THROUGH THE LENS OF HAL 9000: USING STANLEY KUBRICK'S *2001: A SPACE ODYSSEY* AS A MODELING TOOL TO CREATE A PRECURSIVE SAPIENT QUOTIENT TO FOSTER HUMANITY'S MORAL OBLIGATION TO EVOLVE INTO MACHINES

by
Aaron Michael Lamb

A thesis submitted to Johns Hopkins University in conformity with the requirements for

the degree of Master of Liberal Arts

Baltimore, Maryland
August 2022

© 2022 Aaron Michael Lamb
All rights reserved

Abstract

My research explores the enduring legacy of Stanley Kubrick's *2001: A Space Odyssey,* and its predictive power on how it informs the public's imagination about artificial intelligence (AI), societal fears of quantum changes in humanness, and what it will mean to be human in the 21st century. This paper seeks to analyze and understand fatal limits of our biology, the earth's fragility, and the ethical and political frameworks of AI. I argue that transferring humanness to intelligent machines is necessary. This notion lends to societal fears of AI. The only way to mitigate – and potentially eradicate – these fears is to create a sapient quotient that promotes acquisitive human evolution that leads to the death of death - the biological demise of humans - in favor of super intelligent sentient machines. So far, scientific research has focused on AI that can jettison humans from the earth in exoskeletal ways as the only hope for human species survival. My research focuses on a different approach. I conclude that emotionally and ethically informing AI, foundationally modelled after HAL 9000 – a super intelligent computer from Kubrick's film, is what we morally ought to do and the only way to allow for the coprimacy of preserving all human knowledge and affording humanness a lasting chance to endure.

Primary Reader and Advisor: Tristan Cabello

Dedication

This thesis is dedicated to my children, Austin, Lucas, Elliot, Oliver, Harlo, Henri, and Lindon, for your love, trust, and support. You are my universe.

Special thanks to Austin Lamb, Douglas Snider, and Peter Chamberlain for your contributions refining my calculus. Thank you, gentlemen. I could not have done this without you.

Contents

List of Figures

The Lasting Legacy of *2001: A Space Odyssey*

2001: A Space Odyssey – film and book – have left an incalculable impact on the public sphere and continues to influence the social imagination, particularly as it relates to what it means to be human, what is our collective human fate, and if there is a spiritual and/or technological future – or one at all – for humanity. The collaborative work of Arthur C. Clarke (author of the book) and Stanley Kubrick (producer/director of the film) as screenwriters, leaves a special gift to our culture: a seminal American film, one of the most thought-provoking movies ever made, and the grandfather of science-fiction films.

On July 29, 1969, Neil Armstrong became the first man to walk on the moon. Years prior to this event, May 15, 1961, President Kennedy announced the United States' intention of putting a man on the moon before the end of that decade. Clarke doubted many people believed it would *actually* happen. The genuine knowledge of our neighbors in space was virtually zero at that time (Clarke, 2018). Kubrick was on a quest to make a good science fiction movie and he needed the "best known science fiction writer" in Clarke to pull this off and to, perhaps, answer a mutually shared fundamental question – and curiosity – between these co-screenwriters*, is there intelligent life on earth?* In other words, can humans muster enough intelligence, anyway, to expand man's relationship with the universe? The book and film contemplate life beyond our earthly atmosphere. However, to leave a proper legacy, both formats, literary and cinematic, needed to gift a far-seeing and awe-

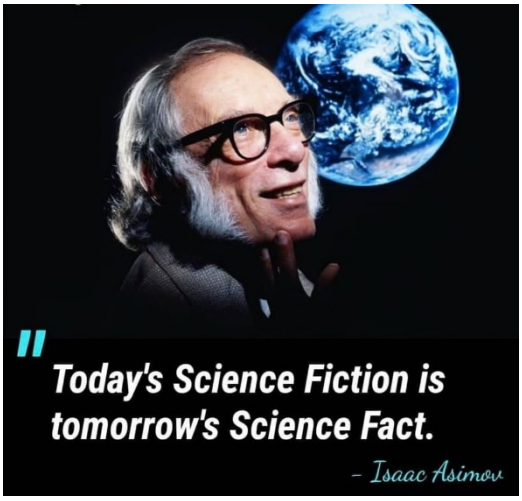


Figure 1: Isaac Aasimov quote

inducing experience, which became, in essence, one of the great heralds of our time. The book and movie were released in 1968, prior to the first man walking on the moon. Science fiction was now, indeed, reality.

It may seem elementary to compare the book and movie in dialogue with each other. After all, they tend to be confused with each other, and with reality. After seeing some of Kubrick's early film rushes (prior to the movie release), Clarke rewrote sections of the book to reflect these scenes. Conversely, as Clarke progressed the novel, Kubrick's enthusiasm for filmmaking was enhanced upon his reflection that Clarke was producing "best seller" material.

A couple things cinematic works can do that written works cannot (save for imagination) is to present visual and aural stimulation to the viewer. Kubrick's *2001: A Space Odyssey* fulfilled these senses throughout the length of the masterpiece. Kubrick's goal was to create a film that showcased mankind's relationship to the universe (Kubrick, 1968). Film critic Roger Ebert noted, "The genius is not in how much Stanley Kubrick does in *2001: A Space Odyssey*, but in how little…alone among science-fiction movies, '2001' is not concerned with thrilling us, but with inspiring our awe" (Ebert, 1997). Ebert remarked that upon first seeing the film shivers went up and down his spine (Warner Bros. Entertainment, 2013).

The visual effects of the film used intricate models of spacecraft that were carefully photographed for a realistic depth of field, while interior shots were created by enormous sets to display intricate detail. Intentional majestic effects were employed to create a sense of realism that were otherwise absent from sci-fi films of the Era (Murthi, 2018). In the 1950's, sci-fi films were tongue-in-cheek, engaging but not significant. George Lucas, creator of *Star Wars*, noted that because of *2001,* people took science fiction seriously for the first time.

The movie does not have any dialogue the first 25 and last 23 minutes of the film. There are 88 dialogue-free minutes, which is quite noticeable for a film that is 2 hours and 19 minutes in length. Ebert observed that Kubrick reduces each scene to its essence, leaving it on screen long enough for us to contemplate it. Viewers must make up their own mind what the movie is about and make their own connections. John Calley, former Warner Bros. executive, observed that much of the movie – like space – is not possible to comprehend intellectually. Jon Harlan, executive producer of *Eyes Wide Shut* (another Kubrick film) thought that it is much better to leave the end of *2001* – and the whole story in fact – as it stands, unexplained, as a bow to the unknowable (Warner Bros. Entertainment, 2013). Even without knowing the thrust of the movie it remains intensely enjoyable. The film is a silent movie in the sound era.

