Intelligent Decision-Making: Reinforcement Learning Applications in the Defense Environment

(Technical Paper)

Artificial Intelligence and the Death of Human Superiority

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

Joseph Banks

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

Rider Foley, Department of Engineering and Society

Rosanne Vrugtman, Department of Computer Science

Prospectus

Introduction

Over the last decade, a positive feedback loop of increased computing feasibility and the development of new revolutionary machine learning models have led to a boom in artificial intelligence. Economic and political interests drive demand for increased efficiency and productivity (Dieppe, 2021; Maier, 1977). Machine learning has become an increasingly popular tool to drive productivity gains, with corporate investment in machine learning tripling in 2017 and a $100 billion market by 2025 (Wellers et al., 2017).

The use of artificial intelligence to improve or replace human labor raises interesting social and political questions such as, will humanity become dependent on these technologies, what innate politics does machine learning possess (Garibald et al., 2018; Susser, 2019; Winner, 1980). This paper focuses on reinforcement learning, an area of machine learning focused on agent decision-making to maximize a cumulative reward. Reinforcement learning is specifically relevant to productivity and labor replacement since it focuses on intelligent decision-making similar to humans (Dayan, 2008). Recent advancement in reinforcement learning has led to agents surpassing human abilities at strategic games and routine labor, providing an avenue for increasing the efficiency and productivity of these human-occupied roles. Examples of this are Deepmind's AlphaGo beating a top Go player and reinforcement learning agents (RL agents) defeating experienced fighter pilots in simulated dogfights (DARPA, 2020; Chan, 2017).

Technical Project

Reinforcement learning is one of the most difficult areas of machine learning for computer scientists to research or use because of the difficulty of transferring it to the real world (Dulac-Arnold et al., 2019). Another difficulty of reinforcement learning is that the best-performing models from cutting-edge research require prohibitive compute costs and frameworks to manage a large number of simultaneous environments. For example, the Rain model required 34,200 GPU hours or 1425 days, thus making it infeasible for smaller research groups (Castro, 2021). Interest in reinforcement learning has gained steam since the emergence of deep reinforcement learning, deep learning in reinforcement learning, and distributed training. These introductions to the field have led to newsworthy breakthrough events such as an RL agent, AlphaGO, defeating a professional Go player (Chan, 2017).

In fall 2020, I worked as a Machine Learning Intern for Heron Systems, a government contractor focused on machine learning. I worked on designing, implementing, and training reinforcement learning models and improving their open-source reinforcement learning library, adeptRL (Adept). In Adept, I changed how the tensors, n-dimensional matrices, from environments are serialized and sent to the model so that more of the time is training and less is data traveling. I refactored the way data preprocessing worked in the library by splitting preprocessing operations into CPU and GPU operations. Because of these changes and the reimplementation of all their preprocessing functions, training speed increased 400%. I also helped refactor the league for self-play. Self-play in reinforcement learning is the process of having a model compete against different versions of itself so that it can train for thousands and thousands of hours without researchers needing to collect data. The league is the idea of saving the model at different stages of training and then selecting from these saved states of the model to compete against. The league tracks the performance of the models against each other and can select models of similar skill to train against. The league also provides for an objective measure of the skill level of particular models and reward schemes.

Heron Systems gained international recognition when it won the DARPA AlphaDogfight trials in August 2020 that focused on building dogfighting artificial intelligence (AI) (Lawler, 2020). To finish the competition Heron System's AI defeated an experienced human pilot fighter pilot in all five dog fights in the simulator (DARPA RSS, 2020). My work on improving Heron Systems reinforcement learning infrastructure continues to help their development of dogfighting AI for the US military and other AI agent applications. Dogfighting AI is a perfect example of the replacement of skilled human labor to increase efficiency and productivity.

One of the important factors in Heron System's ADT victory was the time spent training in self-play. My work on Adept and improving its preprocessing and league contribute to increasing the time models spend training. Helping lead to superior agents that are more likely to outstrip their human counterparts and replace them.

