AI for Social Good: Responsible Use of Artificial Intelligence to Solve Modern Social Issues

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

John Radossich
Spring, 2021

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

Signature _____ Date _____
John Radossich

Approved _____ Date _____
Hannah Rogers, Department of Engineering and Society

Abstract

As humanity progresses through different eras, a trend has emerged which closely correlates the technologies of the time and the direction of historical development. While these technologies have guided society to a peak in terms of human comfort, cutting edge advancements in artificial intelligence and machine learning have the ability to drastically change the functions of society. While there is no doubt that these technologies serve a multitude of purposes, the repercussions of fully committing to an automated world are uncertain and potentially dangerous. This paper aims to address some of these concerns, analyzing certain algorithms that have been irresponsibly developed and implemented into society. On the other hand, this paper will highlight several ethical uses of artificial intelligence for the benefit of a collective society. As billions of dollars are committed to research and development on these technologies, it is imperative to initiate public policies that not only inhibit these algorithms from harming society, but simultaneously incentivize engineers and their responsible integration of artificial intelligence into society.

AI for Social Good: Responsible Use of Artificial Intelligence to Solve Modern Social Issues

Introduction

Physicist Albert Einstein once said, “It has become appallingly obvious that our technology has exceeded our humanity” (Liou, 2018). As humanity transitions further into the Age of Information, the dichotomy of engineers will emerge in a polarizing manner. As technological capabilities break barriers of what was once possible, the engineering field will grow in both size and popularity. Accompanying this spike in technical interest is the division of engineers into two groups; those who answer the “how” question, and those who answer the “why” question. The latter are more focused on the underlying reasoning and purpose that drives their research and development. The prior, those who answer the “how,” are more focused on innovation without question, whether it be for monetary benefit or a true passion for engineering. As a result, this separation of motivation has led to an increasing ignorance for how this technological revolution can fix some of society’s larger issues.

In a developed country such as the United States, engineers are granted the privilege of working with the highest level of cutting edge technologies. This trend is highlighted by an emerging technological obsession with artificial intelligence. Endless resources are being allocated towards machine learning and deep learning algorithms to optimize the functions of society (Pohler, n.d.). However, very few of these solutions extend their capabilities to complex social issues that need to be addressed. With worldwide artificial intelligence spending anticipated to double from roughly $50 billion to $110 billion over the next four years, the application of these methods should be focused on amending various failing mechanisms of society (Soohoo, 2020).

As of August, 2020, the primary use of artificial intelligence technologies were in the enhancement of customer experience and facilitation of employee productivity. The deployment of these solutions are estimated to cover nearly a third of all artificial intelligence spending in the year 2021. More specifically, the banking and retail industries are anticipated to spend the most on AI solutions under the same forecast (Soohoo, 2020). These technologies certainly solve problems for the associated parties implementing them, however, there are more impactful areas of application that would benefit the collective society. Instead of being treated as a luxury product in advanced industries, artificial intelligence could serve as a catalyst for developing societies.

While catering to consumers and employees fits the capitalist nature of the United States, there are larger societal issues that can be reformed using cutting edge advances in technology. Analyzing this idea through the political technology STS framework, this paper uncovers several irresponsible uses of artificial intelligence. Conversely, several examples will be provided of ways that artificial intelligence can be implemented to solve complex social issues. Finally, there are many ethical considerations that this topic initiates, especially when viewing through a political scope. While the hype surrounding artificial intelligence certainly holds merit, to what extent should rapid advances in technology, namely artificial intelligence, overshadow our collective ability to use these technologies for social benefit?

Artificial Intelligence & Machine Learning

In order to untangle the intertwined aspects of technology, society, and policy, it is important to understand these issue's associated technical elements. For the purpose of this evaluation, these terms will be primarily limited to artificial intelligence and machine learning. It

is typical for these buzzwords to be used interchangeably, however; this misclassification of terms will cause issues when discussing regulatory policies that should be put in place. Thus, definitions and examples of these terms are provided in this section to ensure a foundation of basic understanding.

