You Still See Me: How Data Protection Supports the Architecture of ML Surveillance

Rui-Jie Yew
Department of Computer Science, CNTR
Brown University
Providence, USA
rui-jie_yew@brown.edu

Lucy Qin
Department of Computer Science
Brown University
Providence, USA
lq@brown.edu

Suresh Venkatasubramanian
Department of Computer Science, CNTR
Brown University
Providence, USA
suresh@brown.edu

Abstract

Data is key to the functionality of ML systems. Data protection has therefore become a focal point of policy proposals and existing laws that are pertinent to the governance of ML systems. Privacy laws and legal scholarship have long emphasized privacy responsibilities developers have to protect individual data subjects. As a consequence, technical methods for privacy-preservation have been touted as solutions to prevent intrusions to individual data in the development of ML systems while preserving their resulting functionality. Further, privacy-preserving machine learning (PPML) has been offered up as a way to address the tension between being "seen" and "mis-seen" - to build models that can be fair, accurate, and conservative in data use.

However, a myopic focus on privacy-preserving machine learning obscures broader privacy harms facilitated by ML models. In this paper, we argue that the use of PPML techniques to "un-see" data subjects introduces privacy costs of a fundamentally different nature. Data may not be used in its raw or "personal" form, but models built from that data still make predictions, and individualize and influence you–and people like you. We illustrate this point with an example on targeted advertising and models built with private set intersection.

1 Introduction

Data is key to the functionality of ML systems. Privacy laws, legal scholarship, and recent policy proposals have emphasized privacy responsibilities developers have to their data subjects.[1] Technical methods for privacy-preservation are touted as solutions to prevent intrusions to data in the development of ML systems while preserving their resulting functionality. However, a myopic focus

[1]The White House's Blueprint for an AI Bill of Rights emphasizes data privacy [White House]; NIST's AI Risk Management Framework more broadly defines privacy as principles that support human autonomy and highlights "anonymity, confidentiality, and control" as privacy values [National Institute of Standards and Technology, 2023]; Article 10 of the EU AI Act centers on "Data and data governance", emphasizing data considerations in the development of ML systems. The Article additionally considers the use of "privacy-preserving measures" like "pseudonymisation" and "encryption". [European Union Legislature, 2021]

Workshop on Regulatable Machine Learning at the 37th Conference on Neural Information Processing Systems (RegML @ NeurIPS 2023).

on the properties of data as part of these techniques–where it is processed, whether an individual is identified through it– can belie privacy harms to subjects of computation in the development and application of resulting models.

There is a significant body of work that discuss tensions between privacy, fairness Chang and Shokri [2021], Gupta et al. [2023], Xiang [2022], and accuracy Suriyakumar et al. [2021] in the development of ML systems. In these works, the analysis is typically focused on the effects of applying privacy-preserving techniques such as differential privacy or data minimization to fairness in machine learning development. Xiang [2022], in particular, discusses a fundamental tension between being "seen" and being "mis-seen" in ML development. Being "seen" is defined as "having images of your face and/or body collected and processed for *developing* HCCV [human-centric computer vision] systems" and being "mis-seen" refers to "experiencing poor performance from a deployed HCCV system: this includes your face/body not being detected, being misrecognized as someone else, someone else being misrecognized for you, or having images/videos of you misclassified or mischaracterized".

Privacy-preserving ML introduces techniques that can be wielded so that data subjects are technically "*un*-seen" in the development of ML systems. With the application of these techniques, subjects are treated under the scope of data protection principles and privacy laws as not having been seen at all. Developers may then be relinquished of the responsibilities of ever having computed on people. We argue that the use of these techniques to "un-see" data subjects nonetheless introduces privacy costs. Privacy harms persist in how the models are applied: often with models that are developed with techniques to"un-see" you, only to then turn around to take a good look.

