ARTIFICIAL INTELLIGENCE: USING MACHINE LEARNING TO IDENTIFY DANGEROUS MILITARY SYSTEMS

ETHICAL AND LEGAL IMPLICATIONS OF MACHINE LEARNING USE IN ART CREATION

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

By
Alex Joon Kim

October 27, 2023

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

Rider Foley, Department of Engineering and Society

Rosanne Vrugtman, Department of Computer Science

Introduction

Machine Learning is a rapidly progressing field that has gained significant interest in the 21st century. Stemming from Artificial Intelligence (AI), Machine Learning (ML) is unique in its ability to “self-learn” from training data and improve upon itself over time without explicit programming. ML technologies learn through automated processes where they are able to recognize patterns from data and learn from them in order to make optimal decisions in the future. As a result, Machine Learning technologies are able to emulate and expand upon human intelligence, creating machines capable of outperforming human beings with average intellect in complex tasks and subtle pattern recognition (Anjila, 1984; El Naqa & Murphy, 2015). The capabilities of ML have motivated many industries to replace human labor with integrated ML technologies to complete their critical operations.

One such industry that could integrate ML technologies is System Safety Engineering within the military. System Safety Engineering is a field focused on the planning, identification, documentation, and mitigation of hazards that contribute to mishaps in military systems. In other words, it is a compilation of engineering analysis and management practices that control dangerous situations (Bahr, 2015). The integration of ML will be critical in streamlining engineering analysis and preventing dangers in military system use. Through automation, ML technologies will be able to handle important responsibilities such as the detection/analysis of dangerous components within a military technical system before its use. As a result, ML technologies can be harnessed by System Safety Engineers to perform their daily responsibilities and revolutionize their work to uphold their values far more effectively.

As with any technology, however, machine learning technologies impose significant moral and ethical implications during its use (Winner, 1980). Depending on the scale of control delegated to ML technologies, misuse or malfunction of such technologies could contribute to devastating consequences including dangers to human life and property (Leveson, 1991). This is seen in automated military software systems entrusted with preserving the human lives, where the malfunction of these technologies typically result in dangerous military accidents leading to the loss of human lives (Ding, 2023). The risks that arise from military software system malfunction/misuse raises significant moral/ethical concerns regarding the use of ML technologies in high stake fields. However, addressing these concerns remain difficult due to the lack of established frameworks provided to developers to implement the technology successfully (Myllyaho et al., 2022). In this paper, I will focus on the implementation of an ML capability oriented towards System Safety Engineers by leveraging and understanding the moral and ethical implications imposed by ML technologies throughout its application.

Machine Learning to Protect Military System Users

Completing the tasks of System Safety Engineers have become ever more difficult due to the complex and dynamic behavior of modern military technological systems (Harvey & Stanton, 2014). Currently, System Safety Engineers rely on system safety accident models that have their roots in industrial safety, a priority that was more applicable to technical systems before World War II (Leveson, 2002). These traditional system safety accident models struggle when analyzing current military systems that have adopted new and unfamiliar technologies, making them less applicable in the field of System Safety Engineering today. In fact, their use against current large-scale military systems has contributed to an overall lack of analytical substance that is necessary for System Safety Engineers to make adequate sense of these systems (Bakx & Nyce, 2017). In order to assess these military systems in a productive manner, it is evident that the field of System Safety Engineering must be technologically revolutionized with ML and adapt to the fast pace of technological change today.

The Potential Hazard Identification via Artificial Intelligence (PHIAI) capability addresses this issue by incorporating high levels of computer automation from ML technologies with high levels of control from System Safety Engineers in order to effectively identify potential dangers within military technical systems and streamline important hazard analysis tasks through automation. These high levels of human control and computer automation are applied through the Human-Centered Artificial Intelligence framework, a set of development guidelines which serve as the basis for the PHIAI capability (Shneiderman, 2020). The capability begins by training several machine learning models on technical system design data supplied directly by System Safety Engineers, providing them full control on how the machine is able to learn. The capability then allows System Safety Engineers to choose which ML models within the algorithm will be used to identify potentially dangerous subsystems of a military system's design. The capability trains and identifies hazards based on the verbiage used in the system's design descriptions, and exists solely in a software domain.

The PHIAI capability's design improves upon the designs of current methodologies by prioritizing usability and efficiency to maximize the performance of the engineers and technology alike. Having high levels of human control on the capability and its training data allow System Safety Engineers to choose and understand what the capability will learn and how it can identify hazards. This will make the capability adaptable to any system design input supplied by the engineers and easily adoptable by new users. High levels of automation as the capability performs its tasks allows System Safety Engineers to prioritize other important tasks while delegating hazard identification responsibilities to the capability. Since anticipating and controlling hazards at the design stages of an activity is the cornerstone of a system safety effort (Roland & Moriarty, 1990), reducing time on this stage will make the daily operations of System Safety Engineers much more efficient.

