Security For Databases and Throughout Computer Science Curriculum
(Technical Report)

An Analysis of Machine Learning and Artificial Intelligence on Climate Change Through Capitalism
(STS Topic)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Sciences
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Franceska Padilla Coo

November 12, 2023

Technical Team Members:
Nada Basit, Ph.D

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.

ADVISORS
Prof. Pedro Augusto P. Francisco, Department of Engineering and Society
Nada Basit, Department of Computer Science

Introduction

The problem I wish to address in my research is the environmental impact of machine learning and Artificial Intelligence (AI). Specifically, I explore the sociotechnical problem of why, despite the drastic and urgent calls for more environmentally friendly actions as climate change heightens, society continuously overlooks the strenuous demand of resources required to support machine learning and AI for the sake of its abilities and its economic benefit. Machine learning and AI have enabled society to accomplish many things. The ability to predict outcomes has drastically improved, and informed decision-making recently often stems from these two rising fields. Machine learning and AI base themselves on mass amounts of data, which is incredibly powerful in the metaphorical sense for enabling society with new capabilities, and in the literal sense, in terms of resource demands. Like other computer technologies, production contributes a significant amount to carbon emissions. However, machine learning and AI eat up even more energy resources due to the databases that store the data used for teaching algorithms and due to the power required to run such high-energy actions. “Data mining and computation evaluations of persons and corporations have far-reaching environmental costs” (Brevini, 2020, 1). The production of machine learning and artificial intelligence mirrors that of computer production in that it generates a significant amount of carbon emissions, but because of how resource-demanding machine learning and artificial intelligence are in terms of power and physical storage, machine learning and AI contribute significantly more to climate change and carbon emission production. Even more, data collection increases exponentially, and the growth requires the construction of physical storage and more physical computer production.

My capstone research applies to this issue of wasteful and insecure storage. One major contribution to the construction and therefore production of harmful chemicals is the failure to maximize the usage of preexisting physical storage spaces by creating smart database structures with minimized redundancy and information security. Hackers can corrupt data, leading to company revenue loss, which prevents continuously improving database upkeep, or to the corruption of the actual physical hardware, eliciting the striking down of one data center and the creation of another. In this way, my research demonstrates the daily role of software engineers and developers to proactively prevent contributing to the creation of harmful chemicals released into the atmosphere.

Despite this contribution to carbon emissions and climate change, the abilities of machine learning and AI often overshadow any of their negative impact in society's eyes. Steps for a more sustainable environment lose their priority because of how powerful machine learning and AI can be. Contributing even more to the societal importance of machine learning and AI is their contribution to the economy. As a new technology, fascination and excitement encourage investment and stimulate the economy. Instead of focusing on how to lessen the carbon footprints of machine learning and AI, the priority becomes generating new models and algorithms and improving them for societal consumption and usage. In my Science, Technology, and Society section of my paper, I explore this societal behavior demonstrating a preference for the immediate economy over technology's environmental impact. According to scholars like Rai, Rawat, and Lippert, this notion must be changed due to the advent of the climate crisis.

To further this already harmful impact, the fascination surrounding machine learning and AI has skyrocketed their usage to levels that prevent the generation of effective plans of action to handle their environmental contribution. The economic importance of machine learning and AI

is also speculated to impact the “black box” notion of AI. Scholars like Lippert and Brevini speculate that major economic powers keep society in the dark of how machine learning and AI operate, especially their contributions to climate change, to continue with its economic stimulation. This prevents the dissemination of knowledge surrounding machine learning and AI’s impact, leaving the technologies’ advancement unobstructed to continuously benefit the economy. Because of the public’s demand for machine learning and AI capabilities to maximize and grow the performance of the economy, society overlooks and diminishes the environmental impact of machine learning and AI.

Technical Topic

My capstone focuses on the use of databases and database security. Because of the prevalence of data, and, especially these days, the degree of confidentiality embedded in data, database security plays an important role in ensuring that information does not get leaked or tampered with. Data collection itself already gets plenty of backlash and criticism from the public, exemplified by the condemnation of TikTok by older generations as spyware material. The public commonly requests the granting of permission from subjects before data collection and transparency as to the usage of that data, now a legal requirement and seen with Google, Amazon, Facebook, and many other large companies. Therefore, even more care applies to data storage. When creating databases, not only security must be considered, but also the physical storage of the data plays a huge role in security and general data formatting. The way we keep data can either eat up at resources or can be optimized to be compact and secure. For reference, the International Energy Agency estimates data centers’ to use around 200 TWh a year, “more than the national energy consumption of some populous countries such as Iran” (Brevini, 2020,

1). As the demand is anything but insignificant, resource allocation, structure, and security should constantly be the priorities with database creation and construction.