Human and Social Dimensions and Connection to Technical

The improvement of reinforcement learning has many far-reaching human and social implications. According to the framework laid out by Langdon Winner in "Do Artifacts Have Politics?" (1980), technology is inherently political. Winner uses the dangers presented by nuclear energy as evidence for why societies have and will erect laws and regulations around the technology to mitigate potential disasters. Similar to that example, AI with the capability of dogfighting presents many risks to society, such as in the wrong hands providing tech-savvy bad actors military capability outside the ability of local police forces to handle. Governments can increase regulation and governmental control to allay the emergent risks. Thus through Winner's technopolitical lens, AI has inherently authoritarian attributes.

AI's current use by authoritarian regimes and its ability to manipulate decision-making highlights its inherent politicalness and lean-to authoritarianism. The Chinese government currently uses facial recognition technology to prevent foreign nationals from appearing on live streams broadcast in China (Everington, 2020). They also are using it to notify police of the presence of Uyghurs, a Turkic ethnic group in China (Bhuiyan, 2021). These are clear examples of the Chinese government using AI’s superhuman abilities to enforce stricter information control by preventing foreigners from streaming to its citizens and more efficient widespread political persecution of unwanted minority populations such as the Uyghurs. There also exist invisible ways that authoritarian regimes can utilize AI to maintain their power. Choice architectures are decision-making contexts, the design of how decisions are presented to users. An example of choice architecture is donor card opt-in versus opt-out. People are much more likely to be donors if it is the default despite the options staying the same (Johnson & Goldstein, 2003). Changes in the choice architecture can influence behavior in desired directions (Susser, 2019). The use of AI allows for adaptive choice architectures, personalized choice architectures targeted at specific users based on their preferences, thus making manipulation of user behavior easier for outside actors such as companies and governments. Adaptive choice architectures can be designed and created to suppress dissent and sway user opinions to the state's desired stance.

The academic and corporate use of AI further highlights its lean-to authoritarianism and tendency to centralize power. Evidence of these tendencies in academia is provided by a paper that reviews the encoded values in machine learning papers (Birhane et al., 2021). The authors created an annotation scheme, manually reviewed and annotated the publications, and confirmed intercoder consensus. The results showed that the values currently encoded in machine learning papers centralize power and disproportionately favor large and powerful organizations and corporations. "The rise of automation: How robots may impact the U.S. labor markets" (2021) from the St. Louis Fed posits that the world may be on the verge of the fourth industrial revolution as large swaths of the labor market can be automated with improvements in AI. The article also reveals that the automation of jobs is currently polarizing the labor markets as job loss is taking place primarily among routine labor. Recent research out of the National Bureau of Economic Research shows that labor automation is the driver of increases in income inequality over the last four decades in the US (Acemoglu & Restrepo, 2021). These are clear evidence of AI's inherent tendency to centralize power and lean-to authoritarianism.

Research question and research

One of the looming questions of the next century is: What will the geopolitical consequences, specifically in the US and China, of large-scale replacement of human labor with AI be? The importance of this question lies in how widespread the application of AI for currently human labor tasks will be. The jobs at risk of automation range from simple truck drivers to more complex and regulated roles such as fighter pilots in the military. Analyzing and attempting to predict the effects will help create more informed regulations and help mitigate potentially dangerous situations.

My research will use Winner's technopolitical framework to answer the question. This framework will allow for the inherent political attributes of AI to be extracted and analyzed. Answering the question satisfactorily, various types of sources will have to be examined and interpreted. For example, the paper mentioned earlier that extracts encoded values in machine learning papers provides insight into the views and values of the current industry (Birhane et al., 2021). There are also a plethora of data points to the political attributes of AI, such as the Chinese government's use of AI to identify Uyghurs (Bhuiyan, 2021) and enforce repressive policies such as banning foreign nationals from broadcasting without explicit permission (Everington, 2020). The evidence of AI's use will be compared and contrasted with its use in other highly influential countries/regions such as the US and the Eupearon Union. The recent comparative success of AI research in China versus the US (Li et al., 2021) will be analyzed in the context of AI's inherent political attributes and their alignment with the two government structures and their expressed values.

Finally, literature on how automation affects offshore manufacturing and labor can provide data on how AI will affect the US and Chinese relations as China is the world's largest manufacturer. Analyzing this may provide insight into the factors motivating the competition for AI dominance between the US and China. There exists literature on the offshoring of labor and manufacturing and its geopolitical ramifications.

