Hashed & Salted | A Privacy and Data Security Update
March 2022

Pivot to PETs: What You Need to Know about Privacy Enhancing Technologies

With increased privacy regulation and regulatory scrutiny, global constraints on cross-border data transfers, and consumer sentiment and enhancements in vendor management and data governance practices, the ability to (lawfully) acquire user-level data has never been more challenging. The risks of doing so have never been higher. This changing landscape has pushed privacy-enhancing technologies (PETs), which have existed in the wings for several years, onto the main stage. Most companies will need to pivot to using a suite of PETs and working with partners and solutions that are powered by PETs. While some of these solutions are very technical, leaders across a variety of functions will need to understand the basics in order to advise their teams.

What are "Privacy Enhancing Technologies"?

PETs enable companies to embed privacy-by-design principles into their data governance practices. By applying PETs, companies can minimize the amount of personal data they collect, use and share, while maximizing data security. Here is an overview of five of the top PETs in use today.

Homomorphic Encryption

Homomorphic encryption makes it possible to analyze or manipulate encrypted data without revealing the data to anyone. The challenge with traditional encryption is that you must decrypt the data to work with it. Once decrypted, it becomes subject to the same privacy challenges as raw personal data. While encryption can be a good tool for security, it may not be as helpful for privacy. Homomorphic encryption addresses this challenge by introducing an algebraic system that allows functions to be performed on the data while it is encrypted. Data is encrypted with a public key, and only the individual with the matching private key can access the unencrypted data after the processing is completed. Homomorphic encryption allows data to be secure and private even when someone else uses it.

Across industries, including adtech, fintech, life sciences and health care, the ability to share data between parties to generate analytics and insights is critical. Homomorphic encryption is used in real-world solutions like clean rooms. Clean rooms are a location where advertisers and publishers can aggregate customer data from different platforms and combine it with first-party advertiser data for measurement and attribution, but which have strict privacy controls that prevent access to consumers' personally identifying information.

Multiparty Computation

Multiparty computation (MPC) allows two or more organizations to work together, while limiting the information that either party can learn. All inputs remain private, and the data is encrypted while in transit, in storage and in use, ensuring neither party can see the other's data. With MPC, the data doesn't need to be

Attorney Advertising

transferred to a central location — it can be processed locally. Processing data locally or on a device is one method that can be used to minimize the amount of data that gets shared between parties.

MPC can be used to report an ad campaign's results or train a machine learning model where two or more parties hold the data. For example, suppose one party has information about who saw an ad, and the other party has information on who makes a purchase. In that case, MPC makes it possible for both parties to combine their data and identify how an ad performs without sharing user-level information data.

In June 2021, the European Data Protection Board recognized multiparty computation as a supplementary technical measure for international personal data transfers from the EU to countries without an adequacy decision.

Differential Privacy

Unlike homomorphic encryption or MPC, which are processes applied to personal data, differential privacy describes the state of a set of personal data. Differential privacy is a mathematical definition of privacy in which the data itself is not considered identifiable. Differential privacy makes data anonymous by injecting noise into a dataset in a way that allows that data to be analyzed without identifying any personal information. Think about a photograph that is increasingly blurred to the point where you can still tell what's happening in the picture, but you can't identify any singular person in the image. If you need to know only what the picture is about, not who is in it, differential privacy provides a path for analysis that should not involve identifiable data. It doesn't force companies to make a trade-off between privacy and utility.

Federated Learning

Federated learning is decentralized machine learning. One of the critical challenges of machine learning is the need for large amounts of data, which creates privacy and security risks. Federated learning trains artificial intelligence models with data accessed on the user's device. In other words, it makes it possible to take advantage of machine learning while minimizing the need to aggregate user-level data in the cloud. For example, Google uses federated learning to improve on-device machine learning models like "Hey Google" in Google Assistant, which learns from users' voice commands.

Synthetic Data

Synthetic data is artificially manufactured rather than generated by real-world events. It is created algorithmically from an existing set of data. The dataset resembles but is not a replica of the underlying data and can be used as a stand-in for testing. Synthetic data provides a privacy-protective dataset that can be used to train machine learning models, which require access to large volumes of data.

Synthetic data is generally considered non-identifiable. While the underlying dataset is personal data subject to applicable privacy laws, the synthetic dataset created does not relate to any individual and should fall outside the definition of personal data under most laws.

In the health care sector, synthetic data enables health care professionals to allow public use of record-level data, while maintaining patient confidentiality. The CNIL (the Data Protection Authority in France) has reviewed and approved a synthetic data generation tool as an acceptable form of anonymization of health data. Likewise, synthetic datasets can replicate the transaction data that many fraud-detection companies rely on to train their models in the financial industry.

What Questions Should You Be Asking?

Here are eight questions to ask your teams about projects involving PETs. If your teams are not coming to you to explore PETs, consider talking to them about which of these technologies can be used to help them further their goals while managing their privacy risks.

- What problem are you trying to solve? What problem does this technology solve?
- What specific data, if any, will be collected and used?
- Where is the data stored (on the device, in the cloud, on the premises, with a third party)?
- What are the risks of re-identification?
- Can we use multiple PETs to address the risks presented by using one PET (e.g., can we use differential privacy in combination with federated learning)?

- Are we sharing any data/algorithms with third parties? If so, what will they have access to?
- How will we test and monitor the efficacy of the PETs?
- What contractual protections is the provider offering?

These questions will help you vet the technology, and they can be used to determine whether to classify the data involved as pseudonymous, de-identified, anonymous or still personal data.

Related Professional

Jessica B. Lee . . . . jblee@loeb.com

This is a publication of Loeb & Loeb and is intended to provide information on recent legal developments. This publication does not create or continue an attorney client relationship nor should it be construed as legal advice or an opinion on specific situations.

6899 02 REV1 03-16-2022