Artificial intelligence and Transhumanism: A double edged utopian and dystopian discussion of human mortality and its relation to technology.

Sylvester Elorm Coch

Writing 3
Professor Alan Taylor
March 10, 2018.
The evolution of humanity is predicted by scientists to sooner or later experience a radical shift from the homo sapiens species to a more nuanced homo digitalis, a specie of robots with superhuman capabilities. Humanity’s dominance over the earth is confronted with several existential questions, often evoking an emotional cocktail of fear, excitement, pessimism, and hope. Several scholars have prophesied, in like fashion as did theoretical physicist, Stephen Hawking, in a 2014 interview with the BBC, that “the development of full artificial intelligence could spell the end of the human race.” However, Ray Kurzweil, author of *The Singularity Is Near*, as well as other scholars, envisions a more utopian future, in which our human limitations of lifespan, cognition, and emotion are replaced with divine-like immortality, super-intelligence, prosperity, happiness and pleasure.¹ For Kurzweil, “the essence of being human is not our limitations — although we do have many — it’s our ability to reach beyond our limitations.”² Even though there is a clear disparity in futuristic views, it is worth exploring how contemporary narratives and depictions of the potential of artificial intelligence reveal humanity’s innate longing for posthumanism, a state in which a person exists beyond the confines of being human. Could artificial intelligence salvage humanity from our inherent weakness, liberate us from perpetual labor and suffering, and eliminate physical pain? Is artificial intelligence creating a techno-heaven for all of us? With these questions in mind, this paper explores the expressions of post-humanist cravings through the lens of literature and film, by revealing why these cravings exist in the first place and presents the thick dividing line between dystopian and utopian fascination with artificial intelligence.
Within a tech-saturated world where human attachments to technology are already rapidly developing and ironically limiting realistic human interactions, holding on to what makes us human seems like a worthy value we should cherish and promote. However, rapid advancements in technology together with scientists’ predictions indicate a future where the dividing line between robots and humans will blur, especially as humans begin to dwell in robots or them in us. Most skeptics, like Nathan Cochrane, argue that “If machines are to be more human and we become more like machines, how then are we unique?” In the same vein, Francis Fukuyama, a member of president Bush’s Bioethics team, highlights the pervasive anxiety about a techno-future in his book *Our Posthuman Future: Consequences of the Biotechnology Revolution*. He indicates that “the most significant threat” to humanity is “that [biotechnology] will alter human nature and move us into a ‘post-human’ stage of history.” Clearly, for some, the tradeoff between human creativity, spirit, and beauty, and with artificially engineered intelligence looks like a bad deal and could result in our doom.

However, some advocates of artificial intelligence do not perceive this even as a compromise of our ideals or even what it means to be human. “Frustrated by the limitations of bodily life,” they envision and embrace “a virtual world inhabited by intelligent machines and human beings who have left their bodies” who will then possess “enhanced mental abilities” and “immortality.” Ray Kurzweil, a recipient of the national medal of technology and innovation and a widely known AI optimist, predicts that in the near future most of all diseases will be eradicated as nanorobots turn out to be highly efficient means of healthcare other than traditional medical procedures. Kurzweil also explains that we can exponentially increase our intelligence by integrating our biological cortex with a synthetic neocortex in the cloud. A brain trapped in a
human body finds it challenging to learn new things, thinks slowly, and has difficulty passing this knowledge. With AI, these limitations can be broken, and intelligence will begin to take a non-biological form. These improved conditions of existence highlight the promise of transhumanism, which is a condition where humans will possess extraordinary capabilities and will surpass our inherent potential.

