Department of Philosophy and Religion

Are Androids Free? Sci-Fi, A.I, and the Free Will Debate.

A Senior Thesis Submitted to the
Department of Philosophy and Religion in
Partial Fulfillment of the Requirements
For the Bachelor of Arts Degree at
Washington College

Patrick T. Salerno

Spring 2021

I pledge my word of honor that I have abided by the Washington College Honor Code while completing this assignment.

Table of Contents

Introduction 2

I. A Personal Aside 3

Chapter One: Overview of Free Will and A.I 5

I. Chapter Introduction 5
II. Free Will, A Primer 5
III. Artificial Intelligence 8
IV. Functionalism, A Cross-section of AI and Free Will 13
V. Religious Objections to A.I. 15
VI. Chapter Summary 16

Chapter Two: Functionalism and Intelligence 17

I. Chapter Introduction 17
II. Expanding Our Conception of Intelligence 17
III. Conditions of Intelligence 21
IV. Chapter Summary 24

Chapter Three: Artificial Intelligence and Freedom 25

I. Chapter Introduction 25
II. Human Freedom and Functionalism 25
III. A.I and Compatibilism 27
IV. Chapter Summary 31

Chapter Four Moral Significance 32

I. Chapter Introduction 32
II. Commander Data and Understanding Difference 33
III. The Emergency Medical Hologram. 36
IV. L3-37 38
V. Chapter Summary 39

Conclusion: 40
Bibliography 42

Introduction

This discussion will explore what Artificial Intelligence (AI) is and argue for the perspective that these machines will one day possess freedom of the will. Furthermore, this discussion will argue that humanity has an obligation to treat these machines as morally responsible agents. I would like to show that science fiction has informed and should continue to inform the conversations we have today about this technological advancement. The questions this paper will be addressing are: 1) under what circumstances can machines be said to have intelligent thought; 2) does having intelligent thought mean they have freedom and moral responsibilities; and 3) do we as humans have any moral obligations when these machines reach said point?

In Chapter One, I will define and explain the important terms and arguments surrounding AI and free will. This will include detailing the current advancement of Artificial Intelligence, exploring how these advancements are bringing AI ever closer to human intelligence, and exploring the significance this problem has on our society and how Science Fiction, specifically *Star Trek* and *Star Wars*, can aid our understanding of said significance. In Chapters Two and Three, I will develop my argument on the topic, focusing on the topics of compatibilism and functionalism. I will argue that the validity of both these theories, and the natural consequence of their acceptance, is the recognition of the eventual free will of Artificial Intelligence. A compatibilist understanding of free will states that it is possible for actions to be determined and still be free. Functionalism is a theory of the mind which argues that mental states be defined by the role they play within a given system and not by its internal constitution. Finally, in Chapter Four, I will turn to the question of significance. I will use Star Trek's Lieutenant Commander Data and Emergency Medical Hologram and Star War's L3-37 from *Solo: A Star Wars Story* to examine the social, moral, and political consequences of A.I. and will harken each example back to our own society. These stories will highlight the dangers of refusing to grant autonomous entities moral

status due to perceived differences. With the gap in technology that does exist, science fiction can be a bridge to provide deep and thought-provoking hypotheticals on these ethical questions.

I. A Personal Aside

Growing up, I was always struck by the fantastical world of Science Fiction. Images of holograms, robots, and spaceships were an escape for me. The stories of Gene Rodenberry and Octavia Butler would fill my mind with dreams I never thought possible. I learned lessons of compassion, empathy, and curiosity from these authors, and carried their worlds with me wherever I went, looking at them as stories and not reality. It seems, however, that science and science fiction are colliding at an astounding rate. Advancements in Artificial Intelligence over the past 20 years have had a substantial impact on our society and global economy. AI technology manufacturing has led to the loss of millions of jobs worldwide; job loss from automation is expected to grow in the coming decade. We are surrounded by smart devices and "Big Brother"[1] style recognition software making Artificial Intelligence part of day-to-day life for billions of people.

With all the investments being put into the advancement of AI technology the reality of the "Asimovian android" may be closer than we think. There are a multitude of approaches being studied to bring us advanced AI. Neural mapping and complete brain simulation are now seen as a reachable prospect in the coming decades. Symbolic AI, colloquially called Good Old-Fashioned AI or GOFAI, are reaching unprecedented levels of computing power as well.[2] Looking at the array of options Artificial Intelligence may take, and how common these advancements are

[1] Radavoi, Ciprian, "The Impact of Artificial Intelligence on Freedom, Rationality, Rule of Law and Democracy: Should We Be Debating It?" *Texas Journal on Civil Liberties and Civil Rights. Vol 25. 2*. (University of Texas at Austin Law Publications, 2020).
[2] Margaret Boden, *Artificial Intelligence: A Very Short Introduction* (Oxford: Oxford University Press, 2018).

becoming, the question of free will and its subsequent questions on morality and responsibility become quite pressing for today's society.

Chapter One: Overview of Free Will and A.I

I. Chapter Introduction

The goal of this chapter is to provide a brief overture of the major concepts this paper will be working with, namely free will, artificial intelligence, and functionalism. These topics are broad and have been the subject of much scholarship; the proceeding sections will provide the reader with a basic understanding of said topics' respective arguments and give the needed context to articulate this paper's own perspective in the latter chapters.

II. Free Will, A Primer

Freedom of the will designates a kind of control over one's actions.[3]This control has historically been seen as a power or ability one has, innate to them. Many philosophers have equated free will with concepts such as moral responsibility, believing this innate power ascribes moral weight to our actions. Others have claimed that free will is a flawed concept, stating that humanity's actions can be predicted, determined, or controlled to a point where that innate power is either negligible or non-existent. For freedom of the will to exist, it requires the ability to make unrestrained and authentic choices.

A restraint can vary in noticeability and severity. Many argue that every choice has some modicum of restraint. In this context, a restraint refers to physical impediments that control or limit choice. For example, if one comes to an intersection, there are a very limited number of options they could take, but the options presented at the moment present a valid and free choice: do they turn left, right, or straight? If a crossing guard were to direct vehicles away, or a passenger were

[3] O'Connor, Timothy and Christopher Franklin, "Free Will", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/freewill/>.

to grab and turn the wheel for without the driver's consent, there would be restraints from making the choice presented. Restraints can be less absolute. Waking up late, for example, might restrain what one can accomplish before leaving for school or work, without limiting any individual choice.

The other facet to free will is authentic choice. When a person makes a choice, and the decision stems from the self, or when the author and actor are the same, the choice is said to be authentic. This is distinct from restraint, as it is about control over the decision, not the options themselves. On a surface level, choices can be influenced and manipulated without our knowing. Richard H. Thaler and Cass Sunstein describe in their book, "Nudge," A term known as a Choice Architect. "A choice architect has the responsibility for organizing the context in which people make decisions."[4] These architects present information in a way that can influence the decisions people make. Describing a hypothetical involving school cafeterias, Thaler and Cass state that by simply rearranging the cafeteria, "Carolyn was able to increase or decrease the consumption of many food items by as much as 25 percent."[5] Nudging is a kind of passive manipulation that can exert some control over another's decision. As such, it raises into question what choices stem from the self.

Dr Harry Frankfurt, when discussing restraint and control, proposed his famed Frankfurt case. The Frankfurt case is a hypothetical example in which a doctor inputs a device into the brain of one of his patients. This patient and the doctor both want the same person dead, and the device remains dormant unless, at the moment of action, the patient decides he will not kill the man. If that happens the doctor pushes the button and the patient's brain stimulates his want to kill and

[4] Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions Using the Architecture of Choice* (New Haven, CT: Yale University Press, 2008).
[5] Ibid.

will then carry out the murder.[6] This highlights questions on how choices may not be authentically our own. The Frankfurt case is an example of Dr. Frankfurt's principle of alternate possibilities which states that a person is morally responsible for their actions if they could have done otherwise. In the example above, the ability to make the decision could not be said to authentically belong to the patient, thus they would not be morally culpable for the murder. The principle of alternate possibilities is traditionally seen as an argument against determinism, a view which states all actions are causally determined.