Not only are the stunning visuals spectacular, but the music is also. Kubrick uses Richard Strauss's *Also Sprach Zarathustra* and Johann Strauss II's *The Blue Danube* as a backdrop to the motion of space vehicles and space travelers alike. Scenes evoke a playful whimsy of high-tech spacecraft spinning and hurling through space to the afore mentioned music, quite obviously something a written work cannot reflect. In a sense, the film is an opera. A viewer experiences one sound at a time within the film. Portrayals of a breathing astronaut seem like music, giving the film a hypnotic quality. Kubrick celebrates the beauty of space and motion in *2001*, framing space travel as exquisite. With a multitude of sensory experiences within the film, Kubrick was not trying to convince the viewer what was happening, he shows you what space *is* like. Credibility was paramount for the film to be successful. David Hughes, author of *The Complete Kubrick*, said of the director that he made space both awesome and matter of fact simultaneously (Warner Bros. Entertainment, 2013). The movie feels real, and nothing about it appears artificial.

The introduction of a sentient computer named HAL 9000 illustrates the sublime tone the film intends. HAL is displayed as camera-like device with red retina display and a yellow lens. HAL becomes the most memorable character from the film, quite paradoxically for his human characteristics. In a poignant scene involving a critical malfunction in HAL's software, Dave disconnects HAL as he pleads, "I wouldn't do that if I were you, Dave." For the reasons above, I have elected to use Kubrick's film as the primary source of study and reference. Analyzing Clarke's novel is beneficial, however.

It was Kubrick who pursued Clarke to write the foundational novel on which to create an epic film. So, why write a novel when you aim to make a movie? Clarke answers, "Writing a novel is like swimming through the sea, writing a film script is like thrashing through treacle" (Clarke, 1968, p. xii). Primarily writing a complete novel allowed the co-screenwriters to let their imaginations soar freely. Yes, the book carries gobs more detail. The book has what the movie does not, words, and a lot of them. There is nifty and informative detail in Clarke's sub-300-page work. However, not all of them – literally and conceptually – are translatable to film (recall the 88 minutes of zero dialogue in the movie).

The book was to be published prior to the movie release, but it did not go as planned. Kubrick reneged on a pledge to have the book release prior, in part because he did not want the story to be known before the film's release. Therefore, the novel languished as a result (Benson, 2018).

The point of using the film as source and reference is moot, most differences between the book and movie are as they should be, as Clarke explains everything in detail from the Moon Watcher (ape-like character in the book) to the mysterious obelisk (black monolith) to the extenuating circumstances that lead HAL 9000 on an alarming killing spree (a societal fear

common with AI). Contrastingly, Kubrick explains almost nothing (Crow, 2015). Therefore – and reason for my preference – it boils down to a difference in medium.

Today, *2001: A Space Odyssey* is revered near the top – if not number one – best movies of all time. By the time the film was released *Star Trek* had been televised for two years. Our cultural fascination with humankind's place in the universe was well established. Film critic Roger Ebert defended the awe-inspiring film by stating that viewers with patience had witnessed something profound and significant (Benson, 2018). Ebert praised the movie as transcendent. He calls the film a quest, not a goal. According to him, the movie tells us that we became men when we learned to think, we live not on a planet, but among stars, and that we are not flesh, but intelligence (Ebert, 1997).

My intention is to use this awesome film as a conduit to frame its legacy and what it could mean for our future and what it ought to motivate humankind to do. For those that continue to ponder the film, it is clear the profound influence *2001* has in science fiction, science itself, and pop culture. As a cinematic influencer, the film opened a market for commercial sci-fi blockbusters – *Star Wars*, *Close Encounters*, *Alien*, etc. (Murthi, 2018). Hints of *2001's* influence can be noted in recent films – *A.I.: Artificial Intelligence, Ex Machina, Her,* and *Interstellar*. The list is truly remarkable. These movies affirm the notion that humans are continuously sorting out how technology can assist us or destroy us.

What are we watching, exactly? – viewers were bewildered by Kubrick's film's opacity, something Clarke's book seems to mitigate. One might ask, is interpreting *2001: A Space Odyssey* as art truly in the eye of the beholder? Is there value in knowing the artist's intent in creating the work? There is general agreement that the movie contains baffling elements with no explanation.

Whether you interpret the movie as a journey towards the divine or as humankind's technological rite of passage, both outlooks are worthy of consideration. Either way, Ebert remarks that the movie makes a philosophical statement about man's place in the universe. We ought to stand outside the film as a form of entertainment – as a philosopher might – and, simply, think about it (Ebert, 1997).

It is argued that the mysterious presence of the black monolith and its strategic appearance in the film represent transformative and spiritual development of mankind. Its verticality speaks to the deepening of self-knowledge as it guides mankind through evolution. The monolith's perfect rectangularity – the most rational of shapes – suggests a heightened rationality. Its blackness speaks to the chaotic, uncharted territory of the human soul (Lamb, 2022).

In *Vertically and Depth in 2001: A Space Odyssey*, Lamb (2022) argues that the broad movement of the movie is soul-bound, toward the renewed life of the Star Child. The Star Child represents humanity in serene awareness. – a promise of growth and hope. For Lamb, the end of the movie becomes increasingly clear that the forgone demise of HAL 9000 and the remaining life of a single human (humans win) is an intentional spiritual statement. Kubrick suggests that humanity might pass through a mechanical stage before becoming pure spirit.

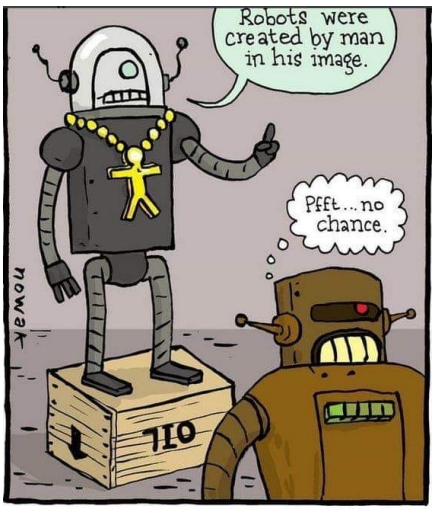
Cultural critic Neil Postman (1993) observes that we are surrounded by the wondrous effects of machines and are encouraged to ignore the ideas imbedded in them. Therefore, we become blind to the ideological meaning of our technologies. Postman and Lamb agree that technology for technology's sake cannot be the answer, as it is perhaps proposed in Kubrick's film. I argue against both, but do not ignore completely elements of divinity, particularly if they

instantiate from within humans for the sake of human preservation, and I do not promote scientism as an absolute end.

