Privacy Enhancing Technologies for AI

Working Draft

Authors:

- Login to your Google account to access full editing permission.
- Change from Editing to Suggesting in the upper right of the Google doc for tracking each author's edits.

Please contact research-support@cloudsecurityalliance.org to request full access to author this document.

Reviewers/Visitors:

- If you have a Google Account, please login before commenting. Otherwise, please note your name and affiliation in the comment you leave.
- Use the Comments or Suggesting features on Google docs to leave your feedback on the document. Suggestions will be written in and identified by your Google Account. To use the comments feature, highlight the phrase you would like to comment on, right click and select "Comment" (or Ctrl+Alt+M). Or, highlight the phrase, select "Insert" from the top menu, and select "Comment." All suggestions and comments will be reviewed by the editing committee.

For more information about Google's Comments feature, please refer to http://support.google.com/docs/bin/answer.py?hl=en&answer=1216772&ctx=cb&src=cb&cbid=-rx63b0fx4x0v&cbrank=1

The permanent and official location for the [Insert WG Name] Working Group is https://cloudsecurityalliance.org/research/working-groups/working-group-name

Acknowledgments

Lead Authors

Add Names

Contributors

Add Names

**This needs to be in agreement with the list of authors contributing to each section*

Reviewers

Add Names

Co-Chairs

If distinct from Lead Author

CSA Global Staff

Add Names

Table of Contents

Section assigned per person – For internal use only (to be deleted from the public peer review version of the document)

Author’s name/email	Section
Marina Bregkou (mbregkou@cloudsecurityalliance.org)	Heading 1
Josh Buker (jbuker@cloudsecurityalliance.org)	Heading 2
Author X	Heading 3

Table1: Example of assignments table for internal use

This is a Heading 1 – author: Please add name of author.

This is a Heading 2 – author: Please add name of section author

This is a Heading 3 – author: Please add name of section author

This is a Heading 4 – author: Please add name of author

This is a Heading 5 – author: Please add name of author of section

[author if author of paragraph is different from the section’s author] This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text.

[author if author of paragraph is different from the section’s author] This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text. This is an example of paragraph text.[1]

Table Header	Table Header	Table Header

[1] This is an example of a footnote

Table Text	This is an example of table text.	This is an example of table text. This is an example of table text.
Table Text	This is an example of table text.	This is an example of table text. This is an example of table text.
Table Text	This is an example of table text.	This is an example of table text. This is an example of table text.

Privacy Enhancing Technologies for AI

(Rocco Alfonzetti)

Privacy Protection Techniques:

Data Anonymization and Pseudonymization:

These techniques modify data to remove personally identifiable information (PII) while still allowing for analysis. Anonymization removes all identifiers, while pseudonymization replaces them with fictitious values.

Both Privacy solutions require robust data discovery and classification in order to identify sensitive data fields within your dataset.

Configuration uses techniques like redaction (complete removal), generalization (replacing specific values with broader categories), k-anonymity (ensuring each record is indistinguishable from at least k-1 others), and noise injection.

Most major DBMS solutions like Oracle, SQL Server, and MySQL offer built-in pseudonymization functionalities. These allow creating and managing pseudonymization keys for data transformation within the database environment.

Cloud providers like AWS, Azure, and GCP offer data anonymization and pseudonymization functionalities within their cloud storage solutions..

Data Masking:

Data masking is a technique used to modify sensitive data within a dataset to protect privacy while still enabling data analysis and other purposes. It essentially creates a fake version of your real data that retains its structure and key characteristics, but hides the actual values. AI data masking solutions are a critical tool for protecting sensitive information while enabling AI development and training. Some of the available Data Masking solutions are:

- IBM InfoSphere Data Masking
- Informatica PowerCenter Data Masking
- Precise DataFlex
- Broadcom Test Data Manager
- Denodo Data Masking (Open-source option)

Advanced Encryption for AI Privacy:

Homomorphic Encryption:

This allows computations to be performed on encrypted data without decryption. This ensures data privacy even while it's being used by AI models. However, it can be computationally expensive for complex models.

- **OpenFHE:** This is a popular open-source library led by a collaboration of industry and academic researchers. It offers various FHE schemes and focuses on usability and security. (https://github.com/openfheorg)
- **SEAL (homomorphic encryption library):** Microsoft SEAL—powered by open-source homomorphic encryption technology—provides a set of encryption libraries that allow computations to be performed directly on encrypted data. This enables software engineers to build end-to-end encrypted data storage and computation services where the customer never needs to share their key with the service.(https://www.microsoft.com/en-us/research/project/microsoft-seal/)
- **IBM Cloud HEaaS (Homomorphic Encryption as a Service):** This offering allows experimentation and development with FHE in a cloud environment. (Availability details are best obtained from IBM directly)
- **Enveil:** This company offers a variety of homomorphic encryption solutions, including secure machine learning capabilities. https://www.enveil.com/

Searchable Symmetric Encryption (SSE):

Imagine a vast medical archive with patient data encrypted. SSE creates a searchable index for this archive, allowing researchers to find specific medical images (data points) using keywords without decrypting everything. They can search for a particular disease using a secret code (search token), and the system locates relevant images while keeping the patient data itself encrypted. Researchers can then decrypt only the specific images they need with a master key. (Think of finding relevant medical images in a secure library without needing to open every file.)