Machine learning technologies prove to be applicable in preserving the lives of military system users. However, machine learning also raises important moral and ethical implications in its use. I will examine these implications within ML's use in the creation of artwork and how they shape the roles of artists and the art industry alike.

Machine Learning's Implications within the Creation of Art

Machine Learning is often associated with technological industries but has experienced increased integration within cultural industries such as the production, curation, and analysis of art (Cetinic & She, 2022). However, ML has brought on significant implications and responsibilities to artists and the art industry alike. These implications will be unpacked using the framework of Actor Network Theory from Bruno Latour's work in *Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts*.

Actor Network Theory (ANT) can be best understood as an analysis of the relationship of responsibility between the human and non-human components (or actors) in a sociotechnical system and the influence non-human actors have in shaping society (Latour, 1992). Latour demonstrates these ideas through his example of the door, where humans delegate the responsibility of allowing them to freely enter enclosed spaces to the door but are entrusted with responsibilities by the door to close it after its use. Like the door, artists and ML technologies must both take on responsibilities in order to allow ML to beneficially impact the art industry. Doing so will require power limitations on ML technologies so they can only serve as a source of creativity rather than as an artist itself and require artists to constantly practice and maintain a conscious understanding of the legal, ethical, and societal implications that arise when using it as such.

The use of ML to create art has shaped the roles of artists and their work significantly by imposing important responsibilities and ethical implications on their work. Prompt-based generative machine learning systems have made significant contributions to art but are inherently limited by their inability to produce original work and rely on creative human input (Hageback & Hedblom, 2022). In ethical, moral, and legal terms, issues arise from these limitations since the ML model’s training data is built off of previously authored art (McCormack et al., 2023). While artists delegate the responsibility of art creation to their ML technologies, its lack of human consciousness causes it to impose the responsibilities of creative input and ethical and legal usage to its users. The issues concerning authorship that stem from the usage of prompt-based generative ML models is synonymous to Latour’s examination of the delegated responsibilities given by the door to the humans. Both present the non-human influence on society from ANT.

Societal implications arise from the increased amounts of ML-generated art that is available for public viewing. Studies have proven that society prefers human-created art over ML-generated art as study participants exercised extensive engagement to derive narratives and emotions from artwork with a “human” label but felt cognitively obstructed when engaging and deriving meaning out of artwork labeled as “AI” generated (Bellaiche et al., 2023). The studies present a societal dilemma where art generated by AI and ML fail to deliver artistic meaning to artist audiences independently. Therefore, in order for artists to use ML to create meaningful art, artists must reshape their relationship with the technology; artists must accept the responsibility to use AI technologies as source of inspiration rather than as an independent artist itself (Lamiroy & Potier, 2022). Artists who rely on AI technologies in their artistic pursuits will inherently receive less praise society for their work, while artists who use AI as a source of inspiration to explore their creative means will be able to use the technology to form a meaningful cultural connection to their audiences (Grba, 2022). Synonymous to the relationship of responsibility formed between humans and the door, artists must understand the implications imposed by ML technology when they choose to use it for artistic creation.

Research Question and Method

It is evident that artists can use the productive capabilities of ML technologies to create artwork. However, these artists must take responsibility for the implications imposed by these technologies as well, raising concerns regarding the legal and ethical usage of ML in art creation. This leads to an important question: To what extent do the ethical and legal implications of Machine Learning technologies affect its productive use in the creation of arts?

To investigate this question, I plan on conducting interviews on artists who have used ML technologies in their work and on artists who haven't in order to examine the relationship between artists and ML technologies. Finding and interviewing artists for my study will be conducted through Reddit, a social platform where users can form communities of like-minded interest called "subreddits" (Proferes et. al., 2021), and direct contact with professors who have a thorough understanding of the art industry at the University of Virginia. By examining subreddits created specifically for art and/or machine learning (such as "r/MachineLearning") as well as professors with experience on the subject, I will be able to begin my analysis by aggregating a list of interviewee artists who have a direct understanding of ML's capabilities in art creation and have different perspectives on the matter. From this list, I will contact and virtually interview each artist through a direct communication technology such as Zoom and ask them three primary questions; How has ML influenced the creation of their artwork and that of the industry, what ethical and legal issues have they experienced as a result of the technology, and how do they believe ML can be integrated differently to be more productive in the creation of art. I will use the data collected form each interview to apply ANT to each artist's relationship and experience with ML technologies and judge them using the same methodologies utilized in Latour's network analysis of the door and its human users. This will allow me to make observations on the responsibilities both parties share with each other that arise from the ethical and legal implications behind ML use in art and how they affect society.

Conclusion

In order for long-term success of the System Safety Engineering field and the preservation of safety within the use of complex military systems, it is imperative that the automative capabilities of ML technologies be integrated into the work of Software System Engineers to improve the hazard identification process and ease the tasks of hazard analysis for Safety System Engineers. The goal of this capability is to make an adaptable, effective, and usable technology that could replace the traditional and outdated system safety accident models used within the field today.

