The Power of AI and its Resulting Ethical and Societal Implications

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Niels van Beek

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

The Power of AI and its Resulting Ethical and Societal Implications

The term "Artificial Intelligence" was originally coined in the 1950s, and was very generally rooted in the idea that machines or computers could one day display human intelligence (Helm, 2020). Now, 70 years later, artificial intelligence (AI) is finally coming to fruition – or so we think. Yes, it is true to believe that artificial intelligence is just over the horizon, but that is only in the over-simplified sense of the term. This is because when people hear "artificial intelligence," most immediately imagine a robot that can act entirely autonomously or a system that can execute a task from start to finish without human intervention. In reality, artificial intelligence is already here and it is currently playing a much larger role in our lives than most people think. It is extremely difficult to apply a singular, holistic definition of "artificial intelligence" as it comes in so many different forms and applications. For the purposes of this investigation, artificial intelligence is defined very generally as the ability of a computer or machine to gather information, determine the proper course of action to take, and execute the action. This generally takes the form of complex algorithms that determine things such as what shows up on your Instagram, Facebook, and Twitter feeds, the advertisements between your Netflix episodes, as well as more commercial purposes such as supply chain and finance. This is an extremely relevant and interesting topic to explore, as AI will continually see evermore usage in our society. There are nearly endless possibilities for the developments that AI could bring to our society in the future so long as we are cognizant of the massive power of this tool and therefore build it and use it in a responsible manner. This research investigated the magnitude and scope of AI, addressed some of its societal and moral concerns in the context of algorithmic biases, and considered possible avenues for solutions.

The Scope and Influence of Artificial Intelligence

Artificial intelligence is utilized in a myriad of different applications across a wide variety of industries, therefore having a massive sphere of influence over thought, society, and the future. This exploration looked at the societal implications of artificial intelligence within the context of how powerful it is, as well as its connection to the framework of sociotechnical systems, which are systems that have been designed with the primary objective of capturing the cooperation of social and technical components in our society. Currently, over 90% of leading companies are investing in AI technology (Lin, 2020). Some of the main ways that AI is being used today are in predictive modeling, supply chain logistics, automated processes, and personalization. For example, AI plays a huge role in determining the advertisements that you see on your media feeds, forecasting for weather services, as well as in more industrial purposes such as manufacturing robotics (Borana, 2016). The many roles that artificial intelligence is going to play in our future will only expand into new developments as well as increase in magnitude in where it is currently being used. Looking to the future, the global artificial intelligence market is expected to reach $267 billion by 2027 (Lin, 2020). AI is certainly here to stay, and it is up to us as the main creators and users to determine if it will foster innovation and equality, or do just the opposite.

Artificial intelligence may be one of, if not the most, complex sociotechnical systems that humans have ever created. Broadly defined, sociotechnical systems consist of the marriage and consideration of both social and technical factors when designing and creating a system. When applied to systems design thinking, Baxter and Sommerville (2011) describe sociotechnical systems as "the underlying premise of socio-technical thinking is … a process that takes into account both social and technical factors that influence the functionality and usage of computer-

based systems (p. 1).” This idea plays a fundamental role in the development of AI systems as our society moves into the future because we want the tools that we create to serve more than a technical purpose, but also a societal purpose. This highlights the very foundational idea of human nature and society because the ultimate goal of the tools that humans create is to progress our society both technologically and equitably. Therefore, we must keep the societal impacts of our creations at the forefront of our design process when developing new technology such as artificial intelligence. When designing our artificial intelligence systems we must remind ourselves that we want to live in a future where these systems help us achieve our goals not only technically but also socially. The idea of using technology to create our version of a truly better and more equitable society is called a sociotechnical imaginary. “The concept of socio-technical imaginaries is based on the study of how political cultures frame the objectives, risks, and benefits of technological developments and to what extent they address fundamental ethical and accountability questions” (Vincente & Dias-Trindade, 2021, p. 710). The sociotechnical imaginary that this research article envisioned was one where artificial intelligence serves its purpose of creating a society in which we want to live — not only by advancing our society technologically, but also by ridding itself of the social, ethical, and moral issues it currently comes along with, which was explored in the context of algorithmic biases.

