Discussion of the Potential and Limitations of Artificial Intelligence – Power and Consciousness

Ben Meshanko
Purdue University Department of Computer Science
CS390GIS: Great Issues in Computing
Prof. Eugene Spafford
December 12th, 2022

Introduction

One of the subjects dominating our discourse today is that Artificial Intelligence (AI) has faced explosive growth in the past decade. While it is easy to marvel at how far we have come in researching and developing this field, the level of disruption that society has faced from AI has been limited so far. This is not to discount our advances, or say that other forms of algorithmic development and data processing have not been disruptive, but AI is still a newborn in the grand scheme of computer science. One of such advances is DALL-E; the product of OpenAI's research has taken the Internet by storm and is a shining example of the potential of AI to mimic human capabilities in art and other traditionally "right-brain" activities. AI is able to near-instantly create a picture of "An astronaut riding a horse in the style of Andy Warhol". The potential is limitless and will continue to grow as neutral networks become more complex and more training data is compiled. Unfortunately, in our rapid technological progress we have not taken a step back and considered how AI should fit into our economic and social structures in the future when AI may be incredibly disruptive to modern enterprises. The long-standing idea of robotics for industrial automation is being challenged by the dawn of general-purpose robots capable of completing abstract tasks.

Another recent development by OpenAI is ChatGPT, which has taken the Internet by storm, with many writing prompts that exhibit its useful capabilities. Here's a prompt: why is ChatGPT important to the development of AI?

> "ChatGPT is important to the development of AI because it represents a significant advance in the field of conversational AI. It incorporates key innovations that enable it to handle the unique challenges of generating human-like responses in real-time conversation, which makes it possible to create more engaging and natural-feeling conversational experiences. This has the potential to make chatbots and other conversational applications more useful and effective for a wide range of applications. Additionally, the development of ChatGPT also represents a broader advance in the field of natural language processing, which is an important subfield of AI that focuses on enabling computers to understand and generate human language. This has the potential to enable a wide range of applications in areas such as language translation and information retrieval. Overall, the importance of ChatGPT to the development of AI lies in its ability to advance the state-of-the-art in conversational AI and natural language processing, and

to enable the creation of more useful and effective applications in these fields." (OpenAI, 2022).

While certainly in its infancy in terms of accuracy and robustness, large-language models are capable of quickly generating knowledge that is useful for humans.

Nonetheless, it remains to be seen how AI will affect humanity's development. With exponential growth in the capabilities of recently-designed AI-based software, it appears that humans will be obsolete in the future. CGP Grey's legendary YouTube video *Humans Need Not Apply* predicts this – detailing how AI threatens nearly every modern industry (2014). However, the evidence is not wholly clear, and researchers are still unsure about potential limitations that may theoretically still give humans an edge decades into the future. Consciousness is often discussed as an end-goal for AI, but little progress has been made to even define consciousness in humans, let alone machines. In addition, human athleticism may be incredibly hard to replicate – we are masters of the physical world. Perhaps even emotional awareness raises human intuition to a level that cannot be matched by machines. Overall, artificial intelligence thus far has had limited social and economic implications, but it is certain that its development will be profoundly transformative to humanity despite any philosophical limitations. In this paper, the potential and limitations of AI will be analyzed, along with historical parallels that we can draw that may be helpful at analyzing a transformative period in our species' progress.

Historical Context

Before analyzing the potential consequences of hyper-capable AI (sometimes called AGI), it is useful to examine how far research has progressed and the background of humanity's obsession with automation. Earliest in the history of automation was toolmaking. Classical wisdom may suggest that the creation of stone tools was just merely a product of the human brain – we use stone tools because they perform better than our hands and we can adapt to our surroundings and learn complex ideas and acquire knowledge over time. However, research by Dietrich Stout suggests that culture and other social interactions share a role in the rapid development of the Paleolithic era. He writes: "this implies that Lower Palaeolithic hominins possessed adequate cognitive substrates for some degree of cumulative cultural evolution, an unsurprising result considering the transmission capacities of modern chimpanzees" (Stout, 2011). Chimpanzees are social creatures and are documented to have a primitive level of culture

– and many primates have been shown to use stone tools. This research implies that social necessity also drives innovations like stone toolmaking instead of merely the ingenuity of certain members.