- Cloud storage providers with SSE capabilities: Some cloud storage providers, like Google Cloud Storage or Amazon S3, offer functionalities that incorporate SSE concepts. These services might not explicitly advertise "SSE" but provide features like client-side encryption with searchable indexes.
- **Paperclip SAFE:** is a data security platform offering "always-encrypted" functionalities. incorporates SSE principles for searchable encryption while data remains encrypted at rest and in use. (https://paperclip.com/safe/)
- **Pysearchable (Open Source Python library):** This library provides various SSE implementations in Python. (https://github.com/ko1o/PYSearch)

Secure Multi-Party Computation (SMPC):

This enables multiple parties to collaboratively analyze data without revealing their individual datasets. This is useful for institutions working together on AI projects without compromising sensitive information.

- **Major cloud providers:** While not explicitly marketed as SMPC, some cloud providers like Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform (GCP) offer functionalities that can be leveraged to build SMPC applications.
- Other companies working on SMPC include:
  - **Ironclad Systems:** (https://ironcladapp.com/lp/contract-securely-with-ironclad/)
  - **Sepior:** (https://www.crunchbase.com/organization/sepior)
  - **Duality Technologies:** (https://dualitytech.com/platform/duality-collaboration-hub/)

Open-source Libraries:

- **OpenSMPC:** (https://github.com/calcom/cal.com) - A popular open-source library offering various SMPC functionalities.
- **MP-SPDZ:** (https://github.com/data61/MP-SPDZ) - Another well-regarded open-source library for secure multi-party computation.
- **TF Enclave:** (https://github.com/tensorflow/tensorflow) - A library from Google focused on integrating SMPC functionalities with TensorFlow, a popular machine learning framework.

- **Federated Learning:** In this approach, AI models are trained on local devices holding individual data. Only the model updates, not the raw data itself, are shared for central aggregation. This keeps user data private while allowing for collaborative training.

Federated learning is especially useful in the following categories.

- Mobile Phone App Development
- Healthcare and Medical Research
- Internet-of-Things (IoT) and Smart Devices
- Additional Privacy-Preserving AI Systems

Local Differential Privacy:

A variant of differential privacy where noise is added to data on individual devices before it's used for training, further enhancing privacy for federated learning applications.Privacy-Preserving Support Vector Machines (SVMs): These algorithms are specifically designed to train SVM models while protecting the privacy of the data used. SVMs are a type of machine learning model used for classification tasks.

Secure K-Nearest Neighbors (KNN): Similar to privacy-preserving SVMs, these techniques allow training KNN models (used for classification and regression tasks) without compromising data privacy.

- **Differential Privacy:** This is a mathematical framework that ensures any information learned from a dataset cannot be linked back to a specific individual. It adds noise to the data in a controlled way, protecting privacy while still allowing for useful insights.

Research and Development Efforts:

- Several tech giants like Apple, Google, and Microsoft are actively researching and developing differential privacy techniques for their AI products.
  - o Apple has mentioned using differential privacy in some of its features like keyboard usage statistics and health data collection on iPhones. However, specifics about the products themselves are not always publicly available.

  - o **TensorFlow Privacy:** (https://www.tensorflow.org/responsible_ai/privacy/guide) This library from Google offers functionalities for incorporating differential privacy into TensorFlow, a popular machine learning framework.
  - o **OpenDP:** (https://github.com/opendp/opendp) This is an open-source project providing various tools and libraries for implementing differential privacy in machine learning.

Zero-Knowledge Proofs

A zero-knowledge proof (ZKP), also called a zero-knowledge protocol, is a mathematical technique to verify the truth of information without revealing the information itself. The method was first introduced by researchers from MIT in a 1985 paper.[1]

A series of cryptographic algorithms are used in the real-world applications of ZKPs to enable the verification of a computational statement. For instance, using ZKP methods, a receiver of payment can verify that the payer has sufficient balance in their bank account without getting any other information about the payer’s balance.

Check out ZKProof, an organization that seeks to standardize and popularize the use of zero-knowledge proof cryptography. For more on data privacy and information security, you can check our other articles:

Goldwasser, S.; Micali, S.; Rackoff, C. (1989), “The knowledge complexity of interactive proof systems” (PDF), *SIAM Journal on Computing*, **18** (1): 186–208.

Additional Considerations:

Transparent and Explainable AI (XAI): XAI techniques help users understand how AI models arrive at their decisions. This fosters trust and allows for identifying potential privacy risks within the AI system. Libraries like TensorFlow Explainable AI (TF-XAI) or DARPA's Explainable AI Toolkit (XAI Toolkit) offer tools and techniques for building and explaining machine learning models.