The research outlined in this prospectus can be expected to reveal insights of how the ethical and legal implications of Machine Learning technologies affect its productive use in the creation of arts. These insights, supported by qualitiative data, will be used to draw conclusions that provide an understanding of the consequences that arise from the complex relationship between artists and ML technologies. They will also provide guidance on how this relationship may be restructured to become more productive in the future.

Resources

Bahr, N. (2015). System Safety Engineering and Risk Assessment: A Practical Approach, Second Edition. *Taylor & Francis Group, LLC*, 2, 4. https://books.google.com/books?id=Z2lYBQAAQBAJ&lpg=PP1&ots=a6ISPu8V58&dq=System%20Safety%20Engineering%20and%20Risk%20Assessment%20Bahr&lr&pg=PA4#v=onepage&q=System%20Safety%20Engineering%20and%20Risk%20Assessment%20Bahr&f=false

Bakx, G., Nyce, J. (2017, August 07). Risk and safety in large-scale socio-technological (military) systems: a literature review. *Journal of Risk Research*, 20(4), 463-481. https://www.tandfonline.com/doi/full/10.1080/13669877.2015.1071867

Bellaiche, L., Shahi, R., Turpin, M.H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., Seli, P. (2023, July 4). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, 8, 42. https://doi.org/10.1186/s41235-023-00499-6

Cetinic, E., She, J. (2022, February 16). Understanding and Creating Art with AI: Review and Outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1-22. https://dl.acm.org/doi/full/10.1145/3475799?casa_token=UZjkwp5UBoYAAAAA%3AQiqD48e5RIcPimkOFd4Se7yJ-RJloh_Aq8KM80_UIR_0EMMB8-9HjbtORMWktTYgYsuWYBVUDXI74w#d1e1539

Ding, J. (2023). Machine Failing: System Acquisition, Software Development, and Military Accidents. *jeffreyjding.github.io*, 1-6. https://jeffreyjding.github.io/documents/Machine%20Failing%20June%202023%20with%20author%20details.pdf

El Naqa, I., Murphy, M.J. (2015). What Is Machine Learning?. *Machine Learning in Radiation Oncology*, 3-11. https://doi.org/10.1007/978-3-319-18305-3_1

Grba, D. (2022, January 5). Deep Else: A Critical Framework for AI Art. *Digital*, 2(1), 1-32. https://www.mdpi.com/1435696

Hageback, N., Hedblom, D. *(2022).* AI for Arts. *CRC Press*, 1, 47-73. https://www-taylorfrancis-com.proxy1.library.virginia.edu/books/mono/10.1201/9781003195009/ai-arts-niklas-hageback-daniel-hedblom

Harvey, C., Stanton, N. (2014, December). Safety in System-of-Systems: Ten key challenges. *Safety Science*, 70, 358-366. https://www.sciencedirect.com/science/article/abs/pii/S0925753514001684

Lamiroy, B., Potier, E. (2022). Lamuse: Leveraging Artificial Intelligence for Sparking Inspiration. *Artificial Intelligence in Music, Sound, Art and Design*, 11(1), 148-162. https://search.lib.virginia.edu/sources/uva_library/items/u10121292

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. *Shaping Technology / Building Society: Studies in SocioTechnical Change*, 19(3), 225-*258.*

Leveson, N. (1991, February). Software Safety in Embedded Computer Systems. *Communications of the ACM*, 34(2), 34-46. https://dl.acm.org/doi/pdf/10.1145/102792.102799

Leveson, N. (2002, June). System Safety Engineering: Back To The Future. *Massachusetts Institute of Technology*, 3. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b2107d4823fa8b3eb83ecc8db006e8aecfe2994a

McCormack, J., Gambardella, C., Rajcic, N., Krol, S., Llano, M., Yang, M. (2023). Is Writing Prompts Really Making Art?. *Articial Intelligence in Music, Sound, Art and Design*, 12(1), 196-212. https://search.lib.virginia.edu/sources/uva_library/items/u10464210

Myllyaho, L., Raatikainen, M., Männistö, T., Nurminen, J.K., Mikkonen, T. (2022 January). On misbehaviour and fault tolerance in machine learning systems. *Journal of Systems and Software*, 183, 1. https://www.sciencedirect.com/science/article/pii/S016412122100193X

Anjila, F. (1984). Artificial Intelligence. *Learning Outcomes of Class Research*, 65.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., Zimmer, M. (2021, May 26). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society*, 7(2). https://doi.org/10.1177/20563051211019004

Roland, H., Moriarty, B. (1990). System Safety Engineering and Management. *John Wiley & Sons, Inc.*, 2, 10.

Shneiderman, B. (2020, March 23). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504. https://www.tandfonline.com/doi/full/10.1080/10447318.2020.1741118

Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus*, 109(1), 121-136.