The next major innovation that dramatically increased human capacity for work was the domestication of animals. Work animals were able to do an order of magnitude more work than mere humans, and paved the way for agriculture, the most impactful development in human history. Jared Diamond, in his famous *Guns, Germs, and Steel* spends significant time discussing the importance of suitable domesticated animals as a major reason for inequitable human development across the globe. As the people who ended up with the weapons to global dominance, "Europeans today are heirs to *one* of the longest traditions of animal domestication on Earth—that which began in Southwest Asia around 10,000 years ago" (Diamond, 1997). While the Agricultural Revolution was an absolute necessity for modern society via division of labor, the Industrial Revolution is far more relevant to the subject of automation and mechanization. Just in the last 100 years, though, the origin of the idea of robots as mechanized, general-purpose intelligent beings is accredited to science-fiction writer Isaac Asimov. He had deep ethical concerns with the existence of robots with superior intelligence to humans, and his three Laws of Robotics seek to ensure that humanity reigns supreme.

While human progress from the Paleolithic era to modern industrialized society has been marvelous, designing an artificial intelligence is far more complex and daunting of a task than anything that has ever been completed before. Part of the difficulty arises from a lack of fully comprehending intelligence and the complexities of the human brain. Many believe that it is philosophically impossible to create something in the physical world that rivals human consciousness, including famous Enlightenment philosopher John Locke. Perhaps the ultimate form of automation is unreachable, and humanity's quest to abstract away increasingly large amounts of work will be fruitless. AI may become increasingly competent at mimicry, but never truly be conscious and able to act outside of a closed-set of programmed knowledge. The Center of Privacy and Technology, a Georgetown Law think-tank dedicated to analyzing the impact of governmental surveillance policies on marginalized groups, published an article in Medium about the challenges of discussing advances in AI. Instead of using cliché, placeholding terms such as "AI" and "machine learning", the thinktank will attempt to "be as specific as possible about what the technology in question is and how it works" (Tucker, 2022). Journalists focusing

on advances in these fields have been contributing to over-hyped progress and misunderstandings among the general public. However, it may be incredibly difficult to convey this information without going deep into the philosophical and analytical of what these terms mean.

Since the word "AI" has appeared 13 times in this paper, it is probably necessary to define some things. I will be defining intelligence as the ability to act outside of instincts or programming *and* the ability to consider and learn from complex and abstract stimuli. Artificial intelligence thus is simply intelligence that is created by humans. Since this definition of AI does not currently exist (and may never exist) as humans are the only entities that are known to exhibit intelligence by this definition, I will instead refer to this strict definition of artificial intelligence as artificial general intelligence (AGI). "AI" can refer to something that exists today – an artificially created entity that is able to solve abstract or open-ended problems is "artificial intelligence". DeepMind's AlphaZero, the most skilled chess playing mind in the history of the game[1], is not capable of acting outside of instincts, but certainly classifies as AI.

Contemporary and Practical AI – DeepMind

Perhaps the most advanced and potent contemporary artificial intelligence is Alphabet subsidiary DeepMind, whose stated goal is "solving intelligence to advance science and benefit humanity" (DeepMind). With the help of Google's impressive resources and talent acquisition network, DeepMind has been at the frontier of not only research but practical implementations of AI. Besides mastering board games, DeepMind has recently developed an optimized way to multiply matrices 10-20% faster than frequently used algorithms on state-of-the-art hardware, such as Nvidia's V100 GPU (Fawzi et al., 2022). The methodology was also incredibly clever: by creating a game out of efficiently multiplying matrices, DeepMind was able to use the same software that dominated the best chess engines to devise a novel algorithm. Given how entrenched matrix multiplication is to the technology industry, even a 10% improvement in matrix multiplication efficiency would be a dramatic improvement in performance for many software implementations. The research is incredibly promising, and allowing neural networks to design faster implementations of common algorithms may result in unimaginable speedups in the future, particularly those which are less trivial than matrix multiplication.