2 Personal privacy harms are legally conceptualized to flow from personal data

Information privacy harms are often treated synonymously with, or as flowing from, *personal* data. The European Union's (EU) AI Act heavily ties privacy considerations to data protections and access to datasets, emphasizing techniques such as anonymisation and pseodonymisation European Union Legislature [2021]. In a similar vein, the EU's General Data Protection Regulation (GDPR) provides a carve-out when anonymisation techniques are used EU Legislature [2018]. In the United States, domain-specific federal privacy laws like the Children's Online Privacy Protection Act (COPPA) and HIPAA state privacy laws like the Illinois Biometric Privacy Act (BIPA) and the California Consumer Privacy Act (CCPA) regulate data collection. There have also been a number of recent legislative proposals for privacy policies, such as the Data Care Act United States Congress [2021] and the New York Privacy Act New York State Legislature [2023]. The Data Care Act largely scopes its protections to "individual identifying data" and the New York Privacy Act to "personal data"[2].

In the legal scholarship, descriptions of wide-ranging, potentially conflicting, privacy harms share roots in individual data. Losing control of individual data or unauthorized access to individual data have traditionally been considered a privacy harm. But even wider-ranging privacy harms, like discrimination or manipulation harms, are described as flowing from the use of personal data. Angel and Calo [2023] embed this conceptualization in the question: "Or to examine, for instance, what makes them different from discrimination or manipulation harms, which also stem from the actual use of *personal data* for decision-making?" Crawford and Schultz [2014] discusses predictive privacy harms as coming from how technical advances in Big Data significantly expand uses of "*personally identifiable data*" to "predictive analysis of an individual's *personal data* without their knowledge or express consent".

3 Personal privacy harms facilitated by ML don't need personal data

When policymakers apply data governance and protection principles to ML systems, they focus on the *data inputs* to model training and examine whether the data handling processes make it identifiable or personal. That determination typically rests on three dimensions: how the data is collected, whether it's personally identifiable, and how identifiable data is processed. Privacy-preserving ML tools

[2]which is defined as "any data that identifies or could reasonably be linked, directly or indirectly, with a specific natural person, or household"

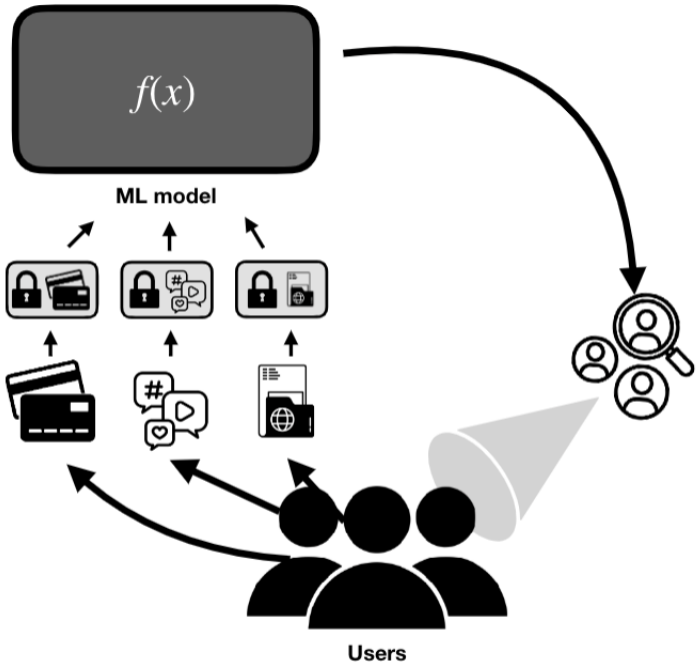

Figure 1: An illustration of secure MPC can be used to facilitate profiling. The encryption of plaintext data (on the left) enables the analysis of results from decentralized sources, which can then be leveraged to target the same individuals.

attempt to make it possible to use data for training so that all three dimensions are deemed "safe for use". [3]

As a result, the legal interpretation of privacy-preserving techniques hold weight in how ML systems are regulated and litigated. As laws for ML systems are trickling in, some of the biggest settlements we have seen for the development and deployment of ML systems have come from privacy and data protection laws, like Illinois' BIPA[Illinois State Legislature, 2008, Yew and Xiang, 2022] and the EU GDPR[EU Legislature, 2018].