The big schism in the Problem of free will is between compatibilist theories and incompatibilist theories, with the central question being whether we can we have Free Will or if our actions are pre-determined. A compatibilist theory would argue we can make our own decisions, hence the term compatible. Different thinkers have various definitions and caveats but agree that we can still claim some sense of free will even if that pure sense of inward choice doesn't exist. Incompatibilism has two main branches, hard determinists and libertarians. Hard determinists claim all our actions are predetermined and as a result, we cannot be free. The libertarian viewpoint rejects the idea of determinism altogether and insists we are intrinsically free. As I will develop later, the conditions of compatibilism, being free from constraint, control, mechanism, and fatalism, is sufficient enough to have for freedom of the will to exist.

The question of free will has significance in society, as moral responsibility is tied to our ability to control and determine our own deeds. Common theories of ethics, such as virtue theory and utilitarianism rely on an ought "– that is they argue there is a correct choice in a given situation that a person should make to live a moral life. Virtue, for Aristotle, existed at the golden mean between a lack or excess of a quality. The virtue of courage for example was at excess the vice

[6] Helen Beebee, *Free Will: an Introduction* (Basingstoke, Hampshire: Palgrave Macmillan, 2013), 139.

brashness and at its lacking cowardness."[7] Utilitarianism commonly argues for the maxim amount of happiness by choosing the option that would lead to the most increase in happiness or pleasure. To be moral in these instances one must choose to do the right thing. Choice being the operative word here, is where free will and ethics intersect. A potential consequence of determinism is the loss of moral responsibility. Knowing the connection between free will and moral responsibility, helps one understand if artificial intelligence can indeed be free has both metaphysical and ethical implications.

III. Artificial Intelligence

III-A. The basics

There are two ongoing conversations with Artificial Intelligence today, one concerns the competing methods in developing the technology, and the other concerns itself with what we do with it. The first spans decades of scientific research and, though it is important to understand how this technology developed and why it advanced in the way it did, is not the central focus of this paper. The second point, "what we do with it," refers to the ethical implications of creating advanced forms of intelligence. Is an AI enabled device alive? Does it deserve rights? How should we use such powerful tools? These are examples of questions in this conversation.

It is easy to think of Schwarzenegger's "Terminator" or Asimov's "I Robot" as what AI machine look like at an extreme level, but AI is, and has been, part of our daily lives. Television shows, like " How It's Made" on the Science Channel, often detail everyday factory machines that utilize AI. More renowned than a canning machine would be the Mars Rover, which utilizes AI to roam the distant deserts collecting and analyzing soil samples. The technology is growing, and AI matching that of the metallic Arnold Schwarzenegger may well be in our future.

---

[7] Aristotle, *Nicomachean Ethics,* trans. Robert William Browne (Penguin Publishing Group, 2020), Book I.

Artificial Intelligence is an Information Processing System (IPS). This IPS is how a machine, virtual or physical, receives and acts on input data. What information a machine can input, how it processes said information, and its available responses to the information determines the sophistication of the AI. A major goal for AI developers, regardless of the method used, is to create a "General Intelligence"[8]. What generality means here is the ability for the IPS to process a wide variety of information in a reasonably efficient way. A general artificial intelligence would have “human-like capacity"[9] to its functioning and would be able to integrate its processes, like learning and language, to handle the tasks before it.[10] “Human-like” is used here because of the perceived flexibility of human intelligence. We have the ability to use existing knowledge to respond to new and unfamiliar situations. In layman’s terms, this is called creative problem-solving. General Intelligence, through its adaptability in information processing, may give them human-like capacity, but does that make them human? Different theorists have provided standards to judge AI and deem when it goes from a machine to an autonomous thinking being. Alan Turing, a renowned computer scientist, provides one view of the problem.

Alan Turing, often seen as the Father of AI, devised the now standard Turing Test. This test is an imitation game, including two humans and one “thinking machine”[11] Each participant is placed in a room, with one human being the questioner, and the other human, along with the machine, being the respondents. A machine has passed the test if the questioner is unable to guess which respondent is the human and which is the machine. A machine, in Turing's thought experiment, is a classic symbolic digital computer, which utilizes a coding system of 0s and 1s

[8] Margaret Boden, *Artificial Intelligence: A Very Short Introduction*, 23.
[9] Margaret A. Boden, ed., *The Philosophy of Artificial Intelligence* (Oxford: Oxford University Press, 1990), 7.
[10] Ibid.
[11] Alan Turing, “Computing Machinery And Intelligence,” *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden (Oxford: Oxford University Press, 1992), pp. 40-66, 40-41.

based 'table of instructions.'[12] The renowned computer scientist framed the question of a thinking machine as such: "Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, [the computer] can be made to play satisfactorily the part of [a human] in the imitation game."[13] Turing, who died in 1954, already foresaw saw this coming and viewed it more as an inevitability than a possibility.

This theory is not without its objections. Philosopher John Searle, a contemporary writer and Professor at UC Berkley, used another thought experiment to undermine the Turing test. Imagine you are in a room and your job is to translate items pages from Chinese to English, but you do not speak either language. Instead, you have a codebook that gives you a detailed list of instructions that informs you which Chinese characters correspond with the corresponding English words. Through one window you get these Chinese characters, and you hand off the translations in English through another. Again, you do not speak either language, you simply match one type of symbol with another and hand them off. Do you understand Chinese? No, you don't Searle posits, you are only following instructions.[14] He uses this analogy to show that a computer program is not capable of thought regardless of how advanced its instructions are. These machines could not be said to have sentience or true thought, because logical syntax alone does not produce understanding.

The arguments for or against the thinking ability of machines have a significant impact on the problem of free will. If a machine can be said to have genuine thought, does that not also imply the ability to choose? Or conversely, if a machine can replicate human thought, does that mean our choices are nothing but a pre-determined program? Searle and Turing provide contrasting views

---

[12] Ibid. 43
[13] Ibid, 48.
[14] John R Searle, “Minds, Brains, And Programs,” *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden (Oxford: Oxford University Press, 1992), pp. 67-88.

as to the nature of artificial intelligence. Chapters Two and Three will explain how, through a functionalist understanding of the mind, AI can achieve the cognition Searle speaks on.

III- B. Artificial Intelligence Today

Thinking Machines have evolved since Alan Turing proposed his imitation game, but despite these advancements, we are not yet capable of creating *Star Wars*' R2-D2. The world, however, has already come to rely on less advanced AI technology. Arek Skuza, a popular tech blogger and AI market strategist, detailed how the international tech corporation, Amazon, has integrated AI on every level of its business model. "AI is not located in a particular office at Amazon, it is more like a ghost cutting across all departments."[15] Starting with the customer experience, Amazon makes use of AI technology through predictive algorithms. When you browse around the site and click on something, the digital machine 'remembers' your search and catalogs it, adding it to its collection of data on you. It uses this data to recommend certain items or promotions. "The role of AI in Amazon's recommendation engine is enormous, as it generates 35 percent of the company's revenue."[16] The automated efficiency created by this AI predictive technology is what is propelling Amazon as a global industry leader. When you do order something, it is a digital machine that catalogs your purchase and sends the delivery information to the appropriate place. At the warehouse, you will find robots and people working together to sort, package and load thousands of purchases. The whole Amazon experience can be tracked and connected digitally, even the delivery vans have amazon's virtual assistant 'Alexa' plugged in to provide directions. The influence this technology is having cannot be denied, it is vital we understand the implications of its advancement.

---

[15] Arek Skuza, "Amazon Artificial Intelligence - How Jeff Bezos Automates the Company," *Arek Skuza Blog*, November 10, 2020, https://arekskuza.com/the-innovation-blog/amazon-and-artificial-intelligence-in-retail/.
[16] Ibid.