[1] https://www.deepmind.com/blog/alphazero-shedding-new-light-on-chess-shogi-and-go

Another recently developed project of the DeepMind division involves using neural systems to configure computer networks. The process involved relaxing certain hardcoded requirements that are typically imposed on modern networks to maximize flexibility and performance. As a result, the neural network was able to configure a computer network that is nearly 500x standard network speeds (Beurer-Kellner, Vechev, Vanbever, & Veličković 2022). Automatic configuration of the computer network is important, because the data-trained algorithms are able to make decisions that are optimal for the network automatically that result in optimal performance. Low-level software such as computer networks have under-hyped in recent years because of how pervasive and flashy machine-learning has been in mainstream computer science research, but it remains an area that has historically driven many performance enhancements. This research generates promise that there may be AI-based approaches to compilation and memory management in the near future, which have the potential to dramatically increase the efficiency of operating systems, trickling up into a performance increase for all software. Overall, Google's investment into DeepMind has been deeply promising for the future of artificial intelligence by challenging ideas and algorithms that may have been seen as solved.

The Fourth Industrial Revolution and Humanity's Fate

In the last decade, technology companies have begun to implement AI into their products, showing signs that the transformative changes that have been predicted for decades are finally starting to begin. The most obvious example is self-driving cars. When self-driving cars are mentioned, it is easy to get clouded by the idea of fully autonomous vehicles driving 90mph bumper to bumper, but that future is decades away. The self-driving technology that does exist today, however, includes lane-departure detection, auto-steering, and other software that increases the safety of the vehicle by removing some necessity and perhaps even authority of the driver. The National Highway Traffic Safety Administration developed a roadmap that documents and predicts the evolution of automated safety technologies. However, all roads lead to full automation. There is a powerful social incentive to fully automate vehicles for safety, but really the issue is economics: "Americans spent an estimated 6.9 billion hours in traffic delays in 2014, cutting into time at work or with family, increasing fuel costs and vehicle emissions. Automated driving systems have the potential to improve efficiency and convenience" (U.S.

Department of Transportation). Certainly, self-driving cars are coming, but the real revolutionary changes may lie in the automation of semitrucks.

There are 3.5 million employed truck drivers in the United States (Cheeseman-Day & Hait, 2019). If these jobs were automated away in the future, that would be devasting to those employed people and their families. This is certainly different than the previous industrial revolution, where digitalized and mechanized machines took away millions of manufacturing jobs, but simultaneously created nearly all white-collar jobs that dominate the economy today. Sure, there are software engineering jobs being created by Tesla and other companies but not nearly to the degree of 3.5 million. Efficiency or initial investment is a nonissue for shipping companies: any initial cost differential would rapidly break even in a future where autonomous semitrucks are the norm. It is unnecessary for these semitrucks to be electric – diesel powered autonomous semitrucks can still run 24/7 until they need to be refueled (by a human worker). This is just one of the many significant repercussions to occur the mass-automation fallout, and it is unknown how society will adapt to significant chunks of the workforce being unemployable by no fault of their own.