Database construction and maintenance must consider the portability of the data, which refers to redundancy and dependencies. Redundancy and dependencies make data confusing and difficult to sift through. Mix-ups occur in the relationships, provide more vulnerabilities because of repeated information and accessibility in several places, and can lead to data loss. Database portability also enables data transfers to different storage locations. The more compact data is, the less throughput on the traveling path for faster collection and retrieval. Well-constructed data minimizes the amount of data transfers and data traffic while still providing a lot of information quickly.

Database security also falls under the subject of database construction. Data insecurities can render the data insignificant, skewing outcomes. If the collected, skewed data becomes the basis for any studies, results of the study could mean nothing or could lead to the implementation of something non beneficial or possibly even harmful. One famous instance of this was Amazon's implementation of a hiring AI, which heavily favored male applicants due to the lack of female applicants in the hiring pool. Other consequences include studies referring to statistics reporting completely incorrect metrics, reporting flawed analyses, and influencing the perceptions of the subject. Data insecurities could also end up wrongly informing people about a topic, leading to useless methods. One example includes data loss impacting an entire pipeline, such as data loss leading to an incorrect target demographic and leading to marketing strategies that do not actually help with advertising. In this case, data transfers through different sources and storage locations serve as a vulnerability as data can be siphoned during transport. Data security ensures that this data remains encrypted and used only by those authorized to do so.

Data insecurity could also provide data to the hands of the wrong people. This could be the revelation of private information, such was the case with the data loss crisis of First American Financial, which leaked the insurance information of around 900 million people, but this has also taken the form of people twisting data to discriminate against different demographics and populations, specifically black and/or indigenous people of color.

I work on enforcing proper data storage, ensuring that data is grouped in a way for easy access and readability. Although common and incredibly basic, attacks on databases occur frequently because of poor management. Still, the structure of the data is well protected using preemptive measures that prevent easy data access. Methods like this include verifying the users attempting to access data, checking for malicious access requests, and restricting the types of data entered. Even more, I encourage database designers to incorporate as many database design guidelines and security measures as possible to set a standard in database creation.
Without these large databases, machine learning and AI would not have enough knowledge to learn and predict from. These operations require mass amounts of data to foresee patterns and account for different anomalies and learn. To continue collecting data, we must continuously prove the safety and proper management of data already on hand.

STS Topic

My topic surrounds the environmental impact of machine learning and AI, but I wanted to explore why the impact of machine learning and AI on the environment is not more well-known and stressed in public discourse, especially due to the increasing amounts of calls for environmental action on climate change. The minimal coverage of green computing contributes to this general lack of information. Green computing, which attempts to make computer

producers more cognizant of the environmental impact of computational technology production on the environment and therefore reduce the rapid usage of resources and provide a more sustainable approach to computer production, only appeared relatively recently in history in comparison to other topics. Green computing emerged in the 1990s with the emergence of the Energy Star Programme when more urgent demands for environmental action arose (Rai, et al., 2023, 112). Because of its more recent introduction, the topic of green computing has not completely pierced into the main computer science field as its own topic. Additionally, some scholars like Brevini and Lippert speculate that green computing and other attempts at reducing carbon emissions in technological production pose a threat to many producers as its initial implementation could potentially harm the immediate economy. This is even more applicable to machine learning and AI as their introduction to society has caused a stir from the public, and many producers hope to capitalize off their success, implementing their own versions of machine learning algorithms and AI.

Machine learning and AI currently have society in a chokehold. As Konig, Wurster, and Siewert discuss, society has handed over a sense of autonomy to AI since its introduction. Because of the notion that AI makes informed decisions, especially because it makes decisions based on measured data and numeric data, society has handed over its ability to choose and make decisions to AI. This reflects the importance society places in AI. The human ability to forgo instinct and choose their next action separates the human species from many other living creatures. However, handing this empowering ability over to a machine indicates trust and a belief in a superior thought process. Society, therefore, holds AI in a high regard, a higher regard than the environment. Konig, Wurster, and Siewert also analyze how society prioritizes AI's contributions over its environmental impact. This means that society will probably forgo the