III.- C. Consequences of AI

Efficiency may be a goal of predictive algorithms, but the consequences of their implementation may be more dangerous than most people realize. AI technology has been integrated into law enforcement across the world with predictive policing. "Predictive policing refers to the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions."[17] This process has become digitized with predictive algorithms and analytical programs. These algorithms have been used to predict crime hotspots and find likely suspects. "Under the mantle of science and objectivity, predictive tools of this type criminalize poverty," as they still disproportionally target and single out disadvantaged communities.[18] In this more efficient and impersonal automation, these predictive algorithms continue to reinforce problematic bias within the system.

An ongoing debate within the AI community is the ethics of implementing such advanced systems in our society. The fear of a "Big Brother" society is real. In China, the government has implemented "The Social Credit System, which scores people on every aspect of their lives, from credit rating to “social sincerity,” involves surveillance and monitoring of the totality of individuals’ lives"[19], this system is possible because of AI predictive and surveillance technology. Similarly, automation is a point of concern regarding AI. As was discussed, millions of jobs have been automated in the past 20 years, and it is expected that a quarter of all jobs can and will be automated in the coming decades[20]. The concern is how said implementation will affect the

[17] Radavoi, Ciprian.“The Impact of Artificial Intelligence on Freedom, Rationality, Rule of Law and Democracy: Should We Be Debating It?” *Texas Journal on Civil Liberties and Civil Rights*. Vol 25. 2. University of Texas at Austin Law Publications, (2020), 9.
[18] Ibid.
[19] Ibid.
[20] Jacobson, Prof. Joseph Mayer. “Freedom of Choice and Artificial Intelligence,” *B'or ha 'Torah* Vol. 26 (Winter 2019), 46-57.

freedom and economic well-being of the lower class. It will be poor communities who are most likely to experience job loss and be the most scrutinized subjects of surveillance.

AI technology is engrained in our global economy and infrastructure; we must understand what the technology we have become so dependent on actually is. We are using AI as an exploitable resource because of the efficiency it can create, but if we are creating intelligence, what should we do when that intelligence surpasses or matches our own? This is where Science-Fiction can help provide some clarity. Science Fiction authors have been using the façade of technology to probe questions of society, ethics, mortality, and acceptance to name a few. Their work, though hypothetical, has not only inspired generations of scientists but technology itself. Chapter Four will expand on the stories of *Star Wars* and *Star Trek*, treating them as philosophical thought experiments to connect the concepts of A.I and moral standing.

IV. Functionalism, A Cross-section of AI and Free Will

To understand how this intelligence relates to our own, let us look to a functionalist theory of the mind. Functionalism posits that "what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part."[21] In this definition, a mental state is a thought, emotion, or other singular condition of the mind. A quick google search defines the mind as the "element of a person that enables them to be aware of the world and their experiences, to think, and to feel; the faculty of consciousness and thought." The mind then is the system that determines the internal constitution of mental states. Functionalism has a few variations, but the following are described as three central claims of classical functionalism. "1) What unites all cognitive creatures is not that

[21] Churchland, Paul M, *Neurophilosophy at Work* (Cambridge, UK: Cambridge University Press, 2007), 5.

they share the same computational mechanisms (their 'hardware'). What unites them is that (plus or minus some individual defects or acquired special skills) they are all computing the same, or some part of the same, abstract sensory input, prior state motor output, and subsequent state functions. 2) The central job of Cognitive Psychology is to identify this abstract function that we are all (more or less) computing. 3) The central job of AI research is to create novel physical realizations of salient parts of, and ultimately all of, the abstract function we are all (more or less) computing."[22] What this means is that the computational requirements for abstract processes can be achieved in a wide variety of ways. Ways in which artificial intelligence can foreseeably achieve.

Mathematician and philosopher Hilary Putnam, for example, posits a "Machine State Functionalism" which argues that any mind within any given organism can be said to be a Turing machine.[23] Like Turing's hypothesized digital machine, the human operating system is advanced but finite. Machine State Functionalists believe that mental states can be put onto a deterministic, or probabilistic, table like the functions of a computer program. Under this view, the connection between A.I and mental causation is quite explicit, as it directly ties both human thought and A.I to program-esque functions.

This connection, though not as explicit, is present in the broader functionalism argument. A mental state is defined by its role, meaning that any program designed to express emotion or respond intelligently can be said to have a mind. This is a bit of an overgeneralization but will be elaborated on in the next chapter.

---

[22] Ibid.
[23] Janet Levin, "Functionalism," *Stanford Encyclopedia of Philosophy* (Stanford University, July 20, 2018), https://plato.stanford.edu/entries/functionalism.

V. Religious Objections to A.I

A popular counter to functionalist or physicalist approaches to AI come from religious objections. This group of objections places the soul, or some sort of unique spiritual essence, as the source for thought. Our mind is not physical, but ethereal. Consciousness then is not a function or role in a system, but the fabric of a human's very being. MIT Professor Joseph Mayer Jacobson poses the question as such, "If people are simply computer programs, executing instructions according to what they have been programmed to do, what is the point of the Torah and Creation in the first place?"[24]. The potential connection between man and machine could invalidate the uniqueness of our creation and 'God-given' freedom. This objection stems from an incompatibilist understanding of free will, with the soul being the factor that makes determinism and free will incompatible with each other. Though we were created, much like we create our machines, our souls were made to express free will. According to some Judeo-Christian and Islamic traditions, the angels of God were created in a similar vein to AI. "[The Angels said] "Glory to You! We have no knowledge except what You have given us. You, only You, are All-Knowing, All-Wise." (Baqarah/30-32)[25] Philosopher Mustafa Cevik claims this phrase is an acknowledgment from the angels that humans are superior. "We can infer the essential divergence between the two species from what angels say. The real difference between angels and Adam is related to the source of knowledge rather than its quantity."[26] The source of knowledge for humanity is our consciousness, which a machine (and apparently an angel) is unable to replicate. In this view an angel and AI are practically equivalent beings: They are both programmed with a subset of 'knowledge' and are limited by it. Knowledge, in this sense of the word, would be the AI's code, or in functionalist

[24] Jacobson, "Freedom of Choice and Artificial Intelligence," 49-50.
[25] Cevik, Mustafa, "Will It Be Possible for Artificial Intelligence Robots to Acquire Free Will and Believe in God?". *Beytulhikme An International Journal of Philosophy*, Vol. 7 No. 2 (Winter 2017), 78.
[26] Ibid.

terms, its internal constitution. Humanity's internal constitution, as designed by God, enables us to have this unique freedom.

The challenges AI face, religious or otherwise, stem from the notion that free will is in some way unique to humanity or biological life. Chapter Three will detail how humanity can maintain its unique notion of free will while still accepting alternate or different conceptions of freedom.

VI. Chapter Summary

The above chapter is meant to give the reader a basic introduction to the concepts of free will, artificial intelligence, and functionalism. Much of what was discussed will be elaborated in the forthcoming chapters as we examine the relationship between these three concepts. Understanding how we view free will and who we ascribe moral responsibility to is an important question as AI technology advances. Functionalism can be a tool to look beyond a human-centered conception of intelligence and help us look at free will in a less biased way. In the next three chapters, I will return to my following three primary research questions, using the theory of functionalism as a guide: 1) under what circumstances can machines be said to have intelligent thought; 2) does having intelligent thought mean they have freedom and moral responsibilities; and 3) do we as humans have any moral obligations when these machines reach said point? In answering these questions, I hope to demonstrate why adopting the mentality that these machines will one day have free will is critical as we continue to advance this technology.

Chapter Two: Functionalism and Intelligence

I. Chapter Introduction

Chapter one detailed the current status of AI technology and the current trends in the free will debate, chapter two will address the first research question, under what circumstances can machines be said to have intelligent thought? This paper takes two parallel approaches in answering this question. The first is to use functionalism and ethical theory to argue why conceptions of intelligence should not be narrow in scope. The second is to re-examine the current achievements of AI technology using culturally acceptable definitions of intelligence. Together, they will show why AI can and should be thought of as intelligent beings.