When privacy harms are conceptualized to flow from the collection, processing, and use of personal data, it may then be reasonable to legally conclude that the use of these techniques then stops further privacy harms in their tracks. This is not the case. As these methods conceal your individual contribution to ML systems, they can simultaneously obsfuscate how you are still targeted as a result of those computations. We illustrate this with a case study on profiling and targeted advertising.

3.1 Case Study: Profiling and Targeted Advertising

"Profiling" and targeting individuals for advertising are within the scope of a number of recent proposals for privacy laws in the United States, such as the proposed federal American Data Protection and Privacy Act (ADPPA) [U.S. Legislature, 2022]and the New York Privacy Act (NYPA) [New York State Legislature, 2023]. The NYPA defines profiling as"any form of automated processing performed on *personal data* to evaluate, analyze, or predict personal aspects related to an *identified* or *identifiable* natural person's economic situation, health, personal preferences, interests, reliability, behavior, location, or movements." Again, there is an emphasis on identification, even so far as to potentially imply that, if the person's identifiable raw data is not processed, then that person was

[3]For example, Veale [2023] highlights that data protection law has facilitated the non-consensual shift to data processing being done locally–challenging the premise of that data having been collected at all.

not profiled. More insidiously, such a legal conception of profiling could also imply that, when that data is no longer considered *personal* or *identifiable*, people no longer have a say when their data is ultimately used to target them.

Secure multi-party computation (MPC) and other technologies that compute on encrypted forms of data can target people in ways that could not be done before without plaintext data about individuals. MPC allows the computation of some function f over encrypted inputs while producing some public output. This enables multiple parties to share encrypted (private) data and then benefit from the public result of the joint computation. In the process, all parties only learn the output of the function. This allows for the analysis of data from various decentralized sources—without *any one* source learning more about *any individual* through another source's data. Multiple parties (such as different tech platforms) could then link data about individuals across their datasets and then perform statistical computations using this linked data, as illustrated in Figure 1.

MPC could therefore enable the joint use of data from financial institutions and online platforms to support targeted advertising: companies can join datasets and link individuals using their financial data *and* internet activity *and* demographic information. Without any entity having access to any other entity's dataset, it can be jointly confirmed that a user who browses an ad on a social media platform later purchases that same product through their ad engagement data and credit card history. This information enables a more fine-grained assessment of the success of an advertisement that can then be used to train better targeting models. And, importantly, a particular user may not be identified throughout this process and this data is never shared in plaintext format. Systems that can perform these types of operations are referred to as "privacy-preserving" but the incorporation of the causal link between a clicked advertisement and a payment—whether done physically or online—enables a new and even more targeted form of individualization.

In other words, "privacy-preserving" ways of computing on encrypted data, and PPML more broadly, can allow for a rebuilding of the entire data collection/processing/training/inference pipeline in a way where no entity along the way is aware of the identity of whose information is being processed. And yet, none of this changes the reality of the surveillance infrastructure that the technology enables and, in some cases, bolsters. Calo [2011] presents a delimiting between subjective and objective privacy harms for the law–with subjective privacy harms being harms that come from a feeling of being watched and objective privacy harms being harms that come from an unexpected use of information about that person toward that person. An environment built up by PPML straddles both. It is an environment where there is not one entity using all available information against you–thus reducing the risk of an objective privacy harm, but where all of that information is nonetheless reflected in a personal way–making even more acute the subjective privacy harm.

