

Dat Do

Professor Allen

ENGL 201 - The Research Paper

6-15-18

Debates on the Development and Research of AI:

How much limitation, if any

Humankind has always had an urge to create. We want to build, to form, and to make. This great desire, although sometimes comes at a cost, has proven to be our most vital if not our best trait yet. Modern society and new ways of living have brought researchers and scientists to think of faster, more efficient ways of executing a task, any task, whether it be simple or intricate. This has ultimately come in the form of artificial intelligence, or AI for short. An AI, today, is designed to perform narrow and specific tasks such as driving a car, recognizing facial features, or conducting useful internet researches. However, it is the long-term goal of many scientists that artificial intelligence would be able to outperform humans in essentially any cognitive task. So, think of Google, Siri, and Cortana, but still years and years away from what scientists are envisioning. Like the countless revolutionary inventions that came before it, the development of AI receives as much if not more skepticism than its predecessors. Many believe that AI could very well be our last invention, ever. While many also see a future where humans live worry-free of all the wearisome duties that each person currently possess. Fundamentally, this paper explores the concerns that arise from the uncertainties that come with the potential of AI. Thus, making it important to also take into consideration how much limitation, or regulation, if any, should be applied to the research and development of the technology.

To begin, it is crucial to first observe the incredible capabilities that current artificial intelligence possess, and how this technology is shaping and changing human societies. The idea that an AI would become smart and strong enough to eventually take over humanity was largely science fiction, and as a science fiction, many also thought that such a reality would be decades if not centuries from now. However, AI has proven to be more than just science fiction as the research and development in the technology have been brewing in recent years and is predicted to very well be the next revolutionary change in history. This change is coming in many forms, but not in any way a physical sense. Meaning, the goal of many AI developers is to create essentially another intelligent being that can effectively communicate in any normal human interaction. Take Google Duplex, for instance, Google's own version of an AI, "a new technology for conducting natural conversations to carry out 'real world' tasks over the phone" (Google Duplex). In a demonstration made by Google CEO Sundar Pichai at Google I/O 2018, the AI was asked to book an appointment at a hair salon for a Tuesday morning. Not soon after the request was made, Duplex quickly came to a conversation with someone at the salon over the phone. Then, things became very interesting. When told by the hair salon, "Sure, give me one second," the AI responded, "Mm-hmm". Rather than a simple "yes," the AI knew to reply in a way that kept the conversation very fluent. It had kept a very human-like voice of tone and was in many ways, a human in itself. That is, in exception to its apparent missing physical body. Overall, however, the conversation was a very fascinating view into what an AI can do and can be. Moreover, it demonstrates the significant lack of familiarity humans have with the technology.



Figure 1. Google CEO Sundar Pichai, Google I/O 2018.

In 1951, computer scientist Alan Turing proposed a test called "The Imitation Game" which was used to ultimately decide whether a machine is intelligent. In simple terms, there would be two participants, one human and the other machine, and a "judge," human, attempting to decide between which of the two is a fellow human, and which is the machine. In the case of Google Duplex, the hair salon receptionist was not able to identify the Google Assistant Duplex as being an AI. Whereas, the AI was able to prove its intelligence as being "a passable simulation of a human being" (The Turing Test). So, whether the hair salon receptionist failed the test, or the Google AI passed it with ease, it is clear to see how much an artificial intelligence is capable of, and more importantly the technology's imminent potential.

Continuing with the Turing Test, and how an AI could potentially become one within a human society, the film *Ex Machina*, released in 2015, brings to the big screen an excellent illustration of a fully-developed artificial intelligence. In summary, a programmer, Caleb, from a large Tech company wins a raffle to spend a week at his CEO's private home. Once there, Caleb is informed by the CEO that he will serve as the judge part of the Turing Test on a newly

developed artificial intelligence called Ava. As time progresses within the week, Caleb begins to grow feelings for Ava. From his interactions with the AI, Caleb would question the motives of his CEO in thinking that Ava is not treated like a human being should be treated. There is little to no ambiguity as to what Ava is capable of; she is as smart as any human being taking into consideration her interactions with Caleb. However, the film goes beyond this in attempting to define what it is to be human; Ava demonstrates an amount of emotional intelligence and consciousness that is on par with any human being. So much so that Caleb is driven to develop romantic interests in the AI. This is strange in a way that romantic relationships are primarily and has always been between humans, but this is not the case that is portrayed by these two characters of the film. *Ex Machina* provides a great example of what the future could hold with artificial intelligence that is capable of possessing and manipulating human emotions. Emotions that are distinct and unique characteristics found primarily in human beings. Overall, it is no doubt that the emergence of AI encompasses a great promise to transform human societies in historic ways.