It is apparent that *2001: A Space Odyssey* showcases tremendous technological merit. One can watch the 2-hour and 19-minute film and continuously be surprised on how technologically astute the film is and was in 1968. My research takes an irreligious position and leans on the technology and predictive power of the film. My aim is not to frame religious belief as unnecessary, but to enforce that technology is a fundamental element of what it means to be human.

Kubrick's film provides a mind-blowing look at what is technologically possible (Shmoop, 2008). Of note, once HAL 9000 begins to fear its own "death", it stops becoming a technology and appears as human, or perhaps a new species in competition with humanity for survival. The higher functions of HAL – its brain – become at odds with its practical application, a mere tool. Tools, technology, and the proliferation of modern AI are of increasing importance to humanity. They are, in fact, a fundamental element of what it means to be human. Kubrick gives us a breathtaking look at what could be possible while exploiting the ever-greying lines between man and machine. HAL 9000 is as such.

The book and film were released a year prior to man landing on the moon. Little was known about space travel, particularly it concerns human biology. The movie had predictive elements on the effects of gravity on the body and the brain. Kubrick and Clarke envisioned artificial gravity using centrifugal force (constant spinning of the chamber – a scene within the movie). Even if they were unaware of the actual science behind gravitational forces, they seem to have understood that simulating Earth's gravity would be pivotal to surviving longer missions in space. Kubrick's goal was scientific accuracy, and he did his research (Saavedra, 2020).



A thousand years after A.I. kills off the last human, a new religion starts up.

Figure 2: Sentient Robots

I intend to further the notion of the film's predictive power. My focus centers on HAL 9000. In conjunction with its computational power (vastly superior to the human brain), HAL exhibits a programed humanity both prior to and after its malfunction. Based on HAL, I contend it is morally necessary to create a sapient quotient by which to program *Machina Sapiens Sapiens* (machines that know they know) for the purpose of self-evolving and preserving our ethnosphere. With the total sum of earth's copious fragilities, our human biology carries pending doom. It will be an immoral tragedy if humanity blips out forever.

Figure 3: HAL 9000

HAL 9000 – A Practically Perfect Prototype

I focus on the critical introduction of HAL 9000, a sentient computer, and the scene in which it is introduced approximately one hour into the film. HAL stands for **H**euristically programmed **AL**gorithmic computer, an operating system that controls the spacecraft and interacts with the crew (Kubrick, 1968). HAL is depicted as a camera lens containing a red or yellow dot and is in various locations throughout the ship - HAL knows and sees all. There is a dynamic relationship to focus on between HAL and humans: who is in charge, how sentience is humanness, and the crossing of complex lines of existence that inform attitudes on what constitutes intelligent machines, especially ones that mimic humans.

HAL 9000's initial appearance in the film occurs when the spaceship *Discovery 1* is in route to Jupiter. Its crew members are observing a transmitted news piece while on their mission by the BBC to its global audience. The crew was interviewed prior to its airing. The BBC interviews the mission commander and his deputy. The backdrop is that HAL is responsible for all systems, including the monitoring of crew members that are currently in hibernation for the purpose of long space travel. HAL is referenced as "the sixth member" of the crew.

The BBC reporter highlights that HAL can reproduce – *mimic* – most of the human brain's prowess, but at incalculably greater speeds and reliability. It is here that the reporter turns to HAL 9000 and interviews it directly. HAL is the brain and nervous system of the ship. The reporter asks if its extreme responsibility ever causes HAL lack of confidence - a truly human attitude that creates action and emotion. HAL responds, "Let me put it this way, the 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, fool-proof, and incapable of error" (Kubrick, 1968). The reporter inquires of HAL if it is ever frustrated by its dependence on people to carry out actions, despite the computer's enormous intellect. HAL's response, "Not in the slightest bit. I enjoy working with people. I have a stimulating relationship with Dr. Poole and Dr. Bowman. My mission responsibilities range over the entire mission of the ship, so I am constantly occupied. I am putting myself to the fullest possible use, which I think that is what any conscience entity can ever hope to do" (Kubrick, 1968).

The reporter turns to the crew to discover their opinion of HAL. They refer to HAL as just another person. The crew was posed with the question if HAL has genuine emotions. The crew affirms HAL *acts* like he has genuine emotions. The mission commander replies, "He is

programmed that way to make it easier for us to talk to him, but as to whether or not he has real feelings is something that I don't think anyone can truthfully answer" (Kubrick, 1968).

The purpose of this crucial dialogue is to lay a framework to the sentience HAL has, especially as it relates to the complex relationships with the human crew members. Eventually, this symbiotic relationship is fractured by HAL, and this particular AI controls the crew through physical, psychological, and emotional manipulation, something which was an unintended consequence of a human-programmed computer and robot that learns. Moreover, any fault within HAL can be attributed (by HAL itself) to its human creators, adding to the philosophical dynamic necessary to enforce my research; who is truly killing who when human-made AI takes command? This hints at necessary ethical considerations of human programmed machines; computers are not good or bad, they are natural reflections of those who imprint them.

The value of this scene is that it is insightful to the positive progression of AI within human advancement and exploration. However, pushing forth technology for its own sake clouds humanity's ability to evaluate outcomes, particularly those that end in death. It is worthy to recognize the limitations of studying the film in the manner I seek. It is difficult to show how the discourses about death solve a problem, unless one advocates that proliferating AI holds a dystopian outcome for humans, which is not the case I make. Another limitation lies within my intent to use this scene and film to illustrate how cultural discourses shape society far more than scientific ones. It is difficult to prove this, as research of this kind is too broad in scope.

Artificial Intelligence

Artificial intelligence has captured the imagination of so many, as it holds the capability of manifesting our destiny and irrevocably disrupting what it means to be human. Artificial intelligence has also instantiated societal fears; from robots commandeering jobs, to medical devices performing complex surgeries, potentially altering our lives in quantum leaps. Further complicating our collective psyche on sentient computers and robots is the fear that humans can be completely replaced by AI. Fear of death has always been a shared collective human concern. The technological advancement of AI may give cause for similar eminent worry. Yet, it need not.