health of the environment for the sake of AI's powers. Notions like these impact our future as the climate crisis worsens, leaving society to "again find themselves at an important crossroad, where decisions taken today and behavioral patterns taking hold now may shape developments and environmental impacts of the technology over the next decades" (Konig, et al., 2022, 1). One thing trumps AI's capabilities, though. Society's view on the value AI provides relies on the easy access to that type of machinery and function, where easy access especially refers to pricing. Society holds free and accessible AI at a higher position than AI locked behind a price. This indicates two things: that society tends to want the most value from their investments and that economy/financials, artificial intelligence, and then the environment, best reflect society's current order of priorities. Velkova also reiterates this notion. Using the example of "waste heat recycling" in European cities, Velkova demonstrates how society readjusts to keep using data centers and AI. Even more, Velkova claims that this importance stems from the fact that society acknowledges the investment of a significant amount of time and money in these technologies, "determined by the interested parties in the production process" (Velkova, 2016, 4). Society therefore wants to maximize their benefit rather than scrap the technology. Again, this further indicates a priority of the economy, then technology like AI, and lastly, the environment.

This problem will be analyzed using utilitarian ethics. The reason why utilitarian ethics will be used is because, based on initial research on the topic, the value of AI depends on how much the public thinks it will help them and how useful the results of machine learning and AI can be to them. Each other topic such as the environment and the economy can also be analyzed within this context, and because of utilitarian ethics, these topics can be in conversation with each other from the perspective of society without having to know the details of how each works. Evidence will be gathered using literature, reports, and media. Literature, reports, and media can

better capture the different feelings across the globe over surveys and interviews which can only really be beneficial locally. Literature, reports, and media also provide information on the economy, AI, and the environment, which may be too much to ask of one person. The societal perspective is preferred in this context as the environment cannot change with just one person's actions, but also value is derived from a collective audience. The information gathered on AI and its value in the context of the economy and of its impact on the environment will be analyzed using case studies, sustainability, and consequences. The case studies provide context on values. Sustainability corresponds with the environmental aspect, especially as this topic pervades over some time, and consequences refer back to utilitarian ethics. Consequences will allow an analysis of what society might find valuable.

Conclusion

Machine learning and AI fall under technologies that could improve with green computing. Not only would it lessen their entire lifespan's environmental footprint, but it would also benefit each specific consumer in the long run in terms of personal finances and maintaining the environmental ecosystem here on earth. However, because this benefit would require time and the adjustment of already established procedures, society prefers the more immediate economic benefit of forgoing sustainable procedures and proceeding with generation of new machine learning and AI technology. Even when society attempts to account for the environmental impact of machine learning and AI technology, the driving force remains economic in that consumers want to get as much value as possible from investments. As machine learning and AI expand and improve, society is at a crossroads to either prioritize economic and

technological stimulation or readjust the current systems to address the impending climate crisis.

References

Awasthi, K., & Sameer Awasthi. (2023). Green Computing: A Sustainable and Eco-friendly Approach for Conservation of Energy (A Contribution to Save Environment). Sustainable Computing: Transforming Industry 4.0 to Society 5.0. 319-333.

Brevini, B. (2020). Black boxes, not green: Mythologizing artificial intelligence and omitting the environment. Big Data & Society, 7(2). https://doi.org/10.1177/2053951720935141.

Gough, C., OAPEN, O'Reilly Online Learning: Academic/Public Library Edition, & DOAB: Directory of Open Access Books (2015). Energy Efficient Servers: Blueprints for Data Center Optimization. New York, NY: ApressOpen.

König, P. D., Wurster, S., & Siewert, M. B. (2022). Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis. Big Data & Society, 9(1). https://doi.org/10.1177/20539517211069632.

Lippert, I. (2016). Failing the market, failing deliberative democracy: How scaling up corporate carbon reporting proliferates information asymmetries. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716673390.

Rawat, Y., & Priti Rai. (2023). Trends and Challenges in Green Computing. Sustainable Digital Technologies: Trends, Impacts, and Assessments (1st ed.). CRC Press. https://doi-org.proxy1.library.virginia.edu/10.1201/9781003348313.

Sharma, R., Śarmā, D., Bhatt, D., & Pham, B. T. (Eds.) (2022). Big Data Analysis for Green Computing: Concepts and Applications. Boca Raton, FL: CRC Press.

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and Policy Considerations for Modern Deep Learning Research. Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), 13693-13696. https://doi.org/10.1609/aaai.v34i09.7123.

Takiguchi, H., & Morita, K. (2009). Sustainability of silicon feedstock for a low-carbon society. Sustainability Science, 4(1), 117–131. https://doi.org/10.1007/s11625-009-0069-1.

Velkova, J. (2016). Data that warms: Waste heat, infrastructural convergence and the computation traffic commodity. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716684144.