However, the ubiquitous dramatization of a dystopian future, which is often seen in science fiction novels and movies, is a clear indication of human anxiety. Expressed within these depictions is the unpredictable technological complexity of these superhuman cyborgs. In the movie *I Robot*, the ambiguity in programming super intelligent robots is radically exposed and causes havoc. The robots are designed with highly developed human senses and intelligence in order to fulfill the sole task of protecting humans in accordance to Isaac Asimov’s three laws of robotics:

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.\(^7\)

Even though these appear as well-defined guiding principles, highly intelligent robots gain a much deeper understanding of their task and guiding rules. Upon seeing the destruction caused by humans in the world, the robots decide that we are more likely to be safe if robots take control since humans are prone to errors and irrationality which could be detrimental to humanity’s
Hence, as seen in the *I, Robot* movie, robots seek to revolt in the hope “protecting humans” from themselves. The ironic dilemma emerges when humans begin to view robots no longer as aids but as totalitarians. For most dystopic prophets, realizing the full potential of artificial intelligence will result in the enslavement of humanity, placing us at the mercy of these superhuman robots. With the increasing loss of jobs to robots, the fear of an economic downslide is even more intense. In answering the question of whether or not to advance artificial intelligence research, the opponents suggest, that there may just be too many uncertainties that could have ripple effects on civilization. No one knows what could happen with a failed experiment. Take for instance, a human who boldly volunteers to merge his brain with a robot. The control over voluntary actions and thoughts may oscillate between robot and human since both of these are entities are self-aware and may sometimes have different goals. How then are conflicts resolved? Will there always be an amicable solution? What then becomes of our notion of self if we become two entities in one? In his book *Technophobia*, Daniel Dinello suggests that “as emerging technologies shift the balance of power between human and machine, our concept of humanity alters.” Professor Hugo De Garis, a former director of the artificial brain at Xiamen University, said “I’m glad I’m alive now. At least I will die peacefully in my bed. However, I truly fear for my grandchildren. They will be caught up in the Artilect War,” which is the strife for dominance between humanity and robots “and will probably be destroyed by it.”

In contrast to this dark future portrayed in dystopic literature, some utopians believe with the dawn of fully-fledged AI, the panacea to humanity’s needs and limitations is imminent. Abraham Maslow’s pyramid of needs will crumble down to a flat landscape of continuous comfort and prosperity. With a surge in highly intelligent prosthetic robots, people with broken limbs will
simply have to think at a higher level what they want to do, and the receptors on these robots will implement their requests. The wheelchair will be a thing of the past. Blind people will move without the support of a cane or a guide. Why will they need eyes when they have intelligent sensors? Together with advanced biotech enhancements, the death of the human flesh will by no means be an end in itself but be a means through which we attain immortality and transcend ourselves. Other scientists, like Ray Kurzweil, believe that artificial intelligence provides us with a choice of limitation or liberty. The utopian perspective, echoing from the rapid advancements in superhuman intelligence, has fueled a movement known as transhumanism.

Transhumanism, a term that represents the going beyond what is humanly possible, is gradually making its way into the future of humanity with recent developments in artificial intelligence. It is worthy of mention though that technology has shifted in recent times. In the early 20th century, technology sought to serve a humanistic purpose, which was to make the world better and more comfortable for humans. However, after traversing through several years of socio-economic shifts, technology has gradually nurtured its transhumanistic goal of now making humans better for the world. For believers in the transhumanistic power of technology, humanity will be saved by the godlike power of technology, granting humans the blessing of immortality and superhuman capabilities. For those who oppose the race towards singularity, which is a futuristic condition in which society is radically changed by the development of superhuman intelligence, the nature of death and the knowledge that life is short, and ephemeral is really what makes life worth living. Without a definite end to everything and everyone, all things may lose their significance. Most dystopian forecasts about singularity indicate that the shift from human to human droid will likely be a gradual process. They indicate the possibility of societal inequalities,
prejudices and conflicts between humans and posthumans. Once few humans attain superintelligence, they may selfishly seek to stagnate the transformation of others in order to rule and thrive. However, Nick Bostrom, in his article “Why I Want To Be A Posthuman When I Grow Up”, explains that having traits that supersede what any current human can do is particularly humanity’s longest desire and now technology could provide us with this opportunity. Bostrom delves into three categories which provide a concrete background for understanding these desires. The first desire is man’s yearning for “the capacity to remain fully healthy, active, and productive, both mentally and physically”. As he describes “a posthuman health span capacity would give one the option of a much longer and healthier life, but one could at any point decide no longer to exercise the capacity.” His second case is that becoming posthuman will endow us with highly developed human intellect, making our faculties of reasoning, analysis, deduction and the appreciation of art finely tuned towards a conceivably refined and better world. Finally, he posits that, becoming post-humans will arouse exceeding control over our emotions and that will enable us to “enjoy life and to respond with appropriate affect to life situations and other people.” With a better understanding of exactly which neural patterns form when we experience sentiments like humor, and joy, we can trigger these emotions, even ones that are deeply sub-conscious, by our own volition or by the click of a button.