It is very clear that AI is playing a fundamental role in driving some of our latest and greatest achievements as humans – and will continue to do so in the future – so we must utilize this sociotechnical design thinking process in order to ensure that AI is not only achieving its technological goals but is truly building a society in which we want to live. As it stands today, the AI systems we are building are only achieving one facet of this sociotechnical imaginary, and that is by advancing our society technologically. We are currently facing serious social issues

that have stemmed from the current way artificial intelligence is being created and built. In his exploration of the social responsibility of AI, Zhao states: “It is a double-edged sword. While letting people enjoy the good life created by new technology, it also allows people to feel its negative effects, such as infringing on human privacy, and bringing new inequalities to human beings” (Zhao, 2018, p. 1). We as a society, alongside our engineers, are the ones building these artificial intelligence systems that are currently bearing these negative side effects. This could stem from a myriad of reasons, whether it be unintended consequences due to oversight, a disregard for one’s impact of their work, or intentional malice. However, this does not mean that the future of artificial intelligence is bleak; because we are the ones building artificial intelligence we can enact positive change in the way we build these systems. For example, automobiles revolutionized human efficiency (like AI) by allowing us to travel faster and further than ever before. The first few decades of automobiles were extremely unsafe, until the seatbelt, airbag, and other safety features were invented. Preceding the invention of these innovations was the recognition of the dangers and the idea that we needed to design and build cars in a safer way. We as a society need to do the same for artificial intelligence, as it has the possibility to revolutionize the world in a magnitude similar to that of cars. In the same sense that the idea of safety in cars has taken effect in multiple areas – seatbelts, airbags, crumple zones, etc. – the revolution of safety and equity in artificial intelligence cannot be tackled one-dimensionally, but rather needs to take place in multiple areas whether it be cultural, educational, or ideological. This research has explored how to take a multi-dimensional approach to remedy one of AI’s dangers – biases.

Evidence of the Moral and Societal Concerns of AI

The widespread impact that artificial intelligence will have in our future means that we need to be hyperconscious about the implications of this technology. "Artificial Intelligence (AI) algorithms are widely employed by businesses, governments, and other organizations in order to make decisions that have far-reaching impacts on individuals and society" (Ntoutsi, 2020, p. 2). Yes, these systems may be mutable – as we are able to change the code – but that does not change the fact that these algorithms can "run off" with a mind of their own when implemented, causing damage before engineers can even recognize or fix the cause. For example, when coding these artificial intelligence algorithms, if we include immoral or unethical practices such as biases in the code, then the system will reflect that with the decisions it makes. For the purposes of this exploration, we will use biases as an example of why AI can be such a dangerous tool, as well as highlight the importance of using safe and ethical practices while designing these sociotechnical systems. Biases are a great example to use for this research because there are both implicit and explicit biases, so we are able to explore why we must not only be cognizant of mistakes or oversight in creating these systems, but also to see why we must enculturate people to care about the systems they create, so as to not build them with malicious intent. Regardless of intent, imparting an entity's, corporation's, or government's biases into a system as powerful and influential as AI is extremely dangerous. Even though it is nearly impossible to eliminate all biases from society, AI provides much scarier implications to these biases. If we code biases – intentional or unintentional – into the algorithms that dictate decisions in the day-to-day lives of billions, then these massive AI systems that we employ will perpetuate these biases to a degree of magnitude far larger than ever before possible.

For example, there is an implicit bias toward gender in the English language, when referring to the order of a list. Generally, the male counterpart is stated first (eg. man and woman, husband and wife, son and daughter), implying that there is a ranking order among these two groups (Leavy, 2018). This bias can be recognized by AI and utilized into perpetuity, possibly prioritizing or listing men before women in cases that may have more dire consequences, such as organ transplant lists, for example. Even though to humans, it may very simply be our diction and not affect the manner in which we create lists, having AI recognize this pattern and utilize it for its decision-making highlights how implicit biases that have been perpetuated by humans can have far more serious implications when employed in the context of AI.