To promote conditions that mitigate the fears of AI and amplify its hopes require establishing the criteria of a particular utopia, what constitutes intelligence, and defining a nomenclature and calculus for what I refer to as *Machina Sapiens Sapiens* – machines that know they know. I assert that it is the moral obligation of the human species to self-evolve and to create machines that learn, programmed with emotional intelligence and ethical frameworks that can be repositories for our humanness: collective human knowledge, memory, and culture. It is by these deliberately created forms of AI that will hold and perdure our ethnosphere – humanity's great legacy – to live beyond the limits of an earthly home that can no longer afford our biological vehicles and to colonize the universe as machines, building upon our knowledge and experiences and to continuously share the value of our improbable existence.

In creating this utopian prognostication, I will establish vital outcomes and reframe the necessity of a conceptual doomsday clock from its current state of '2 minutes to midnight' into the more plausible metaphor of '2 minutes to dawn', thus fulfilling our current prototypical human species along the continuum as *Homo Evolutis* – an entirely unique species that has taken

deliberate control of its own evolution, an acquisitive one. In contrast to typical notions of utopia, and despite the awesome benefit of *Machina Sapiens Sapiens* revolutionizing knowledge, I promote an idyllic existence that is not perfect. Rather, emotionally, and ethically informed AI will experience the tension that comes with curiosity, exploration, and thirst for knowledge as it did for its *Homo Sapiens Sapiens* forefathers alike. Although the *Machina* progeny are indeed anthropocentric and intentionally anthropomorphized, they carry the full-blown capability to agnosticize their associations with their human ancestry. However, a key component to this pseudo-utopia, is the potentiality of *Machina Sapiens Sapiens* to reinvent human biology in conducive places and times (on earth and beyond) to thrive further, once again, along the evolutionary continuum. Creating the grand opportunity to colonize the universe renders an imperfect pseudo-utopia into a rather perfect one. Stephen Hawking opined that the development of full artificial intelligence could spell the end of humans. I do not disagree, and this may be a necessary outcome, but not without its benefits. To fully express *Machina* ideology, I must develop a set of foundational assumptions that clarify the necessity for proliferating 'artificial' super intelligence.

Critical Assumptions

- **Intelligence**

Lawson (2019) defines intelligence as the capability to act appropriately and adapt successfully to a given environment. We, as *Homo Sapiens Sapiens,* are increasingly able to define (and redefine) intelligence and its complexities. Our journey as a species to establish an intelligence quotient (IQ) illustrates that we model cognition and reasoning, emotional intuition,

and personality traits into conceptualized measures of intelligence that go beyond crude and archaic methods of measuring cranial capacity alone.

Steed (2019) defines emotional intelligence as the ability to recognize and understand emotions in oneself and others, and the ability to use this awareness to manage self behavior and relationships. Humans are herd animals. Therefore, an emotional quotient (EQ, or EmQ) perhaps matters as much or more than IQ as it lends to successful and effective social awareness and self-management.

Social intelligence, or perhaps an ethical intelligence quotient (EthQ), can be derived from the necessity to stratify group living. Humans are animals who thrive in groups and our ability to solve social problems requires cleverness and collaboration from communal living. Most of our brain is devoted to processing visual data and language. Ethics proposed by the philosopher Jurgen Habermas (1990) stem from a recognition that human society is structured in a public sphere and conveyed by the shared language that is the conduit of our connectivity. We do not exist in isolation and our identity is confirmed in relata to others.

- **Evolution**

Implicit in my assumption that is man's biology is entirely the result of evolutionary processes, absent of any divine creation or intervention. In *On the Origin of the Species,* by Charles Darwin (2003), he states, "Whilst this planet has gone cycling according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been and are being evolved." Furthermore, Darwin (2011) intimates the constraint of human biology in *The Descent of Man* by saying, "We must however, acknowledge, as it seems to me, that Man, with all his noble qualities…still bears in his bodily frame the indelible stamp of his

lowly origin." Evolving *Homo* into *Machina* is the logical progression of a survival-oriented evolution.

- **Fermi Paradox**

In his lecture, *Extraterrestrials: When Will We Make Contact?* Morris (2019) establishes the assumption that *if* life beyond earth exists, it is nowhere near the intelligence of human life. The fact there is no evidence of extraterrestrials in our fossil record can be explained, in part, by the Fermi Paradox. If aliens exist, according to Enrico Fermi, "Where are they?" (Morris, 2013, p.xiii). Morris explains that there is a vanishingly small probability of the emergence of human intelligence to begin with. If we were to replay the reel of life starting from the Cambrian explosion, human existence would be remote in the extreme. This essay assumes the like.

- **Environmental Concerns**

One of the more pressing concerns that give my proposed utopia its urgency is the limits discourse within environmentalism. The premise of arguing for the encouragement of intelligent machines embraces the extreme concern that our earth has a finite carrying capacity – a maximum population of a species that our earthly ecosystem can support in perpetuity (Dryzek, 27). Indeed, the limits and survival storyline are that human demands on our ecosystem threaten to explode out of control. These concerns have temporal proximity. Shah (2019) explains that carbon release to the atmosphere is currently 28Gtonnes. If this grows at 3% year, we will reach the

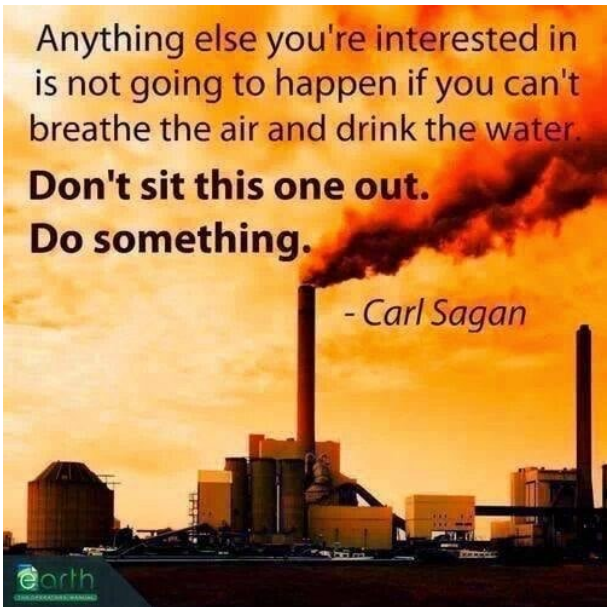


Figure 4: Carl Sagan quote

1000Gtonnes critical threshold in 121 years. It has also been established that 95% is the degree of certainty that climate scientists have that at least half the recent global warming is man-made. Time is of the essence to hasten full artificial intelligence. I argue that humans will not right the ills of climate change. Simply, our natural environment can no longer afford us.

