MISSOURI S&T

S&T's Peer to Peer

Volume 1 | Issue 2 Article 4

May 2017

Exploring Potential Flaws and Dangers Involving Machine Learning Technology

David Nicholas Skoff

David Nicholas Skoff
Computer Science at Missouri University of Science and Technology

EXPLORING POTENTIAL FLAWS AND DANGERS INVOLVING MACHINE LEARNING TECHNOLOGY

Abstract

This paper seeks to explore the ways in which machine learning and AI may influence the world in the future and the potential for the technology to be misused or exploited. In 1959 Arthur Samuel defined machine learning as "the field of study that gives computers the ability to learn without being explicitly programmed" (Munoz). This paper will also seek to find out if there is merit to the current worry that robots will take over some jobs based in cognitive abilities. In the past, a human was required to perform these jobs, but with the rise of more complex automation a person may not be necessary. Many of the sources cited throughout this paper focus on the innovation of machine learning and AI and how dangerous the over automation of the world could be. Machine learning and the resulting AI's have their place in the world and more than likely they will do nothing but push the world towards a more fruitful future. Looking at potential risks of letting lines of code make important decisions is crucial given the consequences that negligence can have. There is a need to explore these topics because losing the human element in decision making can have some big implications if the AI is not programmed correctly. Machine learning has one of the greatest opportunities to impact the world. The need for caution however cannot be understated because of the potential dangers it may pose to jobs, security, and the overall stability of an ever changing world.

Exploring Potential Flaws and Dangers Involving Machine Learning Technology

Humans are always looking to evolve and automate tasks. Programming has come a long way since the early programming languages of FORTRAN and the like. Programming is now a complex task which creates complex solutions to problems plaguing all aspects of humanity. One of the complex solutions is artificial intelligence or AI. Machine learning and AI have created the potential for complete automation at home and in the workplace. There are of course problems with removing a human element from complex tasks. The potential effect on the workplace cannot be understated. Complete automation may even lead to more pressing issues. While the possibility of rogue AI seems straight from a science fiction film, the dangers of full automation are extensive. This danger could come from someone intentionally creating malicious AI or from a simple and innocent error in algorithm construction. In the future, there may need to be certain restrictions and sanctions targeting algorithms that could be used to create powerful AI's that could impact more than just the workplace. As the world nears complete automation in some sectors, security becomes paramount in ensuring safe execution of tasks. Machine learning can be a great tool for shaping the future, but its potential perils cannot be understated.

Workplace Impact

AI taking over the workplace removes the human element from decision making and introduces the potential for malicious attacks upon critical systems. Carl Frey and Michael Osborne explored the fact that jobs that usually require high cognitive ability are being replaced by an automated solution. They say, "Text and data mining has improved the quality of legal research as constant access to market information has improved the efficiency of managerial decision-making" (Frey & Osborne, 2017). This means that in the near future, tasks believed to require a human may become automated. Frey and Osborne specifically mentions such tasks as legal

writing and truck driving may be taken over by computerization (Frey & Osborne 2017). Darrell West from The Center of Technology Innovation at Brookings says, "Telemarketers, title examiners, hand sewers, mathematical technicians, insurance underwriters, watch repairers, cargo agents, tax preparers, photographic process workers, new accounts clerks, library technicians, and data-entry specialists have a 99 percent chance of having their jobs computerized" (West, 2015). This does not necessarily mean that more complicated jobs such as those in the medical and legal fields can be computerized. In fact, West says that these jobs have a less than one percent chance of being replaced (West, 2015). If phased out by robots then the workforce potentially gains efficiency and accuracy but loses the human element.

Another concerning factor is the potential breach of algorithms that dictate AI for critical systems. In these situations, a real person would be unaffected by such malicious attacks on critical systems. These types of attacks may become more probable as time goes on. Researchers from Stanford and Georgetown dissected the fact that making viruses has never been easier. They state, "To complicate matters, writing malicious programs has become easier: There are virus kits freely available on the Internet. Individuals who write viruses have become more sophisticated, often using mechanisms to change or obfuscate their code to produce so-called *polymorphic viruses*" (Kolter & Maloof, 2006). Surely, the security on critical systems which house essential AI would be strong. This however, has never stopped determined hackers from trying to crack through every firewall and security protocol. The computerization of certain jobs is coming and being prepared for such a future would be beneficial for the whole world.

Security Concerns

The potential dangers and pitfalls of machine learning are vast, and include potential attacks on algorithms themselves that are deliberate and destructive. In a paper from scholars at the