Artificial intelligence (AI) is a technical concept that seems to have exploded in magnitude and popularity over the past decade. Artificial intelligence is the notion of creating highly technical and functional machines that contain aspects of human intelligence. Compared to machine learning, AI is the highest level of classification, with wide spread applications and subclasses extending into various facets of life. While there does not exist an agreed upon standard to the extent that these machines must "think" like humans, functionality is primarily dependent on technological progress in areas of data accessibility, storage capacity, and graphics processing units (Copeland, 2009). Facial recognition algorithms and image processing tools are examples of narrow AI; at face value they are classified as artificial intelligence, however they are only working parts of machines that are not classified as such (NarrowAI, 2019).

Building on the premise that AI is currently limited by technology dependence, machine learning is a frequent approach in designing AI applications. Silicon Valley journalist Michael Copeland defines machine learning as "the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world" (Copeland, 2009). Machine learning methods are trained from a collection of data to understand the state of the system and perform a task as opposed to referring to specific, individually coded sets of instructions. This serves as a channel of efficiency as the machine does not require explicit "teaching", but a constantly developing, passive understanding of how to perform a task. Over the years, machine learning has most effectively been used to enhance a computer's "vision."

One example of a machine learning algorithm is the edge detection technology integrated into bank applications' check scanners. While the applications of these algorithms are still fairly primitive and restricted, improvements in data collection and learning algorithms will lead to pivotal practicality.

With the progress of this technology relying heavily on levels of computational power, it is safe to assume that these integrations are inevitable. With computational power doubling at a rate seven times faster since 2012, artificial intelligence will experience an imminent expansion into most industries as a means of optimizing the accuracy and efficiency of task execution (Hao, 2019). However, it is necessary to exercise caution when integrating such algorithms into a fragile society. With any emergent technology, knowledge gaps are unavoidable, making the synthesis between humanity and technology unpredictable. Engineering requires unpredictability to be handled carefully, assuming risks and uncertainties as barriers to an introduction to society. Furthermore, the stakes are high when it comes to artificial intelligence due to the strength of the technology and steep repercussions that may follow.

Political Technology Framework: Misuses of AI

In this section, the misuse cases of artificial intelligence and machine learning will be analyzed using the political technology framework. Specifically, this section will cover the unintentional introduction of politics on the lives of those wrongly affected by machine learning integrated systems. Currently, machine learning is implemented to perform a high variety of tasks, including: stock market and sports modeling, day trade execution, chatbots, media recommendations and much more (Mills, 2020). These applications have such flexibility as they require training data to "teach" the machine the interactions of life. However, this data is often collected, cleaned, manipulated, and processed by humans, adding a layer of inherent bias to the

model (Murray, 2019). Although typically unintentional, these methods of data manipulation lead to an ironic addition of human error to machine-driven algorithms. However, independent of the developer's goal, these models often make decisions that affect the lives of others. Again, the errors in these practices come back to the human element of them. Jason Bloomberg, President of Intellyx, captures this phenomenon when he says, "As human behavior makes up a large part of AI research, bias is a significant problem, data sets about humans are particularly susceptible to bias, while data about the physical world are less susceptible" (Murray, 2019). Thus, artificial intelligence misuses typically stem from the reflection of real-world biases that find their way into the way these models are trained.

One example of incorrect implementation of machine learning algorithms is the US court and corrections system. Specifically, a machine learning algorithm named COMPAS has been implemented in a Wisconsin court system to gauge the likelihood that a convict will commit another crime if released. While this may seem like an appropriate and socially beneficial place to integrate AI, it was found that the algorithm was biased against African American prisoners. In an investigation by ProPublica, it was found that the model incorrectly flagged black prisoners as more likely to reoffend by over 20% compared to those of a different race (Murray, 2019).

These biases have found their way into various mechanisms of the United States in similar forms. It has been proven that certain algorithms have biases against race, gender, religion, socioeconomic status, and geography (Müller, 2020). The next example of inappropriately integrated AI algorithms involves the new release of the Apple Card, a credit card designed by Apple to facilitate user spending through connectivity with their iPhone. While the card was highly praised upon its release, people began voicing their frustration for the inequality of credit decisions that the card's algorithm displayed. In many cases, men with low

income and low credit scores were receiving higher credit limits than females with high income and perfect credit scores (Wiltz, 2020). Despite the company denying credit preferences on factors such as race, age, and gender, there are many accounts that disprove the validity of their claims. In a more objective sense, a married couple with shared assets and perfect credit scores should theoretically receive close to identical credit limits. However, in these cases, husbands are given credit limits 10-20 times higher than their female partners (Vigdor, 2019). As previously stated, the underlying cause of these biases is a direct reflection of the biases that exist in our lives and are translated to the data that trains these algorithms. Thus, it is recognized that many machine learning processes are being inappropriately applied and require political regulation and testing before decisions are to be made from them.