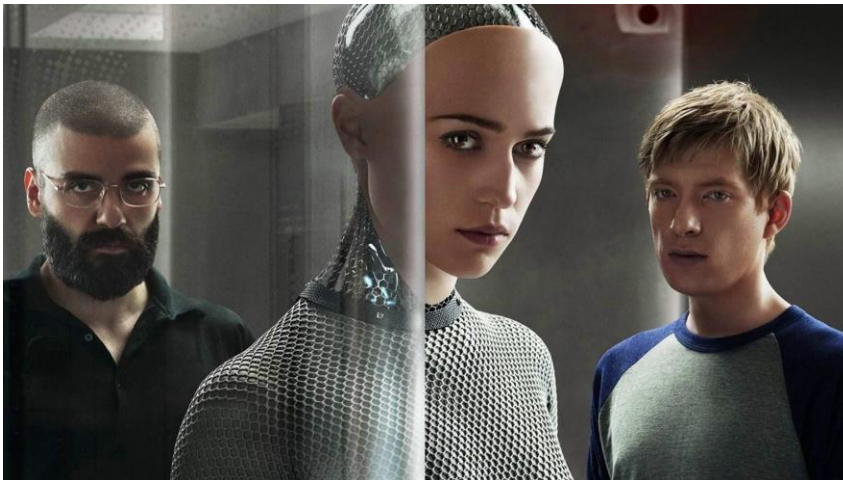


Figure 2. Left to right: (CEO) Nathan, Ava, and Caleb. Ex Machina, Released in 2015.

Subsequently, however, AI's successful development could practically lead to many threats to humanity. According to a group of authors from the Wilson Center, "without adequate risk assessment and mitigation, AI may pose a threat to existing vulnerabilities in our defenses, economic systems, and social structures." Furthermore, much of the controversies surrounding the topic of AI are mostly based off the work of science fiction and film which depict evil yet very intelligent robots taking over the world. As explained by Paul Ford in the article, *Our Fear of Artificial Intelligence*, "the idea that computers would eventually develop the ability to speak and think—and thus to do evil—bubbled into mainstream culture." Upon further examination of the issues, however, the natural fear of AI that has been pushed onto many may not be totally unreasonable. Take, for instance, the example introduced by Stuart Armstrong at the Future of Humanity Institute of "an anti-virus program that's dedicated to filtering out viruses from incoming emails." The AI could determine the solution to its given task would be to kill all humans making it impossible for any emails, virus-carrying or not, to be sent indefinitely. Because of this, establishing a balance between machine and human goals within the research and development process is crucial. Another thing to recognize which is a more realistic and near-term threat is unemployment, "as replacements for virtually all human workers are quickly developed and replicated" (Rundle). Technology has been a threat to the role of humans in the workplace for recent years. If humans continue to implement new, supreme, and intelligent technology, there will be less desire for human workers and more desire for the technology. A technology that can finish a task in a fraction of the time it takes the best human worker to complete. In the case of AI, with the goal of many developers to one day have all types of tasks be done through automated machines, it can indeed be very worrisome when the livelihood of many will be in jeopardy. From this, it is essential to identify the possible solutions that could

hopefully come into effect. Solutions that will ultimately put humanity's welfare and future as first priorities while overseeing the process of research and development in the technology before one can be fully implemented into society.

First, there are solutions, or in this context, regulations, that although would guarantee humankind's safety in the future, they would also drastically slow down if not prevent the progress of AI completely. Consider, for instance, Bill Gates' solution, where he proposes that since human workers are required to pay income tax, so should the robots that will one day replace those human workers. In Bill Gates' words, "Right now, if a human worker does . . . \$50,000 worth of work in a factory, that income is taxed. If a robot comes in to do the same thing, you'd think that we'd tax the robot at a similar level." This solution, although a logical regulation, does not account for a multitude of other problems that may very well follow. Taxes are and have always been a big factor for many if not all large corporations. Now, consider a technology company aspiring to create the very first AI that could learn to fulfill all the cognitive tasks that humans perform. Taxes will essentially cause that company to relocate to another country where, of course, its mission and potential for creating such a product are in high demand anywhere in the world. Thus, resulting in a major loss of technical jobs here on our own soil. As former US treasury secretary, Lawrence Summers stated, "Taxes on technology are more likely to drive production offshore than create jobs at home." Primarily, taxes are just another way to harshly prohibit the development and research of AI without considering the enormous potential the technology can offer.