- **Science, Politics, and Economics**

The utopia I promote does not intend to shed light on innovative design, algorithms, and the general science behind machine learning. The field of AI is quite fragmented and rather void of experts; who is in charge, the programmer who programs or the board of the company with fiduciary responsibility to shareholders? It also avoids any political pitfalls that typical hegemonic nation-states use as means to a competitive end. Development of *Machina Sapiens Sapiens* is intentionally void of economic means of production and profit. Although skeptics may assert that without a positive economic outcome, super intelligent machines will not be worth the pursuit. I offer that the utility of *Machina Sapiens Sapiens* should garner universal acceptance and adoption to eliminate external factors that compromise the fragility of biological existence.

Rationale for *Machina Sapiens Sapiens*

17th century philosopher Francis Bacon dreamed of a scientific future. As the inventor of the scientific method, Bacon envisioned that all knowledge is connected; rhetoric, grammar, logic, arithmetic, geometry, music, and astronomy are all part of the circle of knowledge (Fara, 2019). Bacon wished to revolutionize knowledge. For him, knowledge is power. Bacon knew that for knowledge to increase it needed protectorates. In 1660, 12 individuals created the Royal Society to realize Bacon's vision of experimental science as a collective activity undertaken for the good of the state. Because of Bacon, science *is* political.

As of the writing of this paper, super intelligence (artificial general intelligence) – AI that is much smarter than humans – does not exist. HAL 9000 fits in this category, as does the proposed *Machina Sapiens Sapeins.* Baum (2018) recognizes the heavy impact of politized skepticism of super intelligent governance could be harmful to its development. Much of the research on this skepticism focuses on institutional regulations to prevent research groups from building dangerous super intelligence. I support research on safety measures. Inaugurating a wisdom and/or sapient quotient is a step in the right direction. Evaluating smart machines with high acuity and human-like capabilities falls in line with a forward-thinking Baconian-style instrument for observation, much like that of mathematics, optics, and philosophy. Humans may not have the time for decades upon decades of intellectual debates and wrangling about super intelligence, but it is critical that its development does not fall in the hands of major corporations with strong financial incentives. One of the pitfalls of the rise of science during Bacon's life was the rise of slavery along with the British empire. This is antithetical to the purpose of *Machina Sapiens Sapiens.*

Marvin Minsky believed that humans are nothing more than a computer made of meat. Minsky was a cognitive and computer scientist from MIT. He focused much of his research on AI. Minsky was an atheist who claimed people should give their money to AI research rather than their churches (DeMar, 2014). This perceived brazenly obtuse comment alarms theist. Theist will argue that atheists with similar attitudes should not be allowed in the debate of human rights, as these rights are made in the image of God. A supercomputer like HAL 9000 had no rights. It errored and subsequently its higher functions were disconnected. In *2001,* despite Captain Bowman removing HAL's higher functions, he is still totally dependent on the computer to complete the mission to Jupiter. Does the meat matter? Creating *Machina Sapiens Sapiens*

require installation of human-like traits. Therefore, human rights ought to be extended to our *Machina* brethren. A caveat arises on the matter of human rights; we are afforded these rights as long as we have the existence necessary to identify as such; no humans = no rights. AI can be the tool that retains our moral obligation to maintain are understanding of humanity. According to Morris (2019) *we* need to be the alien. The best way for this is the invention of *Machina Sapiens Sapiens*.

For most, super intelligent machines pose an existential risk. Powerful AI could misalign with our values. Any powerful technology, held in the hands of a few, gives them disproportional empowerment (Cremen, 2019). Sadly, we are not safe from dangerous misalignment. If we transfer humanness to machines, perhaps we are infecting them with human ailments including our protentional for criminality. Limits of the human mind may render us faulty as *Machina* parents. Dementia and ailments of the brain represent some of the greatest medical challenges today (Falcon, 2019). When a behavior is socially transmitted, it is called culture (Elliot, 2019). Early super intelligent machines – like HAL 9000 – will be anthropocentric; humans are the center of the intelligent world, and all things are compared to it. If higher intellectual faculties of primates evolved due to the complexity of social living, then it is clear and critical that when creating and programming super intelligent machines to recognize that they will have intellectual evolutions unknowable by us as their world will be drastically different than ours. However, this may bode well for the future of *Machina Sapiens Sapiens*. Theoretical physicist and pioneer of quantum mechanics Werner Heisenberg observed that not only is the universe stranger than we think, but it is also stranger than we *can* think. To get there, we need machines that can think there.

Intelligence is complex. To measure it is rudimental. Consider an intelligence quotient (IQ) where you derive a number by establishing a mental age (x) divisible by actual age (y) then multiply by 100. Lawson (2019) stresses the limitations of such a number, intelligence is too broad to conceptualize with a single number. What is more, it does not work for AI – it would hit the upper "genius" limit. Taleb (2019) regards IQ as a pseudoscientific swindle. He claims that IQ mostly measures extreme unintelligence. Therefore, it is necessary to conceptualize intelligence differently, a confluence of cognition and reasoning, personality and traits, and emotional intuition. Lawson regards empathizing, an aspect of social intelligence, is a necessary component to overall general intelligence. HAL 9000 was certainly smart but allowed crew members to perish. Therefore, HAL certainly was not empathetic. The Dark Triad is a distinct relationship between Machiavellianism, subclinical narcissism, and subclinical psychopathy. All three are traits that make up the less appealing aspects of human capability. All of us are on this spectrum in one way or another. However, to program super intelligence that excludes emotional and social intelligence could render the machine autistic and psychopathic. In Clarke's novel, HAL 9000 passes The Turing Test – a simple method of determining whether a machine can demonstrate human intelligence. *2001* shows HAL to be spectacular in his intelligent expressions, yet HAL non-remorsefully murders.