Conclusion

Political and economic interests have an insatiable hunger for efficiency and productivity. The rise of superhuman machine learning provides an avenue for vast gains in efficiency and productivity across the economy and the government, similar to the industrial revolution (Bharadwaj, 2021). The contributions made during my internship are helping one of the companies at the forefront of human replacement with AI. The demand for these technologies and their widespread potential impact make them of the utmost importance to the United States and the world. This research paper should shed light on many of the potential effects and dangers of AI and its replacement of human labor, by analyzing the encoded values, inherent political attributes, and impacts on the labor markets. It can hopefully help illuminate a way forward for the technology that minimizes the harmful effects while capturing the enormous benefits.

References

Acemoglu, D., & Restrepo, P. (2021, June 7). Tasks, Automation, and the Rise in US Wage Inequality. *NBER Working Papers*(28920), National Bureau of Economic Research, Inc. https://www.nber.org/papers/w28920

Bharadwaj, A., & Dvorkin, M. D. A. (2021, April 10). *The rise of automation: How robots may impact the U.S. labor market: St. Louis Fed*. Saint Louis Fed Eagle. Retrieved from https://www.stlouisfed.org/publications/regional-economist/second-quarter-2019/rise-automation-robots.

Bhuiyan, J. (2021, February 09). Major camera company can sort people by race, alert police when it spots Uighurs. Retrieved from LA Times website: https://www.latimes.com/business/technology/story/2021-02-09/dahua-facial-recognition-china-surveillance-uighur

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2021). The values encoded in machine learning research. ArXiv:2106.15590 [Cs]. http://arxiv.org/abs/2106.15590

Castro, P. S. (2021, July 13). Reducing the Computational Cost of Deep Reinforcement Learning Research [web log]. Retrieved from https://ai.googleblog.com/2021/07/reducing-computational-cost-of-deep.html.

Chan, D. (2017, October 20). *The AI that has nothing to learn from humans*. The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450/

DARPA (2020). *AlphaDogfight Trials Foreshadow Future of Human-Machine Symbiosis*. DARPA RSS. Retrieved from https://www.darpa.mil/news-events/2020-08-26.

Dayan, P., Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience* (8)**,** 429–453 https://doi.org/10.3758/CABN.8.4.429

Dieppe, A. (2021). *Global Productivity : Trends, Drivers, and Policies.* Washington, DC: World Bank.https://openknowledge.worldbank.org/handle/10986/34015

Dulac-Arnold, G., Mankowitz, D., & Hester, T. (2019). Challenges of Real-World Reinforcement Learning. *ArXiv:1904.12901 [Cs, Stat]*. http://arxiv.org/abs/1904.12901

Everington, K. (2020, July 9). Douyin's facial recognition tech blocks foreigner from livestream. Retrieved from Taiwan News website: https://www.taiwannews.com.tw/en/news/3963271

Johnson, E. J., & Goldstein, D. (2003). Do Defaults Save Lives? In Science (Vol. 302, Issue 5649, pp. 1338–1339). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.1091721

Lawler, R. (2020, August 21). Heron Systems' AI pilot just beat a human in a simulated dogfight. Retrieved from Engagdet website: https://www.engadget.com/alphadogfight-ai-f-16-pilot-025617519.html

Li, D., Tong, T. W., & Xiao, Y. (2021, February 18). Is China Emerging as the Global Leader in AI?. *Harvard Business Review*. Retrieved from HBR website: https://hbr.org/2021/02/is-china-emerging-as-the-global-leader-in-ai

Maier, C. S. (1977). The Politics of Productivity: Foundations of American International Economic Policy after World War II. *International Organization*, *31*(4), 607–633. http://www.jstor.org/stable/2706316

Susser, D. (2019). Invisible influence. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3306618.3314286

Wellers, D., Elliot, T., & Noga, M. (2017, May 31). 8 Ways Machine Learning Is Improving Companies' Work Processes. *Harvard Business Review*. Retrieved from HBR website: https://hbr.org/2017/05/8-ways-machine-learning-is-improving-companies-work-processes.

Winner, L. (1980). Do Artifacts Have Politics? In Daedalus, Vol. 109, No. 1, *Modern Technology: Problem or Opportunity?* (pp. 121-136). The MIT Press.