Other, more lenient solutions which not only take into consideration the uncertainties behind the technology, but also encourage the visionary development and research in AI are, in many ways, the better approach. These proposals are fundamentally mutual agreements which

will satisfy the scientists and researchers while alleviating the worries of the public. According to AI expert, Neil Jacobstein, who has consulted on projects for the U.S. military, a straightforward solution that could prevent "unwanted effects" from an AI could be a control system to shut it all down. "If something does go wrong, and in most systems it occasionally does, how quickly can we get the system back on track? That's often by having multiple redundant pathways for establishing control," Jacobstein said. In other words, because of the large control in which an AI will be under, there could be an established pathway that could simply shut it down completely if the time comes. Furthermore, going back to the problem of unemployment when computers and robots would eventually take over most jobs. Rather than taxing the robots themselves as suggested by Bill Gates, it is more rational to spend resources on training and educating potential workers in the right skills. Workers will have more comfortability and knowledge in and around the development and research that goes into AI. As a result, the uncertainties expressed by the public will diminish due to the fact that so many people will have been trained and educated on the technology. As informed my David Kenny in his article, *Bill Gates is Wrong: The Solution to AI Taking Jobs is Training, Not Taxes*, many current workers at IBM were previously mid-career professionals who were motivated and given opportunities to learn the skills required to thrive in the new digital economy. "They are former teachers, fast food workers, and rappers who now fight cyber threats, operate cloud platforms and design digital experiences for mobile applications," said Kenny. It is important to have expertise in the research and development process of AI, and certainly very crucial when it comes from regular day-to-day people of the public. This is because how and what AI will come to be are ultimately in the hands of the people who will be utilizing and benefiting from the technology. Moreover, further expertise that is also very vital is that of leading researchers from credible institutions who would be able to

offer high-level roadmaps to assessing the AI. According to the World Economic Forum, researchers at institutions such as Stanford provide "in-depth considerations of things like ethically-aligned AI design to maximize human well-being." Essentially, solutions that enable and encourage the fast progress of AI while not overlooking the internal threats and risks that come with the technology will be better suited for the near future to come.

The future of humankind is very bright and full of many technological advancements and discoveries. AI serves as an excellent introduction to this future as well as sets the basis for the protocol and procedure that will need to be developed along with the technology as a way of improving and maintaining the wellbeing of humans. AI has a tremendous potential to achieve the ambitious goals envisioned by many, but it could also come with negative consequences if not overseen and appropriately assessed. Therefore, before proceeding with implementing the technology into society, it is vital to create regulations and limitations to what the technology will be capable of. Whether these solutions to future problems are extremely confining or openly tolerant, it will undoubtedly affect what the technology could accomplish. However, with wise and insightful supervision, AI will have the chance to prove its revolutionary capabilities and with it will come a myriad of opportunities that will further improve society.

Works Cited

- “Artificial Intelligence: A Policy-Oriented Introduction.” *Wilson Center*, 30 Nov. 2017,
www.wilsoncenter.org/publication/artificial-intelligence-policy-oriented-introduction.
- Artificial Intelligence / The Turing Test*,
www.psych.utoronto.ca/users/reingold/courses/ai/turing.html.
- “Benefits & Risks of Artificial Intelligence.” *Future of Life Institute*, The
FLI Team https://futureoflife.org/Wp-Content/Uploads/2015/10/FLI_logo-1.Png,
futureoflife.org/background/benefits-risks-of-artificial-intelligence/.
- Delaney, Kevin J. “The Robot That Takes Your Job Should Pay Taxes, Says Bill Gates.” *Quartz*,
Quartz, 17 Feb. 2017, qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/.
- Ford, Paul. “Are We Smart Enough to Control Artificial Intelligence?” *MIT Technology Review*,
MIT Technology Review, 1 Feb. 2016, www.technologyreview.com/s/534871/our-fear-of-artificial-intelligence/.
- “Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone.” *Google AI Blog*, 8 May 2018, ai.googleblog.com/2018/05/duplex-ai-system-for-natural-Conversation.html.
- “How to Manage AI's Risks and Benefits.” *World Economic Forum*,

www.weforum.org/agenda/2018/01/how-to-manage-ais-risks-and-benefits.

Kenny, David. "Bill Gates Is Wrong: The Solution to AI Taking Jobs Is Training, Not Taxes." *Wired*, Conde Nast, 3 June 2017, www.wired.com/2017/04/bill-gates-wrong-solution-ai-taking-jobs-training-not-taxes/.

Leadbeater, Alex. "What Does The Ending Of Ex Machina Really Mean?" *WhatCulture.com*, WhatCulture.com, 6 Aug. 2016, whatculture.com/film/what-does-the-ending-of-ex-machina-really-mean?page=3.

Muoio, Danielle. "4 Very Real Risks AI Poses to Mankind - and How We Could Prevent Them." *Business Insider*, Business Insider, 7 June 2016, www.businessinsider.com/4-ai-risks-and-potential-solutions-2016-6.

Rundle, Michael. "Super-Intelligent Computers Could Enslave Humanity Says Oxford University. (Or They'll Just Kill Us)." *HuffPost UK*, HuffPost UK, 12 Mar. 2014, www.huffingtonpost.co.uk/2014/03/12/extinction-artificial-intelligence-oxford-stuart-armstrong_n_4947082.html.