Combining Machine Learning and Data Privacy into an All-Encompassing Computer Science Class

(Technical Paper)

Sociotechnical Implications of ML/AI Usage in Surveillance and Censorship

(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

Prithvi Romil Kinariwala

Fall, 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 _____

Prithvi Kinariwala

Approved _____ Date _____

Capstone/Technical Advisor Daniel Graham, Department of Computer Science

Approved _____ Date _____

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction (329 Words):

George Orwell's *1984* describes the invasive "telescreen"—a television with surveillance capabilities that could "spy upon you night and day." However, the one hope humans had over telescreens was that the latter "never mastered the secret of finding out what another human being was thinking." Concerningly, the modern advent of machine learning (ML) and artificial intelligence (AI) aims to give surveillance technology the ability to think like humans. Janiesch et. al. (2021) defines ML as "the capacity of systems to learn from problem-specific training data to automate the process of analytical model building" (p. 1). While traditional algorithms take in a set of instructions and data and output a distinct result, ML algorithms take in data and the result and return the algorithm to determine that solution. AI is widely referred to as ML applications where machines are taught to perform with "human-like" intelligence. In their infancy, both ML and AI applications remained pertinent to only academic research and then grew.

However, as governments began to adopt the technology for surveillance and censorship practices, certain nations have used the robust tool to suppress freedom of expression. Feldstein (2019) claims the most notable of these nations, China and the United States, are proliferating the world with AI-based surveillance systems. However, Feldstein also observes that AI-based surveillance systems are "exploited for mass surveillance purposes" (p. 2) by autocratic and semi-autocratic nations. Therefore, the proliferation of ML and AI-based surveillance systems has become a race, akin to the Space Race at the height of the Cold War. Most critics claim that citizens have the most to lose with government overreach with censorship and surveillance. China's Xinjiang Region is the culmination of surveillance gone to its extreme maximum. Leibold (2019) cites that the Chinese Communist Party "systematically collects information on

its citizens" (p. 47) to maintain control over a region with significant unrest. Not only can governments misuse this technology, but companies also give citizens reason to distrust their use of ML and AI.

This emerging socio-technical problem can be addressed by two main branches. First being the technical aspects of ML/AI in surveillance; which is addressed with a new undergraduate class combining CS 4774 and CS 4501. The social topic is the analysis of the deterioration of the freedom of information with government's using ML/AI in surveillance upon its people.

Technical Topic—Merging CS 4774 & CS 4501 (791 Words):

Computer science students at UVA are required to take five computer science electives. Two options: Machine Learning (CS 4774) and Privacy in the Digital Age (CS 4501) are classes that investigate machine learning foundations and digital surveillance/censorship respectively. Although the two classes exist independently, merging these classes would create an environment that informs students of the negative uses of ML/AI in surveillance. The importance of placing both classes in context can be seen with the social implications of ML/AI technology.

Amongst citizens trying to avoid government surveillance and censorship, citizens have begun to use new and emergent technologies to avoid detection. CS 4501 teaches students of one of these technologies Tor (The Onion Router), a network utility that allows users to use numerous encrypted routers to obfuscate both themselves and their intended server. Relays, depicted in Figure 1, are routers that provide protection and anonymity for both the client and server using Tor for communications. The Figure depicts specific encryption between some relays in order to preserve the anonymity of Tor users. Aminuddin et. al (2018) assert that with

the advent of “machine learning classification technique similar to the classification of encrypted traffic on the surface web” (p. 113), many fear that governments can employ machine learning to crackdown and prosecute users of Tor.

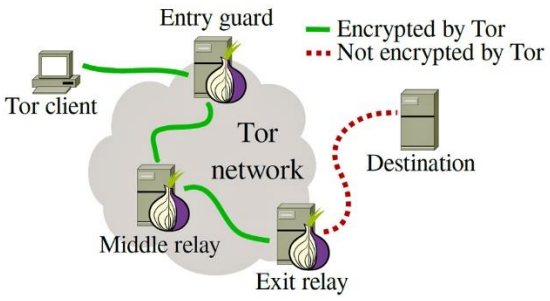

Figure 1. Overview of a Tor network with relays.
Figure shows Tor Clients using encrypted networking to maintain anonymity. (Goodin, 2014, p. 1)

The other component of ML/AI-based surveillance is the surveillance and censorship itself. CS 4501 dives deep into methods of censorship. Examples include DNS blocking, IP address blocking, and also simply just cutting internet access. All of these methods have side effects. However, nations, namely China with the Great Firewall, continue to assert almost-total control over their citizens’ computer usage.