University of California, Berkeley, researchers explore the potential dangers of machine learning and its uses in modern technology. They say, “Use of machine learning opens the possibility of an adversary who maliciously ‘mis-trains’ a learning system in an IDS” (Barreno, Nelson, Sears, Joseph & Tygar, 2006). An Intrusion Detection System (IDS) will monitor network traffic and identify potential threats. The authors identify some real dangers of unchecked machine learning algorithms, and this shows why the technical community cannot sit idly by and let potentially dangerous technology run rampant. This fear of a potentially dangerous AI may seem farfetched and even impossible given the current level of technological advancement. This however, may be more feasible than once expected. Two researchers from the University of Louisville explored a way to intentionally create a malevolent AI. They say, “Just like computer viruses and other malware is intentionally produced today, in the future we will see premeditated production of hazardous and unfriendly intelligent systems” (Pistono & Yampolskiy, 2016). This leads to a world where the average person can manufacture these unsanctioned malevolent AI’s that eventually compromise vital systems and databases. Even if someone were to make a secure machine learning algorithm that cannot be exploited, there could easily be someone deliberately making an algorithm that nullifies the good actions of ethical programmers. The ethical issues associated with compromising machine learning algorithms and making malevolent AI’s from scratch are monstrous. Both sets of authors seem convinced that the possibility of malevolent AIs are worth some level of concern. Security should always hold a top priority, but the potential creation of an AI which has no inherent conscience could be very destructive. Being aware that there could be ways to specifically target algorithms dictating AI’s becomes essential in preparing certain systems for the future.

Malicious AI

Malicious AI has been a topic of science fiction since the inception of the concept of an autonomous artificial intelligence. Artificial intelligence could eventually completely overtake some areas of labor as the need of a human becomes negligible. The potential for an AI to become malevolent is not an immediate problem, but with machines taking over there is a real potential for this sort of thing to happen. A researcher from Louisville University says, "Just because developers might succeed in creating a safe AI, it doesn't mean that it will not become unsafe at some later point. In other words, a perfectly friendly AI could be switched to the 'dark side' during the post-deployment stage" (Yampolskiy, 2015). The potential for malevolence is a concern to a professional environment with large amounts of sensitive data. A single AI which was deemed safe may suddenly "go rogue". This prompts the question if these AI can be regulated and what implications such regulations would bring.

Researchers from Stanford explored the possibility of certain regulations on algorithms. They state, "On the one hand, overlooking ethical issues may prompt negative impact and social rejection,' but 'on the other hand, overemphasizing the protection of individual rights in the wrong contexts may lead to regulations that are too rigid, and this in turn can cripple the chances to harness the social value of data science" (Floridi & Taddeo, 2016). This shows that there may need to be certain restrictions, but individual rights are jeopardized by the implementation of these regulations. One individual however can throw things into chaos. Stephen DeAngelis, CEO of Enterra Solutions, says, "It only takes one 'evil genius' to undo the best laid plans that more ethical scientists put into place" (DeAngelis, 2014). There will always be that one evil genius. There are currently no ways to truly regulate algorithms created by people with malicious intent.

These issues will always be around and the emerging field of artificial intelligence may need heavy regulation in an ever changing world.

Conclusion/Discussion

All of these factors combined lead to a world where AI could truly take over. Robots are taking over jobs soon, people can make intentionally malicious AI's and regulations for algorithm construction are almost nonexistent. The potential impact for this study could bring about greater awareness to the ordinary computer scientist of the dangers of AI if not handled properly. For individuals outside of the disciple, it will also lead to greater awareness that the world is changing. Change is not necessarily a bad thing, but understanding the change and its purpose is important to live in an ever changing world. People may have to be prepared to be replaced by robots and accept that while the robots may be able to do their jobs more efficiently does not mean that they are necessarily better workers. Looking at the potential for infiltration of algorithms is something that needs to be further researched and studied. The potential positive impacts advanced artificial intelligence can have on the future may outweigh the negatives. This however does not mean that the potential dangers can be ignored because the the good may seem more important the potential bad. Awareness is half the battle and knowing how to combat problems before they flare up is an important skill. Being aware that new technology can have alarming consequences is important because oversight of such consequences could lead to disaster. Overall, there need to be regulations within the field to temper the impacts of malicious individuals. Professionals and the common person alike need to be aware of the quickly shifting technological climate and the potential impact that the shift may have in the workplace and overall security.

References

Barreno, M., Nelson, B., Sears, R., Joseph, A. D., & Tygar, J. D. (2006). Can machine learning be secure? *Proceedings of the 2006 ACM Symposium on Information, computer and communications security - ASIACCS '06*. doi:10.1145/1128817.1128824

DeAngelis, S. F. (2014). Machine Learning: Bane or Blessing for Mankind? Retrieved from https://www.wired.com/insights/2014/06/machine-learning-bane-blessing-mankind/

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254-280. doi:10.1016/j.techfore.2016.08.019

Floridi L, Taddeo M. (2016) What is data ethics? *Phil. Trans. R. Soc.* A 374: 20160360.

Kolter, J. Z., & Maloof, M. A. (2006). Learning to Detect Malicious Executables. Advanced *Information and Knowledge Processing Machine Learning and Data Mining for Computer Security*, 47-63. doi:10.1007/1-84628-253-5_4

Munoz, A. (n.d.). Machine Learning and Optimization. Retrieved from https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf.

Pistono, F., & Yampolskiy, R. V. (2016). Unethical Research: How to Create a Malevolent Artificial Intelligence. Retrieved from https://arxiv.org/abs/1605.02817.

West, D. M. (2015). What happens if robots take the jobs? The impact of emerging technologies on employment and public policy. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/robotwork.pdf.

Yampolskiy, R. V. (2015). Taxonomy of Pathways to Dangerous AI. Retrieved from https://arxiv.org/abs/1511.03246.