II. Expanding Our Conception of Intelligence

We understand from Margaret Boden that the goal of Advancing A.I is to reach a "general intelligence," a kind of processer that has the flexibility to be applied to a multitude of differing situations. When this happens – I will say I do believe the technology will reach that point eventually, but for now let us say 'when' for the sake of argument – Is that intelligence of the human sort? First though, what is intelligence? The first entry from the Merriam-Webster website is seen below.[27] The Merriam-Webster entries provide an accessible and culturally relevant conception of how intelligence is viewed by society.

1 A (1): the ability to learn or understand or to deal with new or trying situations: Reason; also: the skilled use of reason

[27] "Intelligence," *Merriam-Webster* (Merriam-Webster), accessed January 29, 2021, https://www.merriam-webster.com/dictionary/intelligence.

(2): the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)

B: mental acuteness: SHREWDNESS

C: Christian Science: the basic eternal quality of divine Mind

We have heard from Alan Turing that the successful imitation of human intelligence is sufficient grounds to call a machine intelligent. His Turing Test seems to match with the above definition labeled A.2, as an objective criteria test involving the application of knowledge. We also touched on John Searle's response against that notion. It is important to note that Searle does not dismiss the notion of man-made intelligent creations outright, rather this critique is specific to digital computers. Searle, as I will detail more later, believes such a program cannot gain a real understanding of the world. Turing and Searle each present valid arguments, but both philosophers seem to imply this intelligent thought must model our own. Alan Turing does so by emphasizing imitation and John Searle does the same thing by critiquing the internal constitution of computers. This does not necessarily have to be the case.

A basic principle of functionalism is that "the relevant function is computable in a potentially infinite variety of ways, not just in the way that humans happen to do it"[28] In psychological terms, this is referred to as the Lashleyan doctrine and argues that a wide variety of brain structures could lead to any of a wide variety of psychological functions.[29] This is supported by the observation of parallel evolutionary patterns. "The analogous point about behavioral similarities across species have been widely recognized in the ethological literature: organisms of widely differing phylogeny and morphology may nevertheless come to exhibit superficial

[28] Churchland, Paul M. "Functionalism at Forty: A Critical Retrospective," *The Journal of Philosophy 102*, no. 1 (2005): 33-50. http://www.jstor.org/stable/3655766.
[29] Block, N. J., and J. A. Fodor. "What Psychological States Are Not." *The Philosophical Review 81*, no. 2 (1972): 159-81.

behavioral similarities in response to convergent environmental pressures."[30] What this means is that functionally similar behaviors are present across a diverse array of species. Just like bats and bees accomplish the function of flight with different wing structures animals can accomplish the function of fear or pleasure with different brain structures. A functionalist takes this notion and applies it to artificial life as well. We already see in nature how diverse a mental state can be, so this extra step is not much of a leap. A machine is just another configuration that could lead to psychological states.

Despite this natural diversity, intelligence or consciousness is classified with the assumption that humanity is the gold standard. Machines, animals, plants, and even aliens would need to communicate on our level for us to consider the notion that they have intelligence. Many have narrowed this conception to exclude entire populations, races, or ethnic groups which has led to countless moral travesties. This is because when something is deemed unintelligent, specifically a human's perceived level of intelligence, they are viewed as lacking in moral standing. This "othering"[31] of fellow humans has been the subject of many philosopher's work, including that of French existentialist, Simone De Beauvoir. While she is best known in the world of feminist theory, her avant-garde approach to philosophy has garnered her high praise alongside her life partner, Jean-Paul Sartre. Their movement, known as French existentialism, grew out of the ashes of war-torn Europe needing to reconcile the atrocities and inhumanity inflicted by Nazi Germany. De Beauvoir's work, like *The Ethics of Ambiguity* and *The Second Sex,* explores the various facets of the other. Gender for example sees women as an other. Patriarchal culture has positioned man as the norm in body, anatomy, strength, etc. and has left women outcasts as the 'other' in society.

[30] Ibid.
[31] Simone de Beauvoir, *The Ethics of Ambiguity,* Trans. Bernard Frechtman (New York, New York: Open Road Integrated Media, 2018).

It is our ability to turn people into an other that has allowed for the subjugation and suppression of their freedoms. De Beauvoir ties freedom to morality directly, arguing that a true moral attitude requires the recognition of your own inherent freedom along with the acceptance of the freedom of others.[32] The process of 'othering,' which involves the dismissal or devaluing of a person/group's intelligence, leads to the suppression of their freedom.

So where does this fit into the AI discussion? As we examine the question "under what circumstances can machines be said to have intelligent thought," we must remember the tragic consequences humans have inflicted by failing to recognize the moral status of the other. It is important to clarify that Simone De Beauvoir was speaking about the ways in which humans were treating each other. Artificial Intelligence was not a focus of her work and may not have been a concern of hers while writing about the other. This paper believes, however, De Beauvoir's concept of "othering" followed faithfully, would include all beings able to recognize their freedom, even non-humans. Similar discussions are taking place regarding the humane treatment of animals, see Peter Singer's writings on equality for animals[33], in which activists and academics are claiming animals are worthy of moral consideration. The circumstances regarding animals and A.I are markedly different but shared between them is a challenge to the human-centric notion of morality. The Lashleyan doctrine, as expressed in a functionalist theory, provides a scientific justification for expanding our conception of intelligence. Simone De Beauvoir's ethics of ambiguity lays out a moral argument against narrowing our conception of autonomy. Coupled together, the two theories provide a strong argument for why machine intelligence should not be dismissed outright.

[32] Ibid
[33] • Singer, Peter, 1990, Animal Liberation, second edition, New York: New York Review of Books.

III. Conditions of Intelligence

Knowing then, that we should approach the question of intelligence broadly, what conditions would a machine need to meet in order to be said to have intelligent thought? Returning to the Merriam-Webster definition the first option they give is the "ability to learn or understand or to deal with new or trying situations"[34] John Searle argues that this kind of intelligent thought requires a degree of intentionality. Through his Chinese room experiment, which we touched on earlier, he attempts to demonstrate how the digital process is entirely syntactic. Searle posits that the syntactic nature means there can be no semantic understanding of the situation. "In the [Chinese room] I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing."[35] The moment this hypothetical Searle leaves the Chinese room, he can no longer communicate in Chinese, because he never knew the language in the first place. Searle believes that the ability to gain this understanding lies in our organic structure, specifically the neural tissue. "Metal and silicone" cannot reproduce understanding, in the same way that they cannot photosynthesize like chlorophyll.[36] Having an understanding allows for acts to be intentional, at least in a 'conscious way.' Here intentionality is narrowed in its scope, as it cannot be denied that in the Chinese room Searle is intentionally matching the symbols he does not understand.

Understanding and intentionality are good indicators of a general intelligence. It implies an ability to respond to situations with the creativity and nuance computer scientists are looking for in AI. As for John Searle's argument against a digital machine, he relies on unwarranted assumptions when he claims that they will never reach that point. I will readily concede that the

[34] "Intelligence," *Merriam-Webster* (Merriam-Webster),
[35] Searle, "Minds, Brains, And Programs," 67-88.
[36] Margaret A. Boden, "Escaping The Chinese Room," in *The Philosophy of Artificial Intelligence* (Oxford: Oxford University Press, 1992), pp. 89-104.

technology is still hypothetical and that advancements in the industry are slower than many ardent supporters would like, but the simple fact is it is too early to rule out, from both a neurological perspective and technological perspective, the possibility of machine understanding. Margaret Boden speaks to this in her paper responding to Searle's Chinese Room thought experiment. She addresses this claim of intentionality by looking to what machines can already replicate, such as vision. "Metal and silicon" she says, can already 'support some of the functions necessary for the 2D - to 3D mapping involved in vision."[37] We have seen progress in expanding the kinds of tasks A.I are capable of completing, but to assume that neurochemical activity is different from the other processes already replicable at this point has no scientific grounding.

