Privacy Risk Minimization in AI/ML applications

By
Pushpak Pujari, Product Lead Manager, Verkada
-
May 9, 2022

Privacy Matters

In this age of data-first organizations, no matter what industry you're in, you're most likely collecting, processing, and analyzing tons of customer data. It could be for fulfilling a customer's service request, for legal or regulatory reasons or for providing your customers with better user experience through personalization using artificial intelligence or machine

learning. However, data breaches are increasing every year, with 1862 reported data compromises in 2021, up 68% compared to 2020, with 83% of those involving sensitive information (as per Identity Theft Resource Center). Such sensitive information falling into the wrong hands could wreak havoc to the customer's life due to identity theft, stalking, ransomware attacks etc. This coupled with the rise of privacy laws and legislations across various states has brought privacy enhancing data processing technologies to the forefront.

Privacy vs data utility tradeoff

For AI applications such as personalization, privacy and data utility can be visualized on opposites sides of the spectrum. Data that doesn't contain anything personal i.e., expose no traits or characteristics of the customers, lend no value for personalization. However, data containing personal information can be used to deliver highly personalized experience but if the dataset, ends up in the hands of any human can lead to loss of customer data privacy. As a result, there is always an inherent tradeoff between privacy risk and utility of that data.

Value of being privacy-first for organizations

Health Insurance Portability and Accountability Act (HIPAA), California Consumer Privacy Act (CCPA), Children's Online Privacy Protection Act (COPPA), Biometric Identifier Act are just a few of the many privacy-centric laws and legislations in the US. Failure to comply with such regulations can cost an organization billions of dollars in fine. For example, recently the state of Texas sued Facebook's parent company Meta for billions of dollars in damages for mishandling and exploiting sensitive biometric data of millions of people in the state. Being privacy-first can help avoid huge fines and not limited to losing the license to operate as a business. In addition, there can be massive loss to the consumer trust and loyalty, brand image and perception. Being negligent about consumer's data privacy can demolish customer lifetime value, affect conversions and renewals. In fact, companies like Apple have flipped the problem on its head and in fact are using privacy as a competitive moat as a differentiator from other technology companies.

Sources of Privacy Risk in data collected by an organization

There are three key sources of privacy risk- a) the raw customer data and any of its derivatives, b) metadata and logs, c) ML models that have been trained on customer data. Raw customer data can be customer entered data such as name, address, age sex and other profile details or data on how customer is using the product such as page visits, session duration, items in cart, purchase history, payment settings etc. Metadata and logs include location of customer, location product website was accessed from, IP address of device, MAC address, service logs, logs of call with customer support etc.

ML models itself can itself seem like they don't contain anything personal, but ML models can memorize patterns in the data it has been trained on. Models trained on critical customer data can retain customer attributable personal data within in the models and present customer personal data exposure risk regardless of whether the model was deployed in the cloud or on edge devices. If a malicious actor gains access to such a model, even as a black box, they can run series of attacks to recover the personal data leading to privacy breach.

An ML model's security classification should be determined based on the data classification of its training data. ML model artifacts can contain plaintext customer data and the ML model itself is susceptible to privacy attacks. If an organization is running a marketplace and sharing ML models with external partners, even under NDA and data sharing agreements, ML models present high risk of privacy attacks.

How to identify the gaps

Here are a few strategies to identify the biggest gaps and start using PET to close the gaps:

1. Follow the data: chart the customer data lifecycle across your organization, right from data collection or ingestion to storage, usage to deletion. A chart will help you visualize the end-to-end flow and formulate an effective strategy for the whole organization
2. Create threat map: identify which humans, processes and systems have access to customer data. Where are the humans in the loop, and is it a business requirement for them to have access? Also, what are the tools used to access data and how.
3. Identify the use cases: for each use case, think about the likelihood of a privacy-risk event and estimate the blast radius and categorize each threat as high, med, and low severity
4. Identify the drivers and define goal success criteria: start with the high-risk items first, regularly measure progress and make your way down the list till all the risk items are low.

Privacy Enhancing Technologies

Privacy Enhancing Technologies is an area of active research with tremendous advancements made in the last 5 years. Broadly PET can be classified under 2 buckets – data sanitization and privacy-preserving computation. Data sanitization techniques focus on detecting and modifying personal information to de-sensitize it. This includes techniques such as direct identifier detection and removal, Pseudonymization, K-anonymization and Differential Privacy. Privacy-preserving computation techniques focus on operating on private data but in a closed environment with no humans

having access to it. This includes techniques such as homomorphic encryption, secure multi-party computation, federated learning, and trusted execution environments (TEE). One of the most efficient ways of managing privacy risk is by doing all the processing on the edge device itself. This mitigates biggest risk of customer data leaving their device, especially for consumer applications and has the added benefit of give a hyper targeted experience in distributed environment without burning cloud compute costs. The biggest challenge is to prevent overfitting since the volume of data may not be sufficient and affect quality if predictions.

Privacy engineering and Privacy Enhancing Technologies is a rapidly evolving field and the best way to stay up to date is by attending conferences, reading research papers, and joining the open-source community to get the latest updates. Lastly, I recommend starting small with the simple techniques of identifying direct identifiers and removing them before moving to more complex approaches.

About Pushpak Pujari

Pushpak leads product management at Verkada, a high-growth Physical Security based in the SF Bay Area where he runs their Cloud Connected Security Camera product lines. He is responsible all of camera software including and not limited to security, privacy and using AI and Computer Vision to improve video and analytics capabilities. Before Verkada, Pushpak led Product Management at Amazon where he helped build an end-to-end privacy-preserving ML platform at Amazon Alexa. He also led product management at Amazon Web Services IoT where he launched a no-code platform to design and deploy IoT automation workflows on edge devices at Amazon Web Services (AWS). Pushpak has also spent 4 years at Sony in Japan building Sony's flagship mirrorless cameras. He holds an MBA from The Wharton School and bachelor's in electrical engineering from IIT Delhi, India. Pushpak is a featured speaker at Privacy Enhancing Technology Summit at Hilton from 18-19 May 2022.