Examining Machine Learning's Impact on Personal Privacy

Abstract

This paper delves into the growing concerns surrounding the use of machine learning and its impact on personal privacy. It highlights the potential for misuse in surveillance technologies and proposes various strategies to counter these threats, emphasizing the need for collaboration between machine learning experts and human-computer interaction (HCI) researchers.

1 Introduction

The intersection of machine learning and privacy has become a significant area of study within the field of computer science. While privacy-preserving techniques such as differential privacy offer potential solutions, some machine learning systems, particularly those designed for biometric analysis or behavioral profiling, inherently compromise individual privacy. Therefore, there is a crucial need to explore methods beyond these traditional approaches.

Although various definitions and frameworks for privacy have been proposed, a universal consensus remains elusive. This paper focuses on specific harms to privacy caused or made worse by machine learning systems. In an era of powerful algorithms and massive datasets, maintaining privacy is increasingly challenging, given that facial recognition systems can identify individuals in public spaces, targeted advertising can exploit user profiles, and predictive policing algorithms can single out individuals for surveillance. This paper addresses these unique threats to privacy that machine learning systems enable.

This research provides an overview of strategies developed to combat privacy-threatening machine learning systems and advocates for increased collaboration between the machine learning community and experts in the field of human-computer interaction (HCI). Two main approaches are discussed: first, challenging the data that feeds these models through obfuscation or data withholding, and second, directly challenging the model itself through public pressure or regulation. This paper suggests that computer scientists have an important role to play in both these approaches.

2 Challenging Data

Machine learning systems depend on data for both training and operation. Data is used to train machine learning models, and new data is fed into the models to generate predictions. These training and deployment stages can be iterative; models can be updated using new data over time. One way to oppose a machine learning system is by disrupting the data it relies on. This involves strategies such as data obfuscation or withholding of data.

2.1 Obfuscation

One method for avoiding machine learning surveillance is by altering either the data used to make predictions or the data used to train the system. For example, research has shown that glasses can be designed to deceive facial recognition systems. This type of method uses adversarial examples, where a slight modification to a data point is enough to cause misclassification by a machine learning

.

model but is imperceptible to humans. Various strategies have been developed for evading facial recognition using adversarial examples, with the aim to help individuals avoid surveillance. However, these approaches often lack strong guarantees.

Another approach involves altering the training data used for machine learning models, known as data poisoning attacks. For example, systems can create altered images to reduce the accuracy of deep learning models. Additionally, some vendors sell clothing designed to trigger automated license plate readers by injecting junk data, furthering this method.

Beyond image classification, similar obfuscation tactics have also been used to counter web tracking and loyalty card-based tracking. Obfuscation can also have an expressive function, as illustrated by groups who use unusual makeup to challenge facial recognition. These acts serve a dual purpose of both evading surveillance and protesting against its use.

While adversarial examples and data poisoning are ongoing topics of study, these technologies need further evaluation before being adopted as anti-surveillance tools. Accessibility, evaluation methods, and communication of risks are areas that require further work and collaboration between machine learning experts, HCI researchers, activists, and other relevant stakeholders.

2.2 Withholding Data

An alternative approach to altering data is to withhold it entirely. This can be achieved through privacy-enhancing technologies that block web tracking. While tracker-blocking browser extensions can provide some privacy to individuals, data can also be withheld collectively. Data strikes, a form of digital boycott, can apply pressure to technology companies. Protest non-use is another way of withholding data, where people stop using platforms due to privacy concerns. These methods go beyond simple evasion, using the act of withholding data as a way to launch broader campaigns against surveillance systems.

3 Challenging Models

While data-oriented approaches are helpful, policy solutions may offer a more effective way to resist machine learning surveillance systems. For example, while strategies can help evade facial recognition, banning the technology would render those strategies unnecessary. There are many forms that regulation can take and many roles that computer scientists can play in this process.

One method of pressuring companies that develop surveillance technologies is through auditing. Research audits of facial recognition systems have shown they perform poorly on darker-skinned subjects, which has led to wrongful arrests. These audits have led some companies to stop selling facial recognition technology. However, audits do have limitations, as they can sometimes normalize harmful tasks for certain communities.

Some technologies are difficult to audit due to restricted access. Nevertheless, these systems can sometimes be reverse-engineered to show potential societal harms. Predictive policing systems, for instance, can amplify existing biases. Algorithmic audits or reverse engineering should focus on broader societal implications of the technology to avoid merely shifting goal posts and algorithmic reformism.

Researchers have partnered with community organizations to resist surveillance technologies, debunking the myth that critics do not understand the technology, and demystifying complex algorithms. It is important for researchers to approach these collaborations with humility, as community organizers bring their own areas of expertise.

It is also crucial to recognize the academic community's role in creating and upholding surveillance technologies. Computer science educators should make computing's role in injustice more visible. Student-led efforts can help educate future computer scientists about the consequences of their work.

4 Conclusion

This paper has outlined various methods for resisting machine learning-based surveillance technologies. It emphasizes the need for participatory methods when developing anti-surveillance technologies.

While these participatory methods are common in HCI research, the machine learning community has paid less attention to it. The impact of surveillance technologies is disproportionately borne by already marginalized groups. Therefore, it is critical that the design of anti-surveillance technologies be led by those who are most affected.