The second part of the definition from Meriam-webster is "The ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)."[38] Machines are acting on the first half of this definition already. Earlier examples such as the predictive policing algorithms are constantly receiving new data and manipulating their virtual environment. The second half, specifically the words 'think abstractly' is where the debate lies. Tests have been made to distinguish levels of intelligence among humans and animals; Alan Turing's imitation game is also a valid intelligence test for machines. This test strips away many elements people expect of sentient robots, such as the ability to speak, to resemble humans, or even to have mobility. All visual or auditory clues are removed, leaving only the ability to respond to questions and imitate correctly.

It is this desire for imitation that can be limiting. Imitation is valuable and has been argued quite compellingly by some of Turing's intellectual descendants, but what is being created is different than us and we must not make sameness the goal. Turing posits that the clever machine

[37] Ibid.
[38] "Intelligence," Merriam-Webster (Merriam-Webster),

would hide its ability to calculate math faster,[39] which creates an ethical dilemma. It is one thing to imbue a machine with the intelligence to lie or hide, it is another to design a test that bases their moral worth on the ability to do so. Remember from chapter one that having free will means having moral responsibility. When a machine passes one of these tests, they are believed to be intelligent and capable of making authentic choices. This would result in them being responsible for those actions. When creating general intelligence, we should understand two things. First: how they perceive and interact with the world will be different than us, this will lead them to act and think differently. Second: We should strive to create artificial intelligence with as clear a moral understanding as possible. They may fail the Turing test because of this, despite possessing remarkable intelligence.

AI have been able to beat humans at imitation games, and the BBC reported in 2014 that a team created an AI program that successfully passed the Turing test at an international competition. This claim has been disputed, but the fact that the program convinced some of the judges is not.[40] Officially, it appears no machine has yet to win an accredited Turing test, but they have won other kinds of tests. IBM's Watson made headlines in 2011 when it beat two long-standing champions in a "Jeopardy!"[41] tournament. Other companies have developed AI systems that have beaten chess grandmasters, curling athletes, and the Rubik's Cube record holder. These tests may seem unrelated, but they show inklings of possibility for where the technology is headed, and how by many standards they may already meet the criteria to be called intelligent.

---

[39] Alan Turing, "Computing Machinery And Intelligence," 43.
[40] "Computer AI Passes Turing Test in 'World First'," BBC News (BBC, June 9, 2014), https://www.bbc.com/news/technology-27762088.
[41] Jo Best, "IBM Watson: The inside Story of How the Jeopardy-Winning Supercomputer Was Born, and What It Wants to Do Next," *TechRepublic* (TechRepublic, September 9, 2013), https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/.

IV. Chapter Summary

In this paper, we are drawing a line between intelligence, choice, free will, and moral standing, which makes any test that categorizes intelligence akin to a moral toll booth. Intelligence, being seen as the faculty with which one examines and makes choices, is connected to free will. As described in chapter one, free will requires a choice to be authentic to the self; intelligent beings are said to have the rational capacity to author their own choices. How intelligence is defined, and who is deemed intelligent, has immense weight on how they are viewed in society. An intelligence test then, must avoid narrowing its scope and othering a segment of the population. On a purely human level, considerations must be made for those suffering from some mental injury or are in a vegetative state. Considerations must also be made for those in the neurodivergent community. Language, age, and physical disabilities are other factors that may make it difficult to communicate the intelligence one has or is capable of. Many humans could and have been excluded from participating in society as a result of narrowing how we define intelligence. When looking at the various definitions and conditions for intelligence, we find that the science of generally intelligent AI is ongoing and, in some cases, showing promising. This is why it is vital we ask these questions regarding free will and moral standing now.

Chapter Three: Artificial Intelligence and Freedom

I. Chapter Introduction

Knowing then, that general artificial intelligence is technologically possible, and that intelligence should not mimic or operate like our own to be considered valid, we now turn to the question of freedom. The following sections will discuss freedom from a functionalist perspective and a compatibilist perspective, looking to synthesize requirements for freedom that exist independently of the human experience of free will.

II. Human Freedom and Functionalism

We've discussed functionalism at length now and understand its basic tenant that there exists a multitude of pathways to achieve a given function. Even if the different beings can complete all the same functions, does a difference in internal constitution affect the ability to be free? A common belief about why this may be so is the notion that humans are in some way intrinsically special. People are afraid of sharing the title of free with machines for the perceived consequences it would have on ourselves, our standing in the universe, the divine's role in our creation, and the status of our freedom. Philosophers have touted humans as having the unique gift of reason and politicians have codified the idea of God-given inalienable rights into the framework of our government. Chapter One addressed several of these religious objections. We are special, and the notion of free AI calls our uniqueness into question. If we can recreate freedom in machines, what does that say about our God-given rights or our monopoly over reason? Skeptics of AI, those who doubt their intelligence altogether, would claim any equivalence between man and machine would devalue the very essence of our freedom.

"There is something about the prospect of an engineering approach to the mind that is deeply repugnant to a certain sort of Humanist."[42] This is a quote from Daniel Dennett's essay "When Philosophers Encounter Artificial Intelligence" and he makes the argument that, right or wrong, this approach to the mind and AI run contrary to the dignified perceptions we have made for ourselves. We view the mind as a purity of essence or a wondrous mystery, conceptions which add to our perceived uniqueness. A machine takes away the mystery or purity and reduces it to sequential mechanical operations.

It is not unreasonable to assume a uniqueness in humanity's intelligence that transfers upon us a kind of free will. Accuracy aside, it has been common practice to view our animal neighbors as largely instinctual beings, a distinction that makes our ability to apply reason and make choices seem different. There is a bit of bias in being the classifier and a subject of classification, but that is beside the point. I propose we adopt a functionalist approach to classifying freedom. We can choose to accept both that our pathway to freedom, the internal constitution, is unique and that there are other acceptable paths to achieve freedom. The following paragraph is a snapshot of one philosopher's attempt to understand choice and functionalism. I am not advocating his view in particular but demonstrating how one might come to accept the freedom of other beings while preserving a sense of uniqueness.

Dr. Paul Churchland in his paper "Functionalism At Forty: A Critical Retrospective" examines how functionalism has evolved and where it has failed in examining cognition, Churchland puts forward a "nonequilibrium thermodynamical portrait of cognitive activity"[43] to be the underlying rules or context which governs mental states. He rejects a full functionalist

[42] Daniel Dennett, "When Philosophers Encounter Artificial Intelligence," in *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen Richards Graubard (Cambridge, MA: MIT Press, 1988), pp. 283-296, 285.
[43] Churchland, "Functionalism at Forty: A Critical Retrospective," 42.

approach in favor of this revised approach. “It is vital to appreciate that the structural and dynamical portrait just painted – of vector coding and vector processing via large matrices with plastic coefficients – is once again a portrait that can be realized in a wide variety of material substrates: in mammalian brains, in octopus brains, in extraterrestrial brains, in electronic chips, in optical systems, and so forth.”[44] Churchland provides a basic structure of cognition that relies on energy flow and information transmission, which can be actualized using several different pathways. This actualization, however it is achieved, leads to fully cognitive beings who can make valid choices. If then, Churchland’s cognitive assumptions are correct, any being that follows his underlying laws are on equal ground to be considered free.

Again, this is one philosopher’s approach to understanding cognition, and in true functionalist fashion, each approach requires its own path to accept my proposal. It relates to Mustafa Cevik’s quote from Chapter One: “The real difference between angels and Adam is related to the source of knowledge rather than its quantity.”[45] His objection is quite different from that of a disillusioned functionalist. Knowing that I cannot respond to every claim, I will say this. Expanding what we accept as a baseline for cognition does not inherently devalue something special in a particular group or belief. The question we should be asking is if we should deny the right to moral responsibility to a being that is functionally intelligent?