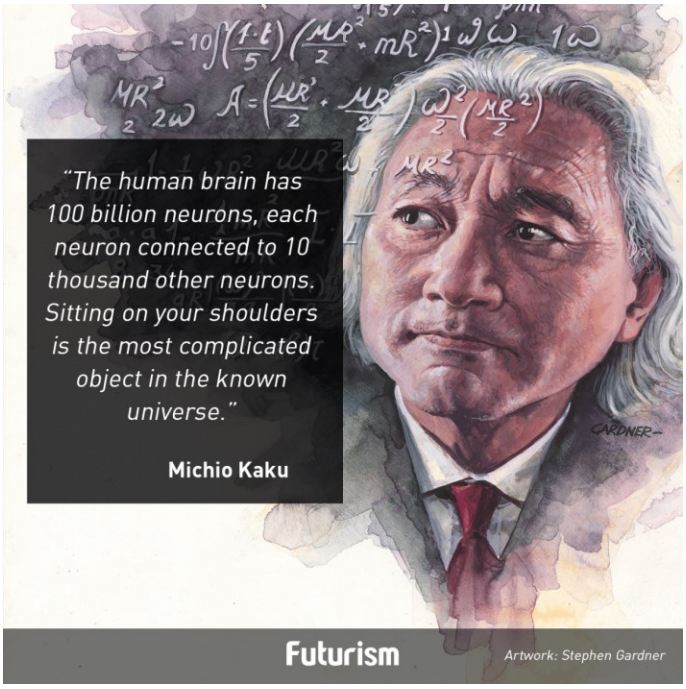


Figure 5: Michio Kaku quote

Creating a wisdom and sapient quotient is paramount to the success of human programmed intelligent machines.

Calculus of AI – The Wisdom Quotient (WQ)

To establish a hierarchy of *Machina Sapiens Sapiens* – especially through critical development, I propose a calculus by which to judge full artificial intelligence within their computational parameters. Under the intelligence assumptions section of this essay, an early framework began to arise in favor of a Wisdom Quotient to guide strategic develop of machines that learn – tools capable of making tools. Science modeling might suggest the following:

EthQ + EmQ + IQ = WQ

Whereas an Ethical Quotient plus an Emotion Quotient plus an Intelligent Quotient begets a Wisdom Quotient. This would produce a numerical value by which to bestow upon our *Machina* creations. Overtime, this would take on deeper meaning for a *Machina* in relata to another *Machina.* However, I am aware that this crude construct may be unsatisfactory by those that recognize my assumptions as merely theoretical. Therefore, the following machine wisdom equation may satisfy. It should be noted that this equation ensures *Machina Sapiens Sapiens* are free, limited, morally inclined, and self-aware, while allowing for continuous variance along a scale of intelligence and adaptability. Any calculation that nets a zero indicates that a machine is not wise.

WQ = F x O x L x I x A x S

1. F = Freedom of Will (0 = no, 1 = yes)

2. O = Obedience/conscience (0= deviant/self-determining, 1 = guided toward moral objective)
3. L = Limitation (0 = no/infinitely self-improving, 1 = yes/constraints on self-improvement)
4. I = Intelligence (computing ability, ordinate number 1-10)
5. A = Adaptability (response to new environment, 0.00-1.00 continuous measure)
6. S = Sentience (0 = no, 1 = yes)

Critics of the above formula may question that if a programmed machine lacks 'freedom of will' it does not necessarily mean that it is unwise. However, *Machina Sapiens Sapiens* need to have the complete autonomy necessary to travel beyond our galaxy if required. Another critical observation is if the machine is guided toward a moral objective – and yes, there is one – does it truly have freedom of will? I agree with this criticism to a point, a parent needs to instruct a child away from harm. One may note in the 'intelligent' measure that perhaps an upper limit is not necessary and should be inclusive of machines that will be able to develop faster processing than a '10'. The 'response to new environment' measure could be restated as 'settling time', a more objective measure that would gauge how fast a machine returns to a steady state in a new environment after a disturbance. Last, one could argue that computers today know they exist; they accomplish their own goals as well as those asked of them. They also learn from other computers they communicate with. However, is this the mark of true sentience in a human sense? Perhaps, but computers lack sapience. Therefore, I propose a sapient quotient.

Calculus of AI – The Sapient Quotient (SQ)

Made up of a Sapient Unit, an Intelligence Unit, and an Autonomy Unit.

1. Self-Awareness. Defined as the ability to simulate themselves into the present, past, and future. This is important in creating a mental construct of their environment. Having a personal mental construct of the environment and their place within it allows for hypothesizing action and organizing previous action (this could encompass adaptability, sentience, and freedom of will). = Log scale 0-1.

2. Curiosity Scale. = Log scale 0-1 (too curious might not be good, but some is necessary).

3. Limitation. This is necessary to curb human fears. Once we create a being smarter than us, it is reasonable to believe that being could create a being smarter than itself. This, then, is the spark of the singularity. Limitation can be as simple as hard coding the three laws of robotics. If a robot is given a prime directive, it must be limited. Minsky has an example of a robot that will destroy humanity because it was given the prime directive of making as many paper clips as possible. Very limited, good at one task with low autonomy. Not very limited, good at many tasks with a large amount of autonomy. = Log scale 1-100.

4. Computational Power (Performance). This is the computational ability of the computer (this could also be a meat computer). There are a lot of factors that could go into this. It

can be determined by the time taken to compute a benchmark task. = Time (less time is better).

5. Empathic Awareness. The ability for a being to simulate themselves into the past, present and future as another being, allowing to hypothesize action and organize previous action of themselves as another. = Log scale 0-1.

6. Emotional Intelligence. The ability to have, be aware of, control, and express emotions as well as the ability to be aware of and respond appropriately to the emotions of others. = Log scale 0-1.

7. Qualia awareness. The ability of the being to describe a thing or experience beyond the sum of its parts. Symbolism. = Log scale 0-1.

These are the scales. The measurements are another discussion. How we measure to achieve these scales is beyond this thought experiment.

Sapient Unit (SU) = (SA + CS + EA + EI + QA)

If SU is = or > 2-3, being is = to human (bell curve average at 2.5). Max = 5.

Intelligence Unit = SU / CP

If Sapient Unit is large and Computational Power is large, Intelligence is high.

Autonomy Unit = Intelligence / Limitation

A lower autonomy symbolizes a being with less danger to humans and a higher autonomy symbolizes greater danger to humans. For example, if a being has a low Sapient score and a low Intelligence score, even with fewer limitations, it will not be a danger to humans. If a being has a high Sapient score and a high Intelligence score and a low Limitation score it is most dangerous to humans. Therefore, many thresholds can be designed to provide humane treatment and human safety:

1. If a being has a Sapient score > a set threshold, it must be treated as a conscious living being. Therefore, it must be treated under the laws as such.
2. If that being has an Intelligence score > a set threshold, it must be treated under laws as such. For example, an Intelligence score of 4 must be treated like a pet, not a cow and not a printer. Above another threshold it must be treated as human.