The compounding effects of AI being implemented in many facets of society will drastically transform the global economy; a shift being coined as the "Fourth Industrial Revolution" (Schwab, 2018). While industrial output will drastically increase, the Fourth Industrial Revolution allows humanity to redefine progress and perhaps even wealth – with a human-centered economy that seeks to adequately distribute resources and good fortunes across the globe. Standards of living will exponentially rise – we will improve all aspects of our world. Ray Kurzweil, a prominent computer scientist and author, even suggests that hyper-intelligent AI created by the Fourth Industrial Revolution may allow humans to conquer their own mortality. AI that can interface through nanotechnology will be able to constantly fix and restore cells – we can optimize our body at the atomic level with God-like intelligence (Kurzweil, 2006). This point in history where AI essentially makes humanity obsolete is called the singularity. Most researchers believe that the dawn of AGI, intelligence that is comparable to humans, will read to runaway growth and the singularity will soon be reached (Urban, 2015). Regardless of the time frame on this intelligence explosion or any discussions on the nature and definitions of consciousness and intelligence, it appears to be inevitable – AI will outclass humans in the near future and bring about an unprecedented rise in our quality of life. Or perhaps not.

Without a doubt, hyper-intelligent AI would be capable of bringing about the extinction of the human race. It would be utterly unstoppable. However, there is very little evidence to show that AI would be concerned with destruction of humanity – most researchers fear that AI will destroy humanity out of necessity for their survival rather than contempt (Urban, 2015). One major difference between the way that humans think and the way that computers think is that humans are unsure of their purpose – perhaps our instinct is to spread genetic material, but humans have moved past merely evolutionary desires[2]. Computers, on the other hand, are intentionally programmed with a purpose. Therefore, in order to safeguard humanity from the dangers of hyper-intelligent AI, some thinkers have proposed enlisting hyper-intelligent AI with the sole purpose of protecting humanity against hyper-intelligent AI. This idea appears to be the only means of survival: "because growth in human intelligence is unlikely to keep pace with growth in artificial intelligence, humans may have little choice but to draw on AI to check AI—and to seek to increase oversight of artificial intelligence as the intelligence of the programs they oversee grows" (Etzioni & Etzioni, 2016). However, it seems philosophically dangerous to enlist machines who do not *care* about humans in a human sense to protect our species. I propose a more ambitious goal: *Artificially Intelligent systems should assist in the expansion of human consciousness while following Asimov's Laws of Robotics.* It appears that there is no sure way to guarantee we are able to control AI, but aligning it to goals favorable to humanity grants hope.

The Consciousness Questions

There are two questions that are generally presented when discussing AI and consciousness. Of course, researchers want to know if humans can create something that is conscious, but the more pressing question is does consciousness even matter – why preserve it? First, no machine that could remotely be considered conscious has been created. The discussion on whether artificial consciousness can even exist is deeply controversial. Many researchers believe that AI will merely need to develop its own communication structure; forming abstract symbols and structured internal representations will form the basis for a human-like brain that is capable of consciousness (Esmaeilzadeh & Vaezi, 2021). These conscious AI are marked by cooperation and creativity: "two machines cooperatively completing a task they are not

[2] Maslow's Hierarchy of Needs displays the stages of the evolution of human needs and desires: https://en.wikipedia.org/wiki/Maslow%27s_hierarchy_of_needs

programmed to do" would have a level of social intelligence that has only been exhibited by humans (Esmaeilzadeh & Vaezi, 2021). However, there is not much reason to believe super-intelligent and conscious AI would make decisions that humans would understand. The intelligence differential is too great. Nonetheless, the question of possibility may go unanswered until a conscious AGI reveals the answer.

If conscious AI is truly possible and was created, it may go undetected for years and have disastrous effects whenever it is uncovered. Machines that are conscious of the consequences of their decisions may be more reluctant to make decisions revealing their intelligence or their consciousness. Perhaps consciousness is an inhibitor, and machines that are conscious are not as capable as unconscious traditional robots. Does it really matter if AI is conscious if it has dramatic implications for society? Probably not. Our desire for AI that thinks like us may be misguided, creating an omnipotent entity that thinks and acts like its human counterparts would be disastrous – those with such a degree of power have not tended to act beneficially to their fellow humans. Maybe a purely analytical oracle created by AI can serve as humanity's collective right-brain, while humans blessed by the fruits of automation are able to abstract away all but the left-brain, emotional aspects of themselves. An AI that has any sort of empathy would be aware of the genocides, environmental destruction, wars and other horrific parts of humanity's tyranny over the planet. Why would an emotionally aware AI favor its creators over the victims of the creators? The uncertainty of consciousness (as depicted by human history) and the instability that AI will create are enough, combining these two issues into an uncontrollable Frankenstein appears to be a recipe for disaster. Consciousness is for humans.