III. A.I and Compatibilism

A digital computer could have all the same functional capacities as a human being, but many would never concede that it has freedom. This centers around the determinism debate mentioned in Chapter One. “Determinism states that our actions are the result of the laws of nature

[44] Ibid 22.
[45] Cevik, “Will It Be Possible for Artificial Intelligence Robots to Acquire Free Will and Believe in God?” 78.

and events in the remote past"[46]. In this case the laws of nature would include the operating system of the computer. All the choices a computer makes could be theoretically predicted with the right knowledge. Free will libertarians reject the notion that human freedom operates this way, believing that regardless of the laws or past humans still can choose in a spontaneous manner. Determinism claims that choices are made, but that the decision you make in any given situation is determined by causal events. AI, because they are machines made by humans, are not often seen as free in the libertarian viewpoint.

It is a compatibilist viewpoint, claiming free will and Determinism can co-exist, where arguments for AI's freedom typically lie. Compatibilists recognize the universal causality of events, but do not believe it is a sufficient reason to claim we cannot be free. The classical compatibilist argument points to four factors that do limit freedom: Constraint, Control, Fatalism, and Mechanism.[47] Constraint and Control are how they sound; is there something or someone physically preventing you from making a choice. Fatalism refers to a different belief, separate from determinism, that posits trying to prevent events from occurring is pointless. The last one, Mechanism, is the belief that determinism reduces us to instinctual or mechanical responses only. The latter two are rejected by compatibilists because even with determinism we still have faculties of reason and emotion that guide our decisions. The kind of freedom developed here is different from a libertarian viewpoint. As a thinking being, one can deliberate and make a choice, even if that choice is made in a causally determined world. For a libertarian, the idea of a causally determined world destroys any sense of authenticity the author of an action may have. Compatibilists believe the standard free will libertarians set is either unhelpful or unrealistic, and that the absence of the four factors listed above is enough for a choice to be authentic.

---

[46] Kane, Robert. *A Contemporary Introduction To Free Will*, (New York: Oxford University Press, 2005), 1-12.
[47] Beebee, *Free Will: An Introduction*, 24-26.

Outlined in this paper thus far is a functionalist case for the autonomy of AI. Knowing that the technical possibility still exists, this paper argued for why we should expand our notions of intelligence and freedom to include different kinds of beings. Nature has shown itself to be naturally diverse, even critics such as Dr. Paul Churchland recognize that a multitude of substances can adhere to the same underlying rules. Various definitions and standards for intelligence already include AI in use today, and as the technology develops it will become increasingly difficult to deny their functionally intelligent capabilities. AI, with the parameters we have set in previous sections, can one day indeed overcome the conditions compatibilists say limit our freedoms. General intelligence and a functionally sufficient mode of cognition would be sufficient for any kind of being to have freedom from the compatibilist standpoint. I want to be clear that there is a distinction between human freedom and the minimum requirements that a being would need to have free will. General artificial intelligence would have the ability to respond to a variety of situations independently, meaning they can be built free from constraint or control. An AI that possess functional cognition can be said to make authentic choices, which would make them free from mechanism. The final factor, fatalism, is an independent issue that would affect both humans and AI. According to compatibilism, beings can overcome fatalism with sufficient capacities for reason and emotion. Thus, sufficiently advanced artificial intelligence would be considered free from a compatibilist standpoint.

The point of this paper is not to prove the freedom and morality of humans, I have spent a good portion of this thesis asking and arguing for why we should look beyond a human-centered model of freedom. Humanity's free will status is not the metaphysical mountaintop. Quite frankly, we do not actually know where we stand on the free will issue. Nor can we reasonably put ourselves at the top of a universal intelligence hierarchy. We lack the knowledge to rank or classify

free will on a purely human level. Humans could very well be undetermined as the libertarians suggest, we just don't know. A libertarian case for Artificial Intelligence would be much more difficult to make, but it would be much more difficult to justify. Even for humans justifying a libertarian conception of free will would be difficult. We do, however, have a reasonable argument for a compatibilist case. As mentioned, AI that meet this standard of sophistication are not facing any outside constraint or control, and they possess functionally sufficient reason and emotional capabilities. By all societal standards – morality, law, culture, etc. they can live and work alongside us.

Knowing that machine free will can be different from human free will, is the argument that sufficiently intelligent AI possess moral standing still valid? This potential objection arises from the connection free will has to moral responsibility. How different does this freedom have to be before it no longer carries moral value? Really the question at hand is Should compatibilism be considered a reasonable standard, regardless of where humans may one day be discovered to lie on this free will scale? Yes. Functionally, meeting the requirements for compatibilist free will is sufficient to be free and bear moral responsibility in society. If one day, humanity was able to scientifically prove that they possess libertarian free will, this would still be the case. The ability of these machines to function within society will not change as a result of this new knowledge, nor should the way humanity respond to them. Consider the opposite scenario, scientists prove determinism is in fact the reality for humans. These machines, being different from us, would still be free in the compatibilist sense. Should they treat humanity as lesser beings or limit humanity's autonomy in this shared society? Most would instinctually say no. On this spectrum of free will, compatibilism has established itself as sufficient in maintaining an authentic enough conception of choice to allow for moral responsibility.

IV. Chapter Summary

This chapter argued that compatibilism is an acceptable conception of freedom and that, regardless of the nature of human free will, is sufficient to have moral responsibility. This chapter also demonstrated how artificial intelligence can be free from a compatibilist conception of free will. The next chapter will elaborate on the significance of this issue and why this question is relevant now.

Chapter Four Moral Significance

I. Chapter Introduction

Having outlined a path in which AI can be considered free, I will now turn to the third question of this paper "do we as humans have any moral obligations when these machines reach said point?" They can be free, but AI are still created with functional purposes. Scientists are not necessarily creating AI humans. We are making ship computers, robot explorers (think Mars rover), or taxis. This is where I will turn to Science-Fiction to help us frame the question. It is hard to grasp the moral significance of the AI issue when no real-world ethical application exists. One could look at animal ethics, but they are biological organisms. Looking to humans would present a whole slew of other moral complications such as the nature of the soul and religion. Having no ideal place in the real world to look for applied ethics, the place to go to examine the consequences of Free AI lie in the world of fiction. The stories these authors create become the dreams of scientists by which new invention is inspired. John Searle, who has expressed some skepticism that AI can be conscious, had this to say about science fiction. "Of course, it is science fiction, but then, many of the most important thought experiments and science are precisely science fiction."[48] The worlds of science fiction are grounded in actionable rules and restrictions which make them the perfect philosophical thought experiments.

For this paper, we will be examining characters and scenes relating to artificial intelligence that highlight humanity's moral responsibility to these entities. There are stories, such as *The Matrix* or *Terminator,* that cast AI in a menacing light. They are valid expositions, but the questions they ask of the audience are usually about the blurred lines between man and machine

[48] Searle, John. *The Rediscovery of the Mind.* (Cambridge, MA: The MIT Press, 1992) 77.

or the de-sanctification of humanity and the body. Thought-provoking questions to be sure, but not relevant to the question at hand. To answer our questions, we will look to the fictional universes of *Star Trek* and *Star Wars*.

Star Trek started as a television series in the 1960s and was written by Gene Rodenberry. The story followed the crew of the Star Ship, Enterprise of the United Federation of Planets as they "Seek out new life and civilization." The Show, and its subsequent movies and sequel series, explored questions of morality, justice, and philosophy through the lens of space adventure. Star Wars is a movie franchise created by George Lucas in 1977. Since the first movie's release, there have been eleven other canonical movies, several TV shows, and dozens of books and comic series. The stories take place in a galaxy far from our own long before humanity set foot on earth. The main series follows the rise and fall of a galactic empire and a fight between the very forces of good and evil. The two worlds could not be more different, but they each feature unique takes on the nature of Artificial Intelligence and how we as humans should respond to it.

II. Commander Data and Understanding Difference

We will start by examining *Star Trek: The Next Generation's* Lieutenant Commander Data, an android who served aboard the U.S.S Enterprise. Data was created by cyberneticists Dr. Noonian Soong in the late 2330s and imbued with the memories of the colonists his creator shared a settlement with. He was left deactivated on his home world for some time, before being discovered and re-activated by a Starfleet landing party. He eventually enrolled in Starfleet Academy, and despite facing social obstacles, graduated in exobiology and probability mechanics.