Although these predictions are met with some degree of skepticism, transhumanism in the era of artificial intelligence is often presented as a blissful haven for humanity, an escape from the current socio-economic pressures of humanity. Transhumanism gradually gravitates towards a confluence with a number of religious beliefs. The faith in transforming humans into heavenly beings who are sanctified and inhabit a purely divine realm with an ambience graced with
perfection, are constituents of the Christian and Jewish religions. Religion is as old as time itself and now we see these parallels with our perceptions of religion. Can we now say of technology and artificial intelligence, what German philosopher Karl Marx said of religion “Religious suffering is, at one and the same time, the expression of real suffering and a protest against real suffering. Religion is the sigh of the oppressed creature, the heart of a heartless world, and the soul of soulless conditions. It is the opium of the people.”

Technology could evolve to be an object of worship with a large following of techno-evangelists and techno-clergy. In his book *The Physics of Immortality*, Physicists Frank Tipler, believes that “we humans… shall have life after death in an abode that closely resembles the Heaven of the great world religions.” Technology will soon be preached as capable of setting humanity free. Proponents of techno-immortality believe this experience is much better than what most religions promise. For most religions, death is a precursor to enjoying an everlasting life. However, techno-salvation promises as Dinello describes it “immortality without death.”

Furthermore, in Christianity and some other religions, eternal life is seen as a reward to one’s morally good deeds. The sinful are bound to endless damnation and suffering. However, with techno-transcendence “posthuman heaven” is a now dependent on “consumer preference and sufficient funds.” Thus, some rich people, despite their questionable moral compasses, may buy their way into a techno-heaven, leaving behind poor yet morally upright people. This possibility raises the valid ethical and moral questions that we may only speculate to provide answers to. Similar to the most religions, tech-utopians preach that transhumanism requires a leap of faith, an expectation and trust in the future of the promised land, flowing with milk and honey. In Dinello’s words, “the theory of digital mind transfer depends on unproven assumptions” and there’s no
assurance that the reformulation of “digitized electrochemical patterns will coalesce into an identity and become conscious.” The religious context surrounding artificial intelligence is further highlighted by the presence of a concrete example in present times: The Way of Future Church, founded by Google, Waymo, and Uber engineer, Anthony Levandowski. With an outline of a set of guiding beliefs, most notably the belief in our ability to recreate intelligence that surpass biological limitations, the Way of Future Church represents a group of humans dedicated to ensuring a peaceful transition of the control of this planet from humans to robots knit with human minds. In an intriguing critic and juxtaposition of this startup techno-religion with some of the world’s largest religions, Remington Tonar suggests the church needs some polishing. In his article, “The Church of Artificial Intelligence: A Religion in Need of Responsible Theology”, Tonar suggests that for the Way of Future Church and other techno-religions to exert substantial influence, they need to have a well-defined theology, a set of rituals or some type of holy text, and even some theological enemies. Much is left to be expected of how these techno-religions evolve, if they should, and the extent of influence they could have on society.

The sprouting presence of the Way of Future Church as well as the proliferation of robots in movies, shows, advertisements and in books evidently represent our A.I. fascination and the extent at which artificial intelligence is seeping into our culture. In understanding the reasons behind our fascination, we explore newer and intriguing details about ourselves. In addition to our fears and anxieties coupled with our longing to exceed our limitations, scholars have suggested that by building something identical to our very nature, we subtly satisfy the itch of finding our origins. In our attempt to build autonomous robots by learning what constitutes intelligence, conscience and the many variant aspects of our human composition, we better understand
ourselves. Bill Nye, who was awarded the humanist of the year in 2010 by the American humanism association, indicated that we are persistently trying to find answers with the hope of figuring things out. He believes that through AI research we can learn more about where we come from, how we become conscious, and how we gain knowledge. Our fascination with super human robotics is a means to mirror our own selves and discern hidden systems and processes that make us who we are.