From a more ethical standpoint, people may purposefully include explicit biases in their AI systems in order to bend users to their will or to realize a specific outcome. Again, this can have extremely dangerous results when magnified with the power and reach of AI. For example, Facebook uses artificial intelligence algorithms to determine the content that its users see every day, in addition to tailoring advertisements toward their interests and preferences. Any sort of bias in this algorithm can change what gets put in front of the eyes of billions of people. A recent development highlighted Facebook's ability to explicitly bend their AI algorithm according to their ideas, making news headlines. Due to developing information, on May 26, 2021, Facebook announced that it would no longer be censoring posts containing the claim that COVID-19 was man-made (Purnell, 2021). By no means has this exploration taken sides, nor has it explored the correctness of this ongoing debate. Rather, what is so alarming about this announcement is not that Facebook decided to stop censoring these posts, but the fact that they were explicitly doing it in the first place – consciously making the decision to censor ideas from the eyes of billions of

users simply because it did not agree with their own ideas. Even though the context in this scenario may be more trivial, the way in which it unfolded holds implications that are extremely worrisome. Using AI to advance one's own interests and suppress diversity of thought – especially at the scale made possible by AI – has extremely concerning implications because diversity of thought is what creates a more equitable, inclusive, and innovative world for us to live in. By employing bias in such a powerful tool to advance their own beliefs and stifling the free flow of diverse ideas to billions of users, Facebook showed us the dangers of an entity actively imparting its own explicit bias upon artificial intelligence.

Exploration of Solutions

In order to remedy the current problems we are witnessing with AI, we must not only think about fixing the systems themselves, but more importantly, we must address our cultural, corporate, and personal mindset as it pertains to building these complex sociotechnical systems. In the words of UVa President Jim Ryan, we must strive to be "great and good," which means to achieve great things while also ensuring that the resulting outcome does ethical and moral good for all of its stakeholders. This idea is very befitting to our development of AI systems because this is how we get closer to achieving our sociotechnical imaginary of a world where AI is able to effectively serve its purpose without the negative side effects we are seeing today. When applying the idea of being "great and good" to our construction of AI it means that we are not only striving to make these systems as technically powerful and productive as possible but that they must also reflect the values of our society by bringing good to the communities and people they serve. This idea has begun to gain more traction in recent years, with the Institute of Electrical and Electronic Engineers (IEEE) creating The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. The focus behind this initiative is to "recognize[d] the

need to have multi-disciplinary participation so as to best identify and address the socio-technical and design-oriented issues related to AIS (autonomous and intelligent systems) and ethics" (Chatila & Havens, 2019, p. 2). There are a few key ways in which people can address the underlying problems contributing to the adverse effects AI is currently causing in our society today. As mentioned earlier, as well as by the IEEE, the solutions will need to be multi-dimensional, or multi-disciplinary, as a one-size-fits-all solution will not be enough to solve the problems contributing to such a complex sociotechnical system.

In order to build our sociotechnical imaginary of a society where AI advancements are both "great and good" for us and our communities we must take a few distinct actions. These key actions we must take as we move into the future include: enculturating people to care about their work's impact upon society, changing corporate mentality to mitigate or account for the negative externalities that their product creates, and teaching engineers to feel a moral responsibility for the things that they create. This multi-dimensional approach is not only important for creating equitable artificial intelligence systems, but it can also be applied to other sociotechnical systems that we build as we forge ahead into the future. By tackling the problem from multiple angles, we are able to make the solution more robust and less susceptible to breakdown or lapses. The first step – or dimension – is enculturating people to care about the societal impact of their work. This will have a great impact on the following dimensions because it plants the seed of making people think beyond how much work they output, and instead think about how the output of their work is impacting their community or society. Secondly, we must change corporate mentality to account for the societal, environmental, or other impacts of their work. Ultimately, the purpose of innovation is to advance our society, not create value for shareholders. However, both of these things can happen concurrently. From a very high-level perspective, a negative externality is an