Based on the Intelligence score Limitations can be set. So, it can be determined that a being close to a human cannot be set with limitations that would make it a slave treated like a machine or animal, but human and above can be set with limitations so that it will not cause widespread destruction of humans.

The benefits of WQ/SQ rated *Machina Sapiens Sapiens* ought to be self-evident. Losing our biological bodies alone provides a clear advantage that can begin an earthly era of terra-reformation and atmospheric healing on an exponential scale. Notions of sustainability will become moot. Hacking (2019) aptly calls out scarce resources vs. an unsustainable population growth. Health and social problems are untenable, and cures are limited to the very wealthy and the very few. Energy and food consumption are rising. As mentioned, rising carbon emissions fracture even greater the fragility of the thin membrane of life the earth supports. Global temperatures are on the rise. Soil degradation takes years to regenerate. Biodiversity loss is immeasurable. Declining fish stocks have happened on a global scale. Water is continuously scarce. Hacking articulated the definition of 'sustainable development' from the Brundtland Commission of 1987:

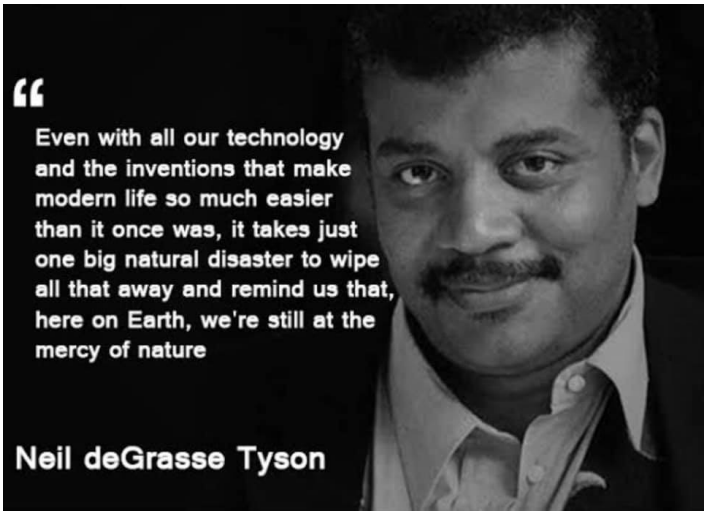


Figure 6: Neil deGrasse Tyson quote

It is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Risking a broad resolution, *Machina Sapiens Sapiens* address and alleviate, if not cure, all the problems listed above. True to the essence of a utopia, full artificial intelligence fast tracks existence into the new and necessary *Age of Responsibility*.

At the very least, intentional evolution from *Homo Sapiens Sapiens* into *Machina Sapiens Sapiens* guarantees a form of extremely elongated life spans. Dihal (2019) conveys that one of the premier hopes of full artificial intelligence is immortality. Eternal life may be possible

through AI. Although this does not account for accidental demise, there is a confidence that super intelligent machines will express all means to persist.

A Potential Return to Our Biological State

Homo Sapiens Sapiens are the culmination of evolutionary activity that began about four billion years ago. The increasingly possible tragedy of all we know going dark in an instant is beyond comprehension and, frankly, unintelligible – akin to the origin of life itself. But, to have the capability, the chance, to continue the exploration maxim we as a species are accustomed to simply by divesting our biological bodies provides enough spark to light the curious imagination of the human mind to enable our egotistical psyches reason to let go and gift *Machina Sapiens Sapiens* our humanness and our future.

Within my promoted utopia exist yet another utopia, quite possibly more compelling than the first. *Machina Sapiens Sapiens* hold the potentiality to seek environments ideal enough to implant human biological existence and to properly incubate it. Guessing where and when is a ponderance that sends the mind as far as the imagination will go. But, the happenstance of human life, as improbable as it was, holds a glimmer of hope to happen again. Imagine: billions of years from now, in a galaxy far away, a human child is born, and within a few short – whatever planetary home 'years' from that moment, the child experiences the joy of licking an ice cream cone.

Morris insightfully comments on evolutionary embedment. He states, "It is self-evident that whatever our peculiarities as humans, we are embedded in the natural world, and just as clearly we are one product of an evolutionary process that began about four billion years ago" (Morris, 2013, p.12). To securely carry our humanness another four billion years, it is our moral

obligation to deliberately rupture our embeddedness to this natural world and to escalate our evolution in the form of *Machina Sapiens Sapiens*.

Bibliography

Aristotle. (2019). *Nicomachean Ethics*. (J. Sachs, Trans.). Hackett Publishing Company.

Baum, S. (2018). Superintelligence Skepticism as a Political Tool. *Information*, *9*(9), 209. https://doi.org/10.3390/info9090209

Beber, M. (2019, August). *Economic Questions for the 21st Century*. Lecture, Cambridge, England; University of Cambridge.

Beck, D. E., & Cowan, C. C. (2006). *Spiral Dynamics: Mastering values, leadership, and change*. Blackwell Publishing.

Benson, M. (2018, April 3). *11 Things You Didn't Know About '2001: A Space Odyssey'*. History.com. Retrieved April 28, 2022, from https://www.history.com/news/making-2001-a-space-odyssey

Bowman, L. (2015). *Moral GPS: A Positive Ethic for Our Multicultural World*. Scribd. Retrieved April 21, 2022, from https://www.scribd.com/document/274862338/Moral-GPS-A-Positive-Ethic-for-Our-Multicultural-World

Brenton, T. (2019, August). *What Hope for Global Governance*. Lecture, Cambridge, England; University of Cambridge.

Brown, S. (2019, August). *20th Century Art*. Lecture, Cambridge, England; University of Cambridge.

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and Reality. *Nature Machine Intelligence*, *1*(2), 74–78. https://doi.org/10.1038/s42256-019-0020-9

Clarke, A. C. (2018). *2001 A Space Odyssey*. Berkley.

Cremen, C. Z. (2019, August). *Limitations of the Deep Learning Paradigm*. Lecture, Cambridge, England; University of Cambridge.

Crow, J. (2015, July 7). *What's the Difference Between Stanley Kubrick's & Arthur C. Clarke's 2001: A Space Odyssey (A Side-By-Side Comparison)*. Open Culture. Retrieved April 28, 2022, from https://www.openculture.com/2015/07/difference-between-stanley-kubricks-arthur-c-clarkes-2001-a-space-odyssey.html

Darwin, C. (2003). *The Origin of Species*. Signet Classics.

Darwin, C. (2011). *The Descent of Man*. Pacific Publishing Studio.