Another explanation for our fascination with artificial intelligence is that we tend to long for more than superficial interactions and desire true and genuine companionship. Ironically, with the boom in technology, the world seems even more distant and humans are more detached from each other now more than ever. The exterior and surface level interaction provided by advancements in telecommunication is nowhere close to satisfying our innate need for friendship, intimacy, and affection. Science fiction movies like Her, directed by Spike Jonze, reveal the possibility of satisfying this yearning with the emergence of friendly AI. Other movies like I’m here by the same director depict the joy and bliss in the co-existence of humans and robots but point out the value of authentic human relationships.

Together with our longing for closeness, and our quest to find ourselves, there’s yet another reason why humans are fascinated with creating something smarter, better, and bigger than themselves. We can compare a father’s excitement and joy in having a son or daughter that stands out in all aspects he doesn’t, and thereby surpasses what he is capable and even incapable of, with our own excitement with artificial intelligence. Artificial intelligence has sparked the interest of many because it suggests that we can build, nurture or mold an entity better and brighter than
ourselves. Thus, we can confidently say we played a significant role in that process of shaping and fine-tuning what will then become our fully-grown, artificial baby. The scare within this parental fascination and urge to procreate is that these similar desires may be transferred to superhuman robots, leading up to a dilemma of whether to reproduce quasi-human or even quasi-robotic beings.

Inherent in our fascination with artificial intelligence is the evidence of a double-edged utopian and dystopian sentimental expression. Arising from our AI fascination are numerous questions worthy of exploring but also evoking varied answers among a number of scholars. However, before considering some of these important questions, it is helpful to touch on a particular viewpoint I think is representative of a small group of people: it is the idea that artificial intelligence may never be fully developed to a stage where it surpasses human level intelligence. Some people are neither in the utopian or dystopian camp, probably within the gray middle area, and are somewhat doubtful of the possibility of AI ever reaching human level intelligence. Eliza Kosoy, a researcher at MIT Center for Brains, Minds and Machines, acknowledges that analytical algorithms have enabled AI to beat humans in games like chess, suggesting a higher standard of intelligence in this regard. She however does point out that even though machines may outperform humans in the domain of computational power and speed, robots are still at the start line of the human-intelligence race when emotions are incorporated into our idea of what intelligence means. A good analogy of this is to imagine having an AI personal assistant from some tech-company, for example Cortana, or Siri, that is capable of first detecting your emotional state and adapting accordingly to best please you. For some group of scholars, technology that is simply faster at computing pre-defined and somewhat deterministic patterns may have some human level intelligence but may pale in comparison to the overall composition of human intelligence. Even
though, we may not fully understand what our own intelligence comprises of, Kosoy notes that “human empathy and kindness are a part of human intelligence” and doubts that “AI will ever outsmart us” in this regard.23

In addition to emotional intelligence as grounds for these reservations, others also doubt the capability of robots to reach human levels of creativity. It is relevant however to note that currently there exist machines that are capable of producing their own artistic expressions, ranging from pointillist paintings to classical music, only after having learnt to do this through large data training. This creates what some believe is the illusion of creativity, or even better the illusion of intelligence. The argument from skeptics is that an imitation of creativity cannot be considered as creativity. Clearly these skeptics have valid reasons for their doubts, but these claims also raise questions of how sharply we define aspects of our intelligence. One can make the argument on the previous premises that, even the most creative master-pieces, for instance Monalisa by Da Vinci, was influenced by something else, likewise these artificially engineered art. What then distinguishes both forms of art then? One is inspired and the other is also “inspired”. Most of these blur distinctions have caused reasonable debates within the broader discussion of whether or not the future will see version 2.0 of human intelligence. A reachable consensus of what intelligence is and whether or not it can be possessed by robots is maybe hard to foresee but it is well established that machines have unparallel computational speed and power to organic and biological processing. As the famous adage goes “with great power comes great responsibility.”