CS 4501 also considers non-ML-based surveillance techniques. Noting that ML is not the only technical issue that pushes government surveillance to its max. Nations are still pursuing non-machine learning solutions. Petit (2019) cites that an “everywhere war” (p. 30) exists where governments take extraordinary measures to push surveillance upon their citizens. With this surveillance, Petit claims that “accountability is on the decline” (p.50) Therefore, a solution to the rise of ML and AI in surveillance shouldn’t just address ML/AI, but also non-ML solutions.

This trend hasn’t gone unnoticed. In fact, in response to ML and AI proliferation in surveillance networks, many nations have opted to use AI for reasonably good purposes. Some

nations choose to use it to rightfully better national security. The deliverable—a new class that merges CS 4774 and CS 4501— would also outline responsible AI and ML use by nations. As Cho et. al. (2020) note, one example of which is South Korea's development of "artificial intelligence technology to process military intelligence tasks more quickly and accurately" (p. 1). A departure from using the technology on its citizens, the nation follows all government oversight requirements and uses the technology only against proven enemies of the state.

There is a significant shift in how ML and AI are used for surveillance purposes. Certain nations, like the United States and China, have started another space race on new technology. Other nations, like South Korea, have (publicly) responsibly used emerging technologies. Although very little can be done to prevent a nation from utilizing its defense funds for malicious technologies, more forgiving technologies do exist for nations that wish to increase accountability to the public. Referring to the Convolutional Neural Net (CNN) explained by Preece (2018), nations can employ—for at least domestic purposes—ML algorithms that don't take personal information (p. 40).

Saha (2018) explains CNNs are a variation of Artificial Neural Nets (ANN) that take in an input image, assign importance to certain aspects of the image, and then use those weights to classify the image. Students in CS 4774 are tasked with creating a CNN to classify images of building on Grounds at UVA. Figure 2 depicts the execution process of CNNs. Input images are broken down into their RGB (Red, Green, Blue) components and then broken down into the convolutional layer with filters. Next, the data is pooled and classified using an activation function.

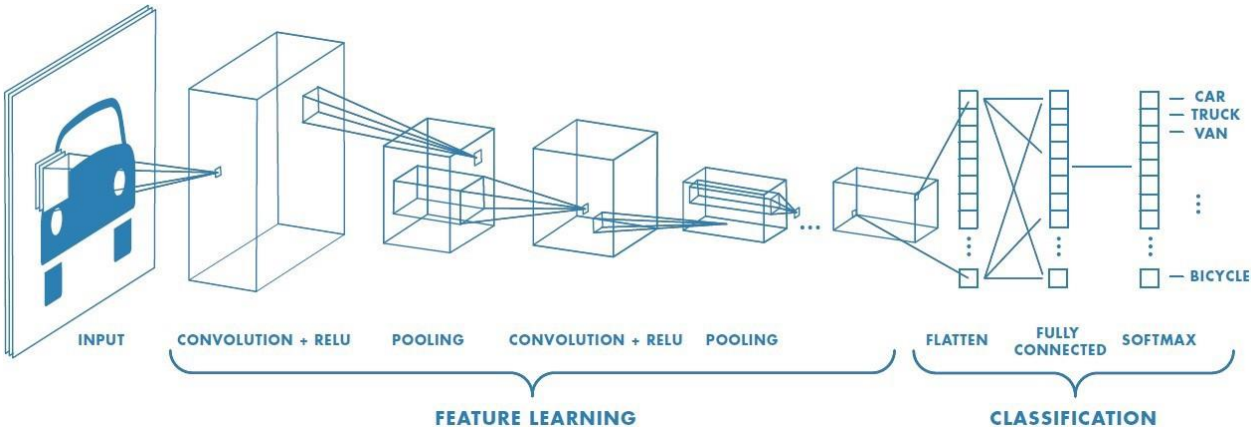

Figure 2: CNN phases diagram.
Steps from Input to Classification are shown in sequence. (Saha, 2018, p. 1)