Political Technology Framework: AI for All

While the irresponsible misconduct surrounding many of these algorithms is an issue in today's society, most would argue that the problem should be addressed at its core. In other words, it seems as though the selection of problems that AI is being used to solve is faulty and these technologies would be better off being applied in other areas. Specifically, there are many channels of sustainability efforts through which machine learning and artificial intelligence could be integrated. Many sustainability efforts are reliant on some third-party energy source which drives the collection of this energy. For example, wind farms, water currents, and solar rays are harvested and converted to raw energy. However, the collection process lacks optimization as these methods rely on a device such as a solar panel to remain in one location. With the political investment of resources into sustainability tools having machine learning implementations, these processes would collect energy more efficiently through the prediction of daily wind, water, and

solar maximums. Hence, these collection devices would employ these predictions and move to these spots accordingly (Haughey, 2020).

Healthcare is another area of application that machine learning has the ability to drastically enhance. Currently, the healthcare system lacks a universal network to store and learn from collected patient data. In fact, in some countries, there is a complete lack of any data collection at all. In Rwanda, Africa, most hospitals do not have a digitized database and their medical records only exist in paper form, impeding a doctor's ability to identify trends across all patients. There are many limitations of non-digital data, as advances in data processing and data analysis become completely irrelevant. A specific use case of machine learning algorithms is the digitization of patient medical records in Rwanda, Africa to work towards optimizing perioperative mortality rates (POMR) and understanding the signs of unsuccessful surgeries. The establishment of this system would enhance these hospital's medical record system which exists to gather patient data to formulate an accurate diagnosis and assign a treatment plan (Rickard, J., Ntakiyiruta, G., Chu, K., 2016).

As previously mentioned, image processing tools are a form of AI and could be used to scan medical records and adjust their grayscale to be easily read by a computer. These tools excel at edge detection, image cropping, and enhancing contrast so that a machine learning algorithm can "see" the data more clearly (Copeland, 2009). If enough training data is collected, it is possible to design a natural language processing classification model that can read handwriting and store the information in a database. Ultimately, through a fraction of the research and development that has already been poured into AI and ML, an entire country can streamline their data into a form that can be utilized to save lives. Connecting this database to the

endless supply of data analysis and visualization tools, the healthcare system as a whole can be studied and improved upon.

Ultimately, machine learning algorithms will always be considered a political technology due to the anatomy of their workings. Artificial intelligence and machine learning tools are only as strong as the data they are constructed from and as long as this data reflects a biased world, these machines are guaranteed to contain aspects of these biases. In reality, these machines are not (for efficiency's sake) explicitly coded to understand every state of a system and are instead geared to predict the world based on self-correcting assumptions and classifications. In some cases, this data is an extrapolation of humanity's bias over the course of history and thus, these machines have understood the world to be biased. A political technology is certainly a product of its inventor or creator, however, it is the technical system blended with modern politics that projects its effects on society (Winner, L., 1980).

Technical Recommendation

Throughout this analysis, artificial intelligence has been defined as a powerful technology, with incredible societal potential despite its ease of political misuse. While there is no correct perspective when valuing the rate at which countries accelerate technology, there is value in intermittent assessment of a system's current state. As research and development continues to grow exponentially, the world will begin to change at a rapid pace, independent of the collective status. Should these technologies continue to advance, the gap between the most and least developed countries will widen. Thus, it would be wise to reconsider this sprint to the future and focus these technologies on the development of the many imperfect systems that already exist.