Data had a 20-year career serving as a Starfleet officer before being assigned to the Enterprise as its second officer.[49]

Mechanically, Data is based on the ideas of Biochemist and author Isaac Asimov. Commander data has a "positronic brain,"[50] which is a fictional technology featured prominently in some of Asimov's literary work. Data is also bound by Asimov's Three Laws of Robotics, detailed below.[51]

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 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 Law.

In addition to the above constraints, Data spends most of his life without an "emotions chip," meaning he cannot feel emotions in the way humans do. Despite this lack of emotion, Data is intrigued by humans and has an apparent desire to learn to be like them. This desire is featured in many of the show's story arcs and is a central point to his character's development throughout the series.

Data, in his struggle to become more human, highlights the need for us as humans to respond empathetically and morally to artificial beings. The best example of this can be seen in the next generation episode "The measure of a Man." In this episode, which took place in the middle of the sequel series' second season, Commander Data is ordered by Dr. Bruce Maddox, a cyberneticist, to transfer from the Enterprise to be studied and hopefully copied in his lab. When Data realizes this study could be harmful, he objects and eventually resigns from Starfleet. Dr. Maddox informs Data and Captain Picard that he cannot resign from his post, as Data is Starfleet

[49] "Data," *Memory Alpha* (Wiki), accessed January 29, 2021, https://memory-alpha.fandom.com/wiki/Data.
[50] Ibid.
[51] Ibid.

property, not a citizen. Outraged by this, Captain Picard takes the matter to court, representing Data and challenging the idea that he is property. [52]

This situation is ripe for philosophical discussion, and the episode's courtroom setup allows the characters to explore the emotional and moral impact Commander Data has had on them while debating the moral impact their decision will have on him. In this hypothetical situation, think of it as a thought experiment. We have an artificial entity that we can prove has a functional level of cognition to be an asset in military and scientific expeditions. We also know, however, that they do not have an equivalent process for feeling emotion. That is not to say they do not form certain kinds of attachments or relationships. This being chose on its own accord to join Starfleet and passed the same entrance exams as other sentient beings. Dr. Maddox claims we are showing overt sentimentality to "it" because it looks and sounds like a human. This is true, but Data's machine-like quirks would cost him dearly in a Turing test. Captain Picard argues that, as a team of explorers seeking out new and different kinds of life, to not show him a moral benefit of the doubt is to risk condemning a unique form of life to slavery and death. To be fair to Dr. Maddox, the moral question here isn't just about Data. The doctor is proposing a study that could benefit the entire human race, creating a whole army of Datas that can serve humanity in almost any capacity. If Picard is right though, it makes the consequences of Maddox's experiment much more horrific.

This is a powerful thought experiment, but how can we apply it to AI today? In chapter one, I highlighted how Amazon has integrated artificial intelligence into its entire corporate structure. Dr. Maddox is suggesting Starfleet do the same thing, but with a computer that spends his free time pretending to be Sherlock Holmes.[53] How should we respond to Amazon if they announce "Alexa" has a functional level of self-awareness? Data's inability to mimic humans

[52] "The Measure of a Man," *Star Trek: The Next Generation* (Viacom, February 13, 1989).
[53] "Data," *Memory Alpha* (Wiki), accessed January 29, 2021, https://memory-alpha.fandom.com/wiki/Data.

means he'd fail a Turing test. When we judge this new "Alexa" would her inability to fool us lead to her subjugation? In chapter two I referenced, Simone De Beauvoir's ethics of ambiguity to highlight the dangers of denying rights to those who are different. The process of othering is dangerous and morally reprehensible. Dr. Maddox viewed Data as an other, and was willing to terminate him because of it. Was Dr. Maddox correct in denying Data moral status?

I will continue this thought by jumping to another character in Star Trek for a point of comparison.

III. The Emergency Medical Hologram.

The Emergency Medical Hologram, also known as the "EMH" or simply "The Doctor" was part of the main cast of *Star Trek Voyager*. The Doctor was a holographic computer program created by Dr. Lewis Zimmerman to serve as a physician in emergencies. Events in the show's first episode left the hologram as the only medically trained entity left alive on the ship. Being promoted to Chief Medical Officer, The Doctor ran not as originally programmed, developing a unique personality and acquiring hobbies due to his prolonged activation.[54] Star Trek Voyager posed many questions over its seven-year run, but one relevant to the previous section can be seen in the season 3 episode "The Swarm."

During this episode, The Doctor is suffering major memory loss. After some investigation, it is revealed that the source of this memory loss has to do with his program's capacity. As an emergency program, The Doctor was only meant for 1,500 hours and the self-expansion of his original programming was also creating issues. The story comes to a head when The Doctor is

[54] "The Doctor," *Memory Alpha* (Wiki), accessed January 29, 2021, https://memory-alpha.fandom.com/wiki/The_Doctor

unable to remember a life-saving procedure. Ultimately, The Doctor decides to undergo a change that wipes most of his memory.[55]

Mustafa Cevik, who I mentioned in chapter one, argued that digital computers could not rise above the limitations of their programming. AI specialists are working to create a general intelligence that can adapt and grow. The Doctor is an AI capable of altering themselves to grow beyond their original purpose. Their sarcastic personality would also lend to their passing the Turing Test. Despite this growth and passability, The Doctor is expected to sacrifice themself and their identity on multiple occasions. I'll ask the same question here that I did for Data: Is it right for the human crew to other The Doctor and deny him moral status? Now the Doctor did make this decision, but not without being pressured. No one would reasonably ask a human physician to wipe their memory to relearn a medical procedure.

Using the criteria we outlined in previous chapters, the answer for both Data and The Doctor would be no, it is not right. The two digital entities, though different in structure and design, are functionally conscious. They are aware of their relative position in the world and can respond uniquely to situations. They make choices that stem from some sort of internal deliberation and can pass various forms of intellectual tests. They possess intelligence and, in their natural state, are free from constraint or control, and possess faculties of reason. Data and The Doctor form their own attachments and have expressed feelings of love. They are made of machine parts and can be switched off, but from a compatibilist standpoint, they are free and deserving of moral standing.

[55] Ibid.

IV. L3-37

To provide a brief contrast to the world of *Star Trek*, we will turn to the droid L3-37, seen in the movie *Solo: A Star Wars Story*. The movie, released in 2018, is the origin story of the franchise's titular character, Han Solo. In it, Solo embarks on a heist to pay off a debt to the criminal organization, Crimson Dawn. Along the way, he meets another original series character, Lando Calrissian, and his co-pilot, L3-37.[56] L3-37 is a custom-built pilot droid with a feminine voice and striking personality. L3, as she is nicknamed, is boisterous and sarcastic and a vocal supporter of robot equality. During the film, L3 causes a riot by turning off the restraining bolts, technological implants that limit a droid's ability to act independently, of the droid workforce during the heist. L3 is injured during the ensuing battle, and her hard drive is uploaded into the Millennium Falcon so the ship can utilize her navigational programming.[57]

L3 does not have seven seasons worth of episodes to pull from like Data and the Doctor, but her story provides an interesting facet of the AI problem to explore. Dr. Bruce Maddox wanted, but failed to replicate the technology found in Commander Data to create an army of robot servants, *Star Wars* has that army. Droids in Star Wars make up a slave labor class across the whole galaxy. Robots are bought and sold to function as house servants, farmhands, factory workers, and more; like in the example above, these droids are often equipped with restraining bolts to prevent disobedience. The restraining bolt was introduced in the very first *Star Wars* movie, being worn by both R2-D2 and C-3PO. When Luke Skywalker removed R2-D2's bolt, the droid ran away to find their old master, Ben Kenobi.

[56] "L3-37". *Wookipedia*.(Wiki), https://starwars.fandom.com/wiki/L3-37
[57] Ibid.