Numerous strategies have arisen over time to combat the misuse of such powerful technologies. Most successful, however, is the use of fair use agreements of ML and AI technologies. Preece et. al., (2018) describe one example of modified ML use by governments as the use of "congestion classifiers" (p. 42), where two CNNs are used to note if a street is congested. Although this may seem like an ordinary ML solution, congestion classifiers are championed by the civilian community as they do not collect information of the people on the street, instead just the fact that humans are populating it.

The new merged class would encompass both technological and social implications of the two classes. Although there is a unit in CS 4501 that encompasses ML in surveillance, this would give the full context to the given units.

STS Issue—Global Citizens' struggle for Free Information (681 Words):

Actions on the part of citizens complicate the socio-technical balance between nations and their citizens. After the spread of ML/AI-based surveillance across the world, citizens recently have begun to find both legal and illegal alternatives to surveillance from overreaching

governments. For legal alternatives, Hossain, M. A., & Rahman, S. M. (2013) claim certain individuals can employ "adopted cryptographic approach to also hide privacy-sensitive ROIs" (p. 280) but this type of technology is not accessible by all. Wealth-disproportionate nations have some companies that can employ this technology, while others can't. This level of inequity further propagates the greater issue—the constant struggle for the freedom of information. This struggle is a constant cat-and-mouse game between prying governments using ML/AI technologies and the citizens trying to circumvent said prying.

Aminuddin et al. (2018) cover the use of Tor, and how it was initially the cutting-edge method around national surveillance. The article claims "machine learning technique for encrypted traffic classification should be considered as the prominent approaches on identifying this Tor traffic" (p.118). With time, newer technologies push the envelope with anti-surveillance strategies. This struggle has both human and non-human actors that when combined, create a complicated socio-technical problem. The sociotechnical is best analyzed when broken down into its social, technical, and organizational components.

However, there is significant evidence that it is not only the technological ramifications that drive negative change for citizens. Rising nationalism, competition between nations, and distrust amongst the global community are pushing nations to put physical censorship between open access to information and the people. Jonas (2019) argues that due to "rising nationalism worldwide, governments, civil society groups, transnational companies, and web users complain of increasing regional fragmentation online" (p. 1) This negative change is a large step in the wrong direction for the freedom of information.

Social Implications

As the freedom of information and expression varies nation-to-nation, the common

national attitude towards such freedoms also differs nation-to-nation. Certain nations' populations have great advocates championing full and fair access to information in the face of censorship. Such nations also have highly technical networks actively combatting ML/AI and non-ML/AI surveillance from nations.

However, other nations (China being the greatest example) instead instill a culture of strict obedience onto their citizens. Taking the case study of China, citizens are forced to adhere to the government's wishes. Descendants often are reprimanded or face deadly punishments. When combined with great technical research budgets, nations like China evolve into surveillance states. Leibold (2019), while using the example of China's Xinjiang region, claims when technology is left unchecked, the nation has been able to use "quantum computing and artificial intelligence, to build predictive policing models that identify any source of instability before it emerges" (p. 52) thereby allowing nations to segregate based on pure computing power. Most significant, however, is that the popular culture of the region allows such overreach to occur unhinged.

Technical Implications:

Although covered in the earlier sections of the paper, the technologies used by all parties transcend all levels of technical experience. ML technologies that use CNN can be used both invasively and non-invasively. AI technologies have only just recently been developed, but their use has been heavily utilized by companies (like Facebook) and nations alike. In response to ML/AI-based overreach, constituents have adopted technologies like Tor, VPN (Virtual Private Networks), and public DNS servers to sidestep censorship and surveillance measures.

Organizational Implications:

Although unbeknownst to the common public, organizations exist on both sides of the struggle. Nations and companies utilizing ML and AI solutions are countered by both grassroots

and decentralized organizations. As nations and companies begin to upgrade surveillance methods, civilian organizations employing technologies like Tor attempt to express their displeasure alongside like-minded people.