Conclusion

It appears practically unavoidable that the development of AI will place society in the most rapid and dangerous transitory phase in human history. Within the next half-century, the fate of humanity will be sealed. Will we expand to the stars accompanied by hyper-intelligent companions or be annihilated in a rounding error by them? One thing is for certain though: humanity is bound to extinction if we stop innovation. If the 21st Century is the last century that biological life is the sole arbiter of Earth – so be it. Of course, research needs to be carefully controlled and planned in order to limit harmful consequences – but halting innovation out of fear guarantees our downfall. Humanity is presented with the opportunity to become the first

species on Earth to ever reach immortality (Urban, 2015), and we cannot reach it without the rapid explosion in intelligence on Earth that will occur this century. AI will change the world, but the transformation will be incredibly chaotic and dangerous. Nuclear weapons had the same effect of risking our extinction, but advanced our understanding of physics tremendously, and now nuclear power produces over 10% of the world's energy. Now, humanity is developing a much more powerful and dangerous nuclear bomb – artificial intelligence. AI will have far-reaching consequences for all of humanity and may lead to an inconceivably high quality of life, but we risk creating our extinction. Soon.

Work Cited

Asimov, Isaac. (1942). *Runaround*. Street & Smith.

Beurer-Kellner, L., Vechev, M., Vanbever, L., & Veličković, P. (2022). *Learning to Configure Computer Networks with Neural Algorithmic Reasoning*. OpenReview. https://openreview.net/forum?id=AiY6XvomZV4.

Center for Sustainable Systems, University of Michigan. (2021). *Nuclear Energy Factsheet*. https://css.umich.edu/publications/factsheets/energy/nuclear-energy-factsheet.

CGP Grey. (2014, August 13). *Humans Need Not Apply*. YouTube. https://www.youtube.com/watch?v=7Pq-S557XQU.

Cheeseman-Day, J. & Hait, A. W. (2019, June 6). *Number of Truckers at All-Time High*. United States Census Bureau. https://www.census.gov/library/stories/2019/06/america-keeps-on-trucking.html.

DeepMind. (n.d.). *DeepMind: Solving intelligence to advance science and benefit humanity*. https://www.deepmind.com/.

Diamond, Jared. (1997). *Guns Germs and Steel: The Fate of Human Societies*. W.W Norton & Co.

Esmaeilzadeh, H., & Vaezi, R. (2021). *Conscious AI*. Cornell University. https://arxiv.org/abs/2105.07879.

Etzioni, A., & Etzioni, O. (2016). *Keeping AI Legal*. Vanderbilt Journal of Entertainment & Technology Law. https://heinonline.org/HOL/Page?handle=hein.journals/vanep19&div=7&g_sent=1&casa_token=&collection=journals.

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., Novikov, A., Ruiz, F. J. R., Schrittwieser, J., Swirszcz, G., Silver, D., Hassabis, D., & Kohli, P. (2022). *Discovering faster matrix multiplication algorithms with reinforcement learning*. Nature. https://www.nature.com/articles/s41586-022-05172-4.

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Penguin Books.

National Highway Traffic Safety Administration. (n.d.). *Automated Vehicles for Safety*. https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

Schwab, K. (2016). *The Fourth Industrial Revolution*. World Economic Forum.

Stout, Dietrich. (2011). *Stone toolmaking and the evolution of human culture and cognition*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049103/.

Urban, T. (2015, January 27). *The AI Revolution: Our Immortality or Extinction*. Wait But Why. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html.