While public policy can regulate the rapid development of these technologies, it is the engineers and decision makers that have the power to make the greatest difference. Regulation cannot be the only solution, but must act as a catalyst for companies and engineers to redirect their attention to these problems. AI and machine learning should not be regulated on the grounds of what it is, but instead on the grounds of how it can be used. These technologies are tools with infinite applications that require governance through the ways that people integrate it into society (Edelman, 2020).

The overall recommendation of this paper is to create a duality of policies: regulations that influence a responsible use of AI as well as regulations that safeguard from the misuses of AI. On one hand, these technologies should be incentivized into solving the world's environmental and societal issues. However, it is equally important to eradicate machine learning algorithms that inflict pain on others, whether it be intentional or not. Companies that choose to adopt an inherent desire to allocate their copious resources responsibly will make this change and enhance our humanity. When these issues and imperfections have been addressed and majorly corrected, then it might be the right time to commit to a world of automation and prepare for the new era of technology.

Counterarguments and Ethical Considerations

Artificial intelligence is a highly debated topic that introduces many ethical considerations for those who employ these tools. However, a reader of this critique might argue the limitations our country faces by restricting the usages of these automated technologies. An economist would argue against these points, claiming that technological progress is highly correlated with economic growth and that to help other countries, we must first be economically dominant. While this point is valid, and it is an unorthodox recommendation to regulate a

country's development, these regulations would aim to develop AI cautiously and correctly instead of rapidly and dangerously. Furthermore, as modern society has reached a relative technological peak in terms of comfort, humanity should focus on correcting the systems that may be flawed or outdated before launching an entirely new era. The recommendation throughout this analysis is to simply exercise caution throughout the pre-mature implementation of this technology; this is an entity that has not been studied before and therefore carries misunderstandings regarding the trickle down socio-political effects of affected systems.

Another critic of this analysis might argue that foreign countries might impose a military threat to the United States if they cease AI development. With AI having many military applications, it would seem a disadvantage to halt research and development in these areas. An answer to this person might include the historical repercussions of international technological races. The Cold War between the United States and the USSR, although never engaging in traditional battle, amplified feelings of distrust and hostility between the countries. While the military view of carrying a large stick is certainly valid, the United States is already the leading force in AI and ML development, delivering more than half of global AI spending - the argument is that this spending should be applied to areas of intercontinental and global improvement (Soohoo, 2020).

When delivering any recommendation regarding a technology of this scale, many ethical considerations involved must be considered. The first of these considerations deals with data, the main constructor of these tools and algorithms. While data can be used to responsibly understand the functionality and optimization of systems, it is important that the rights of humans are acknowledged and that their data is kept private. Unethical collection of data will only work against the trust that AI needs to build in order for it to be fully accepted and integrated into our

societies. Building off this point, it is critical to never lose sight of the larger scope relating to these problems. While it is certainly efficient to determine prison sentences or loan qualifications through an algorithm, there is a definitive extent to which these algorithms should be trusted over human judgement (Matthew Stewart, 2020). Lastly, the unintended consequences of these technologies beg ethical questions of the recipient of blame. Should an autonomous vehicle take the life of an innocent bystander, who is to blame? The technology, or the non-malicious computer scientist whose only fault was using the data they were given?

Conclusion

With artificial intelligence and machine learning having grown an obsessive following over the past few years, it is safe to say that humanity is entering a shift in eras. The technological transitions that will take place in the near future will affect society on a magnitude of great uncertainty. Being on the brink of this uncertain future, it is imperative that decisions are being made for the collective good of humanity, not for the fame that comes with the hype surrounding these technologies. Furthermore, it is important to understand that uncertainty typically increases the potential downfall of unforeseen outcomes. Having barely scratched the surface of this technology's capabilities, there are already plenty of accounts regarding faulty autonomous vehicles, discriminatory prison sentencing algorithms, and sexist credit loan algorithms. If these are the unforeseen downfalls at this stage of development, it begs the question of the future shortcomings these technologies will leak when scaled in proportion and function.