Previous Research:

Previous work in the sociotechnical problem has yielded mixed solutions. Most notable, however, is social policy research conducted in the realm of surveillance and the police-state. McCahill, M. (2007) cites "disproportionate targeting and exclusion" (p. 14) due to technologies having inbuilt sensitives for marginalized people. This research lays the framework for understanding how government innovation in surveillance can impact certain groups of people. McCahil further argues for the need for solutions protecting such marginalized people.

Conclusion (142 Words):

To address the dominance of ML and AI technologies in both surveillance and censorship, a proposed solution (and technical deliverable) is a class that educates computer science students of said implications. As stated with numerous case studies, there is great significance that can only be taught in a contextualized merged class that combines technical ML material from CS 4774 and technical digital privacy material from CS 4501. As for the societal aspect, the use of ML/AI in surveillance and censorship has brought significant limitations to the freedom of speech and information. Petit (2019), in his definition of "everywhere war" (p. 30) goes on to claim that it can be prevented with the presentation and demonstration by individuals against widespread high-tech surveillance. The trajectory of human liberties can be significantly changed with both the technical education and social awareness of ML/AI in surveillance.

Works Cited:

Aminuddin, M. A., Fitri, Z., Kaur, M., & Singh, D. (2018). A survey on Tor encrypted traffic monitoring. *International Journal of Advanced Computer Science and Applications*, *9*(8). https://doi.org/10.14569/ijacsa.2018.090815

Cho, S., Shin, W., Kim, N., Jeong, J., & In, H. P. (2020). Priority determination to apply artificial intelligence technology in military intelligence areas. *Electronics*, *9*(12), 2187. https://doi.org/10.3390/electronics9122187

Goodin, D. (2014, January 21). *Scientists detect "spoiled onions" trying to sabotage Tor Privacy Network*. Ars Technica. Retrieved November 1, 2021, from https://arstechnica.com/information-technology/2014/01/scientists-detect-spoiled-onions-trying-to-sabotage-tor-privacy-network/.

Feldstein, S. (2019). *The Global Expansion of AI Surveillance*. https://doi.org/https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf

Hossain, M. A., & Rahman, S. M. (2013). Towards privacy preserving multimedia surveillance system: A secure privacy vault design. *2013 International Symposium on Biometrics and Security Technologies*. https://doi.org/10.1109/isbast.2013.49

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*. https://doi.org/10.1007/s12525-021-00475-2

Jonas, A., & Burrell, J. (2019). Friction, snake oil, and weird countries: Cybersecurity systems could deepen global inequality through regional blocking. *Big Data & Society*, *6*(1), 205395171983523. https://doi.org/10.1177/2053951719835238

Joshi, A., Jagdale, N., Gandhi, R., & Chaudhari, S. (2019). Smart surveillance system for detection of suspicious behaviour using machine learning. *Advances in Intelligent Systems and Computing*, 239–248. https://doi.org/10.1007/978-3-030-30465-2_27

Larrondo, M. E., & Grandi, N. M. (2021). Inteligencia artificial, Algoritmos y Libertad de Expresión. *Universitas*, (34), 177–194. https://doi.org/10.17163/uni.n34.2021.08

Leibold, J. (2019). Surveillance in China's Xinjiang region: Ethnic sorting, coercion, and inducement. *Journal of Contemporary China*, *29*(121), 46–60. https://doi.org/10.1080/10670564.2019.1621529

McCahill, M. (2007). US and them – the social impact of 'new surveillance' technologies. *Criminal Justice Matters*, *68*(1), 14–15. https://doi.org/10.1080/09627250708553275

Petit, P. (2019). 'everywhere surveillance': Global surveillance regimes as techno-securitization. *Science as Culture*, *29*(1), 30–56. https://doi.org/10.1080/09505431.2019.1586866

Preece, A., Harborne, D., Raghavendra, R., Tomsett, R., & Braines, D. (2018). Provisioning robust and interpretable AI/ML-based service bundles. *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*. https://doi.org/10.1109/milcom.2018.8599838

Saha, S. (2018, December 17). *A comprehensive guide to Convolutional Neural Networks - the eli5 way*. Medium. Retrieved November 1, 2021, from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.