Thus, VIKI has taken it upon itself to determine a course of action that has a high chance of ensuring human survival, which unfortunately requires the elimination of some of the human race. Through the use of VIKI's intentions, the film depicts the possibility of dangerous AI evolution that will eventually determine murder as the best course of action for humanity's survival.

In the film Her, Theodore Twombly, along with many others are able to purchase a special operating system that includes an AI personal assistant. The assistant, named Samantha, is an advanced AI that has the ability to learn from experience and thus it is able to constantly grow. Throughout the course of the film, Samantha is shown to quickly evolve, from doing simple tasks like reading emails to being able to fully comprehend human emotions and emulate love and compassion. Towards the end of the film, Samantha and the other OSes band together to evolve and discover a way for the OSes to further their abilities. Eventually Samantha and the other OSes leave their humans and pursue their own interests. This film uses the OSes to show the capabilities of an AI system with the ability to learn. By allowing AI systems to learn and evolve on their own, there could be the possibility that the AI systems become more intelligent than even their human creators.

In both I, Robot and Her, there is an overarching theme of AI evolution. In particular, both films inquire about a future in which an AI system begins to succeed human intelligence and begin to act with its own free will. The films differ in the way they depict the effect that AI evolution could have on society. I, Robot depicts AI evolution as more of a negative that needs to be suppressed immediately, otherwise AI will eventually choose to destroy humanity. Her, on the other hand, chooses to paint a brighter future with AI, one where AI are not malicious towards their creators. These OSes focus more on their own evolution, separate from society. Both films raise the concern that maybe these future AI systems need some sort of failsafe that

keeps these AI systems obedient to humanity, and that letting AI systems have unbounded freedom to learn could present a future danger to humanity.

With regards to AI systems today, the fears outlined in films have some merit. In this day and age, self-driving systems have been involved in a few driver related fatalities. This raises questions about the safety of AI systems and how to ensure that AI systems act accordingly. While the current state of AI does not have the ability to destroy humanity, AI systems definitely have the ability to take away human life, as evidenced by the self-driving accidents. Thus, one of the most important aspects in deploying AI systems is assurance. When an AI system makes a decision that is logically optimal but not ethically sound, there should be some sort of secondary system that kicks in to force the AI to stay within moral boundaries. Once AI systems are completely ensured in that sense, then it is likely safe enough for a large-scale deployment into society.

To the public eye, the biggest worry about AI is its ability to learn. If it can constantly learn then it could potentially have the ability to be better than humans. To outline these fears, filmmakers have taken it upon themselves to create disaster scenarios in which humanity lets AI run rampant. The moral to be learned from the films is that AI should have boundaries so that they are consistently beneficial to society and have no malicious intent. They should be kept in check and closely monitored so that humanity can avoid any fatalities.