AI needs to be considered on a case-to-case basis, not as an individual entity. As machine learning algorithms find their way into United States systems, it is important to regulate the areas of their integration as opposed to regulation of the technology itself. Building on this, it is

imperative that algorithms are thoroughly tested and rid of implicit biases that naturally exist in data being processed by humans. A biased dataset will inevitably lead to a tool that yields biased results as the power of these technologies relies on the methods through which they are taught. Understanding the research and development growth trends surrounding AI, now is the time to consider the social ramifications that their implementation will produce. If these ethical concerns are delayed until post-integration, it will be too late to slow the momentum of their irresponsible effects. Alluding to the renowned physicist Albert Einstein, the true power of these rapidly developing technologies can only be optimized with an equal growth in our own humanity.

References

Copeland, M. (2019, November 06). The difference between ai, machine learning, and deep learning?: Nvidia blog. Retrieved March 02, 2021, from https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

DeepAI. (2019, May 17). Narrow AI. DeepAI. https://deepai.org/machine-learning-glossary-and-terms/narrow-ai.

Edelman, R. (2020, January 11). Perspective | Here's how to regulate artificial intelligence properly. Retrieved October 16, 2020, from https://www.washingtonpost.com/outlook/2020/01/13/heres-how-regulate-artificial-intelligence-properly/

Hao, K. (2020, April 2). The computing power needed to train AI is now rising seven times faster than ever before. MIT Technology Review. https://www.technologyreview.com/2019/11/11/132004/the-computing-power-needed-to-train-ai-is-now-rising-seven-times-faster-than-ever-before/.

Haughey, C. (2019, June 27). AI for Social Good: 7 Inspiring Examples. Retrieved October 16, 2020, from https://www.springboard.com/blog/ai-for-good/

Liou, C. (2018, February 7). Has Technology Exceeded Our Humanity? ST112 WA2018. http://web.colby.edu/st112wa2018/2018/02/07/test-3/#:~:text=It%20has%20become%20appallingly%20obvious%20that%20our%20technology%20has%20exceeded%20our%20humanity.&text=The%20ambiguity%20of%20the%20word,in%20a%20multitude%20of%20ways.

Matthew Stewart, P. (2020, July 29). The Limitations of Machine Learning. Retrieved October 16, 2020, from https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6

Mills, S., Baltassis, E., Santinelli, M., Carlisi, C., Duranton, S., & Gallego, A. (2020, October 08). Six Steps to Bridge the Responsible AI Gap. Retrieved October 16, 2020, from https://www.bcg.com/en-us/publications/2020/six-steps-for-socially-responsible-artificial-intelligence

Müller, V. (2020, April 30). Ethics of Artificial Intelligence and Robotics. Retrieved October 15, 2020, from https://plato.stanford.edu/entries/ethics-ai/

Murray, J. (2019, May 08). Racist Data? Human Bias is Infecting AI Development. Retrieved October 16, 2020, from https://towardsdatascience.com/racist-data-human-bias-is-infecting-ai-development-8110c1ec50c

Pohler, L. (n.d.). A Technological Perspective on Misuse of Available AI. Retrieved October 15, 2020, from https://www.consciouscoders.io/wp-content/uploads/2018/08/A_Technological-Perspective_on_Misuse_of_Available_AI-1.pdf

Rickard, J., Ntakiyiruta, G., Chu, K. (2016, April 4). Associations with Perioperative Mortality Rate at a Major Referral Hospital in Rwanda. World journal of surgery. https://pubmed.ncbi.nlm.nih.gov/26546186/#:~:text=Conclusions%3A%20The%20POMR%20at%20a,and%20transfer%20from%20district%20hospitals.

Soohoo, S. (2020, August 26). Worldwide spending on artificial intelligence is expected to double in four years, says IDC. Retrieved March 31, 2021, from https://www.iot-now.com/2020/08/26/104337-worldwide-spending-on-artificial-intelligence-is-expected-to-double-in-four-years-says-idc/

Vigdor, N. (2019, November 10). Apple Card Investigated After Gender Discrimination Complaints. The New York Times. https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

Wiltz, C. (2020, June 22). The Apple Card Is the Most High-Profile Case of AI Bias Yet. designnews.com. https://www.designnews.com/electronics-test/apple-card-most-high-profile-case-ai-bias-yet

Winner, L. (1980). Do Artifacts Have Politics? In Modern Technology: Problem or Opportunity ? (pp. 121-136). Cambridge (Mass.), MA: Daedalus.