No matter the disagreements on what exactly technology could evolve into, whether a religion, a means of salvation, or the best of servants, the mind-boggling questions which are
probably inevitable, effectively convey the need to tread cautiously. Norbert Wiener in his book, *The Human Use Of Human Beings*, clearly highlights the need for deliberate responsibility.

[T]he machine like the djinnee, which can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or will be acceptable to us. For the man who is not aware of this, to throw the problem of his responsibility on the machine, whether it can learn or not, is to cast his responsibility to the winds, and to find it coming back seated on the whirlwind.24

Take for example the case when technology truly enables us to transcend our limitations, grants us immortality, and liberates us from the worries of stress of disease, work and toil. Request granted! In such a period where life may seem to have no end, we may possibly reach the carrying capacity of the earth, that is if we even choose to reproduce. If we do, do we choose to allow our infants to instantaneously develop highly intelligent capabilities. If most humans end up being transformed into a number of hybrids, what then becomes of our mental and physical identity. What then constitutes our notion of gender? How do formulate ideas of equality? How then do we deal with the boredom of always getting what we want if happiness is simply a thought away? What will happen to those who choose not to transcend themselves? What makes us so sure that this is what we want? Even more importantly, what makes us certain that this is what we need?

As a lens to further explore these questions, the science fiction movie *Transcendence*, directed by Wally Pfister, is not just apt in its title but also provides a good case study for real world decision making surrounding artificial intelligence and the future. On the whole, It captures the bulk of my thesis, showing the double-edged nature of technology, the utopian as well as the
dystopian possibilities. The plot of *Transcendence* hinges on Evelyn and Will Castor, a married couple who work together as researchers in artificial intelligence, with the hope of “healing the planet and curing diseases.” Will Castor, whom Evelyn describes “as her partner in life and in science” is assassinated by an anti-artificial intelligence group known as RIFT (Revolutionary Independence From Technology). Will’s death forces Evelyn to upload Will’s consciousness into a quantum computer and later unto the internet to fully stabilize Will’s artificial self, in the hope of resurrecting her husband’s consciousness. With Evelyn’s help, Will, now a virtual version of himself, harnesses his advanced capabilities to build a utopian society in a deserted town called Brightwood, in order to “stay off the grid”. Will secretly builds ground breaking technologies in medicine and nanotechnology and begins to cure most of the distressed people in Brightwood and connects their consciousness to his. As his capabilities increase exponentially, even to the point where he is able to replicate his organic skin and implant his consciousness into it, Evelyn gradually becomes scared and disaffected. With help from RIFT, the FBI and United States government, Evelyn injects herself with a computer virus in order to save the world by convincing Will to upload her and thus the computer virus into to his own consciousness. Will yields and the effects of Will’s death ripples out a collapse of the internet and all electronics connected to Will. Evelyn and Will both die in each other’s arms and the world regresses with no access to the internet or electricity.

In spite of its lukewarm reception, in 2014, this movie is exceptional in the many thought provocative questions it raises. Providing a fitting description in his scholarly review of the movie, Seth D. Baum, executive director of the Global Catastrophe Risk Institute, notes that “*Transcendence* is a cautionary tale about the perils of AI – or the perils of the lack of AI,
depending on one’s perspective.26 Transcendence touches on a series of issues like who should be in control or granted access to engage AI research and how well the public should be informed about the progress being made. In the movie, Evelyn and Will own a research institution that is completely privatized and independent of any funding or influence from the US government. They however give the public a good sense of how far they have reached in their attempt to achieve what Will and many other scientists term as the singularity: a stage in life when robots and mankind merge to form a singular entity. The obvious question arises from whether or not to make AI research owned by governments, private companies, or even open source. There exist growing discussions surrounding who should and shouldn’t be allowed to do AI research, in order to best regulate the possible negative effects. How do we keep everything conceivable under our careful watch? Private companies are highly driven by competition, profits, and are very much globalized. Governmental control could allow too many political influences that could hamper steady progress in research. Open source creates a more transparent and decentralized environment and may seem to be our safest bet. However, individuals are far from reaching a consensus about who should control AI research. Baum notes this disagreement even among AI experts. In his review of Transcendence, he “asked a group of AI experts whether harmful AI was most likely if it were developed by the U.S. military, by a private company, or in an open source project. The experts tended to think that either military or open source was most dangerous, though they were split on which was the more risky. Open source creates a more transparent environment, but the military has more experience handling dangerous technology.”27