economics term where individuals, communities, or any third parties bear a cost from the production of a good or service even though they are not included in the production or consideration of production (Kenton, 2020). By changing corporate mentality, it is possible that corporations begin to make it the norm to account for the negative externalities that their product creates while performing their calculations of profits and returns to shareholders. This theory is also very often applied to pollution and the idea of a carbon tax. The premise of this idea is that companies need to pay some sort of compensation, or tax, because they are making a profit from their activities while the pollution it causes hurts the communities or environment around them. Instead of a tangible and measurable externality such as carbon pollution, we are now tackling a societal externality that is even more difficult to quantify. However, we are clearly able to see in the examples that these negative externalities exist with artificial intelligence systems, and we understand some of the underlying causes of it, so it is up to the creators of these systems to identify the causes and account for them. The creators of these systems, primarily corporations, employ many thousands of engineers to write and maintain the code for these intricate AI systems. These engineers are the next angle of attack in the pursuit of our sociotechnical imaginary. Engineers are the primary creators of artificial intelligence due to the fact that these systems are so complex in both hardware and software. Engineering education needs to be rooted in the idea that every individual has a responsibility for what they are creating, whether it be moral, ethical, or otherwise. By educating engineers to feel responsible for and truly care about the things they create, they ideally become less likely to follow immoral conduct leading to these negative externalities, or they may remain more diligent in identifying and avoiding oversight leading to unintended externalities.

By taking a multi-dimensional approach to our solution for mitigating the pitfalls of artificial intelligence, we are able to get some step closer to our sociotechnical imaginary of a world where these massive sociotechnical systems are able to advance the interests of humans without causing the seriously harmful negative externalities we are seeing today. It is clear that artificial intelligence is an extremely powerful tool that has immense promise and possibilities for human advancement – but with great power comes great responsibility. The biases – both implicit and explicit – that we encode into AI can have implications far beyond our intentions or our knowledge because we are now working with a tool that goes far beyond ever before possible, and for the majority of us – goes far beyond our understanding. This is why it is imperative we be hyperconscious of the methods in which we both create and employ artificial intelligence systems.

References

Baxter, G., & Sommerville, I. (2011). *Socio-technical systems: From design methods to systems engineering.* Interacting with Computers, 23(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Borana, J. (2016). *Applications of Artificial Intelligence & Associated Technologies*. 4.

Chatila, R., & Havens, J. C. (2019). *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.* In M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, & E. E. Kadar (Eds.), Robotics and Well-Being (Vol. 95, pp. 11–16). Springer International Publishing. https://doi.org/10.1007/978-3-030-12524-0_2

Choi, L. (n.d.). *Financial Stress and Its Physical Effects on Individuals and Communities*. 3.

Godsted, D., & McCormick, M. H. (n.d.). *National K-12 Financial Literacy Research Overview*. 20.

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). *Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions.* Current Reviews in Musculoskeletal Medicine, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Kenton, W. (2020, October 26). *Understanding Externalities*. Investopedia. https://www.investopedia.com/terms/e/externality.asp

Leavy, S. (2018). *Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning.* Proceedings of the 1st International Workshop on Gender Equality in Software Engineering, 14–16. https://doi.org/10.1145/3195570.3195580

Lin, Y. (2020, August 22). *10 Artificial Intelligence Statistics You Need to Know in 2021 [Infographic]*. https://www.oberlo.com/blog/artificial-intelligence-statistics

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). *Bias in data-driven artificial intelligence systems—An introductory survey.* WIREs Data Mining and Knowledge Discovery, 10(3), e1356. https://doi.org/10.1002/widm.1356

Purnell, N. (2021, May 27). *Facebook Ends Ban on Posts Asserting Covid-19 Was Man-Made.* Wall Street Journal. https://www.wsj.com/articles/facebook-ends-ban-on-posts-asserting-covid-19-was-man-made-11622094890

Vicente, P. N., & Dias-Trindade, S. (2021). *Reframing sociotechnical imaginaries: The case of the Fourth Industrial Revolution.* Public Understanding of Science, 30(6), 708–723. https://doi.org/10.1177/09636625211013513

Zhao, W.-W. (2018*). Improving Social Responsibility of Artificial Intelligence by Using ISO 26000.* IOP Conference Series: Materials Science and Engineering, 428, 012049. https://doi.org/10.1088/1757-899X/428/1/012049