The droids mentioned here can reasonably be said to match or surpass the technological requirements this paper has put forth for intelligence and would be free in a compatibilist view. What is presented then, is the consequences of accepting the objection mentioned in chapter Three. If droid free will is different from human free will, are they still subject to moral standing? This world has decided no, and that because these beings are manufactured, they can be bought and sold and forced to work in conditions they express objection to. When treated well, droids are almost like pets. It would be like if your family dog had a Ph.D. in astrophysics and could fix all the household appliances themselves. In most cases though, they are akin to slaves, exactly what Captain Picard warned against in his own world. Droids are different, yes, but should that difference deny the pain and fear they express? No, and L3-37 has every right to fight to be recognized.

V. Chapter Summary

The cases presented in this chapter are hypotheticals, but they provide a needed context to examine how this technology may develop. Knowing we cannot separate the idea of freedom from moral responsibility, we must look at the question of free will through both a metaphysical and ethical lens. To forget one would be to deny the complexity of the problem before us. By saying a being is intelligent, you as the viewer are placing moral value on the choices being made. This is only possible if the choice is authentically their own, meaning they have to have the freedom to make these choices.

With how engrained AI technology is in our society, we cannot ignore the insight these stories can prove. Data is a being that would easily be identified as artificial. Should his internal constitution determine his validity and status or should his functional capacities? Given that we are nowhere near understanding our own internal constitution, we should not be so quick to judge.

Conclusion:

Through this thesis, I explored the current conversation surrounding Artificial Intelligence and argued that accepting that these machines will one day have free will is the moral choice. I also demonstrated how science fiction has informed and should continue to inform the conversations we have today about this technological advancement. The questions this paper addressed are 1) under what circumstances can machines be said to have intelligent thought; 2) does having intelligent thought mean they have freedom and moral responsibilities; and 3) do we as humans have any moral obligations when these machines reach said point?

Chapter One provided a basic introduction to the concepts of free will, artificial intelligence, explaining the significant terms and ideas that are central to this paper, such as compatibilism, general intelligence, and mental states. Chapter Two continued with this idea of mental states and functionalism and demonstrated why we should look to expand our conception of intelligence beyond a human-centric model, and how AI could one day meet it. Chapter Three reintroduced the notion of freedom and how a functionalist approach could be used to address potential objections to AI being free. In this chapter, I also argued why compatibilism should be accepted as a sufficient standard of free will, regardless of humanity's status as free or determined agents. Finally, Chapter Four examined major figures in science fiction, using their stories as thought experiments to demonstrate the moral significance of this problem.

My goal in writing this thesis was to demonstrate the need to look beyond a human-centered conception of freedom and morality when judging the cognitive abilities of artificial intelligence. Difference should never be the sole determinator for intelligence; it was dangerous when we demarcated each other this way and it will be dangerous if we demarcate potential life this way. I asked the reader to expand their notion of intelligence and to expand their conception

of freedom. I did so to, hopefully, demonstrate how difference should not preclude one from having either. We do not know what will become of Artificial Intelligence, but we do know what it is being used for now and where we wish to see it progress. Whatever it becomes, it will likely be different from us. Knowing difference is not a barrier to morality, we should endeavor to tread carefully in the creation of artificial intelligence, knowing at any point we could be dealing with a free entity worthy of moral value.

Bibliography

1. Turing, Alan. "Computing Machinery and Intelligence." In *The Philosophy of Artificial Intelligence,* Edited by Margaret A. Boden, 40-66. New York: Oxford University Press, 1990.

2. Kane, Robert. *A Contemporary Introduction To Free Will,* 1-12. New York: Oxford University Press, 2005.

3. Jacobson, Prof. Joseph Mayer. "Freedom of Choice and Artificial Intelligence ." B'or ha'Torah. Vol. 26 (Winter 2019): 46-57. https://borhatorah.wordpress.com/.

4.Hanley, Richard. *The metaphysics of Star Trek, 40-70.* New York: BasicBooks, A division of HarperCollins Publishers Inc., 1997.

5.Radavoi, Ciprian. *The Impact of Artificial Intelligence on Freedom, Rationality, Rule of Law and Democracy: Should We Be Debating It?.* In "Texas Journal on Civil Liberties and Civil Rights." Vol 25. 2. University of Texas at Austin Law Publications, 2020.

6.Cevik, Mustafa. *Will It Be Possible for Artificial Intelligence Robots to Acquire Free Will and Believe in God?*. Beytulhikme An International Journal of Philosophy." Vol. 7 No. 2 (Winter 2017).

7.Levin, Janet, *"Functionalism", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL* <*https://plato.stanford.edu/archives/fall2018/entries/functionalism/*>.

8. Burkett, Dan. *Mindless Philosophers and Overweight Globs of Grease: Are Droids Capable of Thought?* In "The Ultimate Star Wars And Philosophy." 231-240 Ed. Jason T. Eberl. John Wiley & Sons Ltd, 2017.

9. Boden, Margaret. *Artificial Intelligence: A Very Short Introduction*. Oxford: Oxford University Press, 2018.

10. Boden, Margaret A., ed. *The Philosophy of Artificial Intelligence*. Oxford , England: Oxford University Press, 1990.

11. Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions Using the Architecture of Choice*. New Haven, CT: Yale University Press, 2008.

12. Beebee, Helen. *Free Will: an Introduction*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2013.

13.Searle, John R. "Minds, Brains, And Programs." Essay. In *The Philosophy of Artificial Intelligence*, edited by Margaret A. Boden, 67–88. Oxford , England: Oxford University Press, 1992.

14.Skuza, Arek. “Amazon Artificial Intelligence - How Jeff Bezos Automates the Company.” Arek Skuza, November 10, 2020. https://arekskuza.com/the-innovation-blog/amazon-and-artificial-intelligence-in-retail/.

15.Churchland, Paul M. *Neurophilosophy at Work*. Cambridge, UK: Cambridge University Press, 2007.

16.Churchland, Paul M. "Functionalism at Forty: A Critical Retrospective." *The Journal of Philosophy* 102, no. 1 (2005): 33-50. Accessed January 29, 2021. http://www.jstor.org/stable/3655766.

17. Levin, Janet. “Functionalism.” Stanford Encyclopedia of Philosophy. Stanford University, July 20, 2018. https://plato.stanford.edu/entries/functionalism/#ConPlaFunThe.
18.“Intelligence.” Merriam-Webster. Merriam-Webster. Accessed January 29, 2021. https://www.merriam-webster.com/dictionary/intelligence.

19.Block, N. J., and J. A. Fodor. "What Psychological States Are Not." *The Philosophical Review* 81, no. 2 (1972): 159-81. Accessed January 29, 2021. doi:10.2307/2183991.

20. Beauvoir, Simone de. *The Ethics of Ambiguity*. Trans. Bernard Frechtman. New York, New York: Open Road Integrated Media, 2018.

21.“Computer AI Passes Turing Test in 'World First'.” BBC News. BBC, June 9, 2014. https://www.bbc.com/news/technology-27762088.

22.Best, Jo. “IBM Watson: The inside Story of How the Jeopardy-Winning Supercomputer Was Born, and What It Wants to Do Next.” TechRepublic. TechRepublic, September 9, 2013. https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/.

23.Dennett, Daniel. “When Philosophers Encounter Artificial Intelligence.” Essay. In *The Artificial Intelligence Debate: False Starts, Real Foundations*, edited by Stephen Richards Graubard, 283–96. Cambridge, MA: MIT Press, 1988.

24.“Data.” Memory Alpha. Wiki. Accessed January 29, 2021. https://memory-alpha.fandom.com/wiki/Data.

25.Singer, Peter, 1990, *Animal Liberation*, second edition, New York: New York Review of Books.

26. Snodgrass, Melinda M. “The Measure of a Man.” Episode. *Star Trek: The Next Generation* Two, no. Nine. Viacom, February 13, 1989.

27. “L3-37.” Wookipedia. Wiki. Accessed March 29, 2021. https://starwars.fandom.com/wiki/L3-37

28. Searle, John. *The Rediscovery of the Mind.* Cambridge, MA: The MIT Press, 1992.

29. Aristotle. *Nicomachean Ethics.* trans. Robert William Browne. Penguin Publishing Group, 2020.