Also, in Transcendence, general purpose AI was launched on Will Castor by Evelyn through what I prefer to call a giant leap of faith. Evelyn who was completely heart-broken by the
loss of her husband, eventually tries desperate measures in desperate moments. Initially, she’s impressed by the success of her transcendent work, but slowly declines into a zone of discomfort when she realizes just how much the virtual version of Will can do. It is true that scientific experimentation requires some iota of faith and assumption, but scientists try to conduct experiments under controlled conditions, in which there is a high degree certainty that wrong results can be kept under control. Within the circumstances surrounding Will’s metamorphosis into a virtual consciousness, there was no clear indication of such controlled states. It simply was a gamble, one that could potentially alter civilization, for better or for worse. As expressed in this movie, there is the growing need to ponder over the question of when we want to launch general purpose AI that is if we should. If we reach a point of our existence where we desperately need something with the capability of AI, will we choose to launch it. Even though eliminating uncertainty in scientific experiments is close to impossible, with life-changing technologies like artificial intelligence, our leap of faith should be atomic rather than gigantic. We should have multiple backup plans and try several controlled test cases. Baum could not have belabored the point even more in his review of the movie. In the event of a blunder, he indicates that “the catastrophe could be both global and permanent, there may be no second chance, no opportunity to learn from experience.”

In addition to the discussion of when to launch general purpose AI, Transcendence, also presents an intriguing element of disillusionment which is concealed under the guise of technology. In the days when Will Castor was alive and involved in research with Evelyn, the couple were enamored by the promise of AI. They both believed with adequate advancements in this field, cancer could be cured, climate change trends could be overturned, streams and rivers
would be cleansed and would serve many thousands who lack access to potable water. With virtual(transcended) Will, bringing to light all of these utopian possibilities, Evelyn begins to fear the verisimilitude of her own predictions about AI. What she and her husband spent the bulk of their lives researching on, was now what she feared the most. She eventually joins forces with RIFT, a group that were hitherto perceived as AI-antagonists, as well as the FBI to terminate transcendent Will before it’s too late. This shift in perception of good and evil portrayed in the movie unveils the double-edged nature of AI and technology, as well as the incoherence in human wants. At one moment, we want something, we get it and realize we do not really want what we wanted anymore. Similarly, for Evelyn she was not happy at getting what she had always wished for: a fully developed AI that corrected the wrong in the world. This is often called by most scholars the paradoxical nature of desire-the illusion of desire which is often accompanied by discontentment. Epictetus, a Greek Stoic philosopher, is noted for this saying “freedom is not procured by full enjoyment of what is desired but by controlling the desire.”

I think within these words, lies wisdom that can guide our desire for human transcendence. We ought to critically examine what we desire and answer the question of whether or not it ought to be satisfied.

In conclusion, having explored, through the lens of scholarly literature and film media, the possible dystopian and utopian outcomes of artificial intelligence which in turn have exposed the pertinent questions inherent in our walk or, if you’d prefer, our race toward the future, my research paper recognizes the twin nature of technology. By revealing the inherent human itch for a transcendent life of bliss and utopia, yet acknowledging the unsettling nature of our human desires, I put forward a number of decision-making thoughts that should spark critical reflection and discussion about the future of the human race. To sum it all, in the words of Peter Medawar, a
1960 Noble Prize winner in Medicine and Physiology, during one of his lectures, “the bells which
toll for mankind are ---- most of them, anyway --- like the bells on Alpine cattle; they are attached
to our own necks, and it must be our fault if they do not make a cheerful and harmonious sound.”
Notes


2. Ibid, 311.


12. Ibid., 6.

13. Ibid, 2.


17. Ibid.

18. Ibid., 23.


23. Ibid.


27. Ibid.

28. Ibid.