4 Re-imagining Privacy for ML Systems

Re-imagining privacy for ML systems requires a centering of how PPML or, rather, technical processes for de-personalizing data, further reduce the level of control that users have in their interactions with ML systems. This is not to say that data access and confidentiality are not important. Rather, the rhetoric and technical infrastructures that support data access and confidentiality can obfuscate what are ultimately surveillant ends.[4]

Limiting the scope of privacy protections to sensitive or personal input data or inferences of that data ignores the personal privacy harms that are fueled by input data to models. Such a conception of privacy emphasizes what is done *to* input data rather than what is done *with* that data and the models arising from them. In their proposal for a duty of loyalty for privacy law, Richards and Hartzog [2021] provide a related argument of data opportunism as a harm that data protection fails to protect– namely, that: "data protection models can miss abuses that do not involve personal data processing, like dark patterns for nudging or the use of knowledge gleaned from aggregated data from other people to manipulate us". The value of your data no longer lies just in its ability to identify or influence you but in its ability to identify and influence others [Viljoen, 2021]. Similarly, the value of data aggregations is not just in providing community-level insight but in using those aggregations to individualize people. PPML techniques call into question whether we are observed,

[4]For exmaple, Kalluri et al. [2023] find that an overwhelming emphasis of computer vision machine learning on targeting people: "90% of papers and patents emphasize it as a strength that their technologies can target human data".

but we nonetheless see ourselves reflected through our devices. As Veale [2023] advocates, the computing infrastructure for privacy-preserving surveillance calls for a shift away from privacy as data access and confidentiality and toward privacy as device accountability to users–and we also argue, a say in how we are reflected through and influenced by our models.

References

María P Angel and Ryan Calo. Distinguishing privacy law: A critique of privacy as social taxonomy. *Available at SSRN*, 2023.

Ryan Calo. The boundaries of privacy harm. *Ind. LJ*, 86:1131, 2011.

Hongyan Chang and Reza Shokri. On the privacy risks of algorithmic fairness. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 292–303. IEEE, 2021.

Kate Crawford and Jason Schultz. Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, 55:93, 2014.

EU Legislature. General data protection regulation, 2018.

European Union Legislature. Eu ai act. `https://artificialintelligenceact.eu/the-act/`, 2021.

Arushi Gupta, Victor Y Wu, Helen Webley-Brown, Jennifer King, and Daniel E Ho. The privacy-bias tradeoff: Data minimization and racial disparity assessments in us government. 2023.

Illinois State Legislature. (740 ilcs 14/) biometric information privacy act., 2008.

Pratyusha Ria Kalluri, William Agnew, Myra Cheng, Kentrell Owens, Luca Soldaini, and Abeba Birhane. The surveillance ai pipeline, 2023.

National Institute of Standards and Technology. Nist ai risk management framework. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf, 2023.

New York State Legislature. New york privacy act, 2023.

Neil Richards and Woodrow Hartzog. A duty of loyalty for privacy law. *Wash. UL Rev.*, 99:961, 2021.

Vinith M Suriyakumar, Nicolas Papernot, Anna Goldenberg, and Marzyeh Ghassemi. Chasing your long tails: Differentially private prediction in health care settings. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 723–734, 2021.

United States Congress. S.919 - data care act of 2021. `https://www.congress.gov/bill/117th-congress/senate-bill/919`, 2021.

U.S. Legislature. H.r.8152 - american data privacy and protection act, 2022.

Michael Veale. Rights for those who unwillingly, unknowingly and unidentifiably compute! 2023.

Salome Viljoen. A relational theory of data governance. *Yale LJ*, 131:573, 2021.

White House. Blueprint for an AI Bill of Rights. `https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf`.

Alice Xiang. Being'seen'vs.'mis-seen': Tensions between privacy and fairness in computer vision. *Harvard Journal of Law & Technology*, 2022.

Rui-Jie Yew and Alice Xiang. Regulating facial processing technologies: Tensions between legal and technical considerations in the application of illinois bipa. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1017–1027, 2022.