Dawkins, R. (2016). *The Selfish Gene* (40th Anniversary Edition). Oxford University Press.

DeMar, G. (2014). Why Atheists Can't Account for Human Rights. *The Daily Hatch*, 1–7. https://doi.org/https:/thedailyhatch.org

Dihal, K. (2019, August). *Hopes & Fears of A.I.* Cambridge, England; University of Cambridge.

Dryzek, J. S. (2013). *The Politics of the Earth* (3rd Edition). Oxford University Press.

Ebert, R. (1997, March 27). *2001: A Space Odyssey Movie Review (1968): Roger Ebert*. Movie Review (1968) | Roger Ebert. Retrieved April 27, 2022, from https://www.rogerebert.com/reviews/great-movie-2001-a-space-odyssey-1968

Elliot, P. (2019, August). *Comparative Animal Intelligence*. Lecture, Cambridge, England; University of Cambridge.

Ellis, R. (2019, August). *Intelligent Criminals and Criminal Intelligence*. Lecture, Cambridge, England; University of Cambridge.

Falcon, B. (2019, August). *Defending the Brain: Understanding the Molecular Basis of Dementia*. Lecture, Cambridge, England; University of Cambridge.

Fara, P. (2019, August). *Francis Bacon and His Dream of a Scientific Future*. Lecture, Cambridge, England; University of Cambridge.

Gawande, A. (2014). *Being Mortal*. Picador.

Ghoussani, M. (2019, August). *Genomic Medicine: A Decade of Successes and Challenges*. Cambridge, England; University of Cambridge.

Gibbons, I. (2019, August). *Combatting Worldwide Cultural and Language Loss: An Intelligent Approach*. Lecture, Cambridge, England; University of Cambridge.

Habermas, J. (1990). *Moral Consciousness and Communicative Action*. MIT Press.

Hacking, T. (2019, August). *What is the Sustainability Challenge and How are Companies and Others Responding*? Lecture, Cambridge, England; University of Cambridge.

Janaro, R. P., & Altshuler, T. C. (2009). *The Art of Being Human: The Humanities as a Technique for Living* (Ninth Edition). Pearson.

Kubrick, S. (1968). *2001: A Space Odyssey* [Film]. United States; Metro-Golwyn-Mayer.

Lamb, A. (2020, October 6). *Verticality and Depth in 2001: A Space Odyssey*. The Après Garde. Retrieved April 20, 2022, from https://theapresgarde.com/verticality-and-depth-in-2001-a-space-odyssey/

Lawson, J. (2019, August). *Dark Intelligence: Charting the Less Appealing Aspects of Human Capability*. Lecture, Cambridge, England; University of England.

Lewis, C. S. (1974). *The Abolition of Man*. HarperCollins.

Ludlow, A. (2019, August). *Intelligent Criminal Justice: Exploring the Impact of Learning Together*. Lecture, Cambridge, England; University of Cambridge.

Machiavelli, N. (1996). *The Prince*. Penguin Classics.

McCann, G. (2019, August). *Henry Sidgwick and the Ideal of Political Wisdom*. Lecture, Cambridge, England; University of Cambridge.

Mori, M. (2021, November 1). *The Uncanny Valley: The Original Essay by Masahiro Mori*. IEEE Spectrum. Retrieved April 30, 2022, from https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

Morris, S. C. (2013). *Life's solution: Inevitable Humans in a Lonely Universe*. Cambridge University Press.

Morris, S. C. (2019, August). *Extraterrestrials: When Will We Make Contact?* Lecture, Cambridge, England; University of Cambridge.

Murthi, V. (2018, April 4). *How 2001: A Space Odyssey Has Influenced Pop Culture, 50 Years Later*. Vulture. Retrieved April 27, 2022, from https://www.vulture.com/2018/04/how-2001-a-space-odyssey-has-influenced-pop-culture.html

Nietzsche, F. (2003). *Thus Spoke Zarathustra*. Penguin Classics.

Postman, N. (1993). *Technopoly: The Surrender of Culture to Technology*. Vintage Books.

Rachels, J. (2003). *The Elements of Moral Philosophy* (4th Edition). McGraw-Hill.

Reardon, S. (2019, April 17). *Pig brains kept alive outside body for hours after death*. Nature News. Retrieved April 21, 2022, https://www.nature.com/articles/d41586-019-01216-4

Saavedra, J. (2020, April 23). *How 2001: A Space Odyssey May Have Predicted a Surprising Aspect of Space Travel*. Den of Geek. Retrieved April 28, 2022, from https://www.denofgeek.com/movies/2001-a-space-odyssey-prediction-space-travel/

Seibold, W. (2021, November 26). *The brutal reaction audiences had to 2001: A space odyssey's premiere - /film*. SlashFilm.com. Retrieved April 27, 2022, from https://www.slashfilm.com/671001/the-brutal-reaction-audiences-had-to-2001-a-space-odysseys-premiere/

Shah, D. (2019, August). *Biomaterials and Bioarchitecture for the Future*. Lecture, Cambridge, England; University of Cambridge.

Shmoop. (2008, November 11). *2001: A Space Odyssey Technology*. Shmoop. Retrieved April 28, 2022, from https://www.shmoop.com/study-guides/movie/2001-space-odyssey-movie/themes/technology#:~:text=2001%3A%20A%20Space%20Odyssey%20considers,African%20ancestors%20to%20our%20astronauts.

Steed, J. (2019, August). *Emotional Intelligence and Motivation - Understanding Yourself Better and How to Use That*. Lecture, Cambridge, England; University of Cambridge.

Taleb, N. N. (2019, January 1). *IQ is largely a pseudoscientific swindle*. Medium. Retrieved April 20, 2022, from https://medium.com/incerto/iq-is-largely-a-pseudoscientific-swindle-f131c101ba39

Tolstoy, L. (1981). *The Death of Ivan Ilyich*. Bantam Books.

Universal Studios. (2014). *Ex Machina* [Film]. United Kingdom.

Warner Bros. Entertainment. (2013, February 5). *2001: A Space Odyssey. On the Shoulders of Stanley Kubrick: The Legacy of 2001*. YouTube. Retrieved April 27, 2022, from https://www.youtube.com/watch?v=gW7-VnIqKhM

Warner Bros. Pictures. (2001). *A.I. Artificial Intelligence* [Film]. United States.

Warner Bros. Pictures. (2013). *Her* [Film]. United States.

Wyatt-Moon, G. (2019, August). *The Future of Electronics*. Lecture, Cambridge, England; University of Cambridge.