ISACA

Exploring Practical Considerations and Applications for Privacy Enhancing Technologies

Privacy

CONTENTS

ABSTRACT

While privacy enhancing technologies (PETs) are not new, the recent proliferation of machine learning systems, the rise of the data economy, support from policymakers, and increased privacy awareness among consumers have increased their popularity. PETs enable enterprises to harness the value of personal data while protecting individual user privacy. Recent technical advances and the availability of commercial and open-source solutions have made PETs accessible more broadly, enabling responsible data use. This white paper introduces popular PETs, offers evaluation guidance, highlights regulatory perspectives for using PETs for privacy compliance, and shares practical applications through case studies.

Introduction

Today, the unprecedented volume of personal data available to enterprises has created endless new opportunities for harnessing this data for insights using technologies such as artificial intelligence (AI), including machine learning (ML). However, these developments have also introduced new threats[1] and increased the risk of privacy harm to individuals whose personal data is utilized by these systems.

In response to the expanded threat surface, consumers are demanding that enterprises be good stewards of their data and use it responsibly. A recent survey[2] of about 5,000 consumers from 19 countries reported that nearly 68% are concerned about online privacy. Enterprises are challenged to find ways to extract value from personal data while respecting privacy.

Privacy enhancing technologies (PETs) are a promising solution. They support personal data analysis, sharing, and use while adhering to data protection principles and without negatively impacting privacy. PETs can help prevent downstream harms[3] by bolstering data protection practices. In the last decade, they have emerged from the research realm and started gaining industry adoption via commercial offerings and open-source solutions, lowering the cost barrier for implementation.

Defining PETs

Early definitions of PETs can be found in the 1995 Information and Privacy Commissioner Ontario report,[4] which described PETs as a "variety of technologies that safeguard personal privacy by minimizing or eliminating the collection of identifiable data," and the 2002 Organisation for Economic Co-operation and Development (OECD) Inventory of Privacy-Enhancing Technologies, which defined PETs as a "wide range of technologies that help protect personal privacy."[5]

While no specific legal definition of PETs in data privacy law exists, recent guidance[6] published by the UK Information Commissioner's Office (ICO) views PETs as "technologies that embody fundamental data protection principles by minimizing personal information use (this covers the legal definition of personal data in the UK GDPR); maximizing information security; or empowering people."

The International Organization of Standardization (ISO) defines PETs as a

> *privacy control, consisting of information and communication technology (ICT) measures, products, or services that protect privacy by eliminating or reducing personally identifiable information (PII) or by preventing unnecessary and/or undesired processing of PII, all without losing the functionality of the ICT system.*[7]

1 Shokri, R.; Stronati, M.; et al.; "Membership Inference Attacks Against Machine Learning Models," IEEE Symposium on Security and Privacy, 2017, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7958568
2 Fazlioglu, M.; "Privacy and Consumer Trust," IAPP, 2023, https://iapp.org/media/pdf/resource_center/privacy_and_consumer_trust_report.pdf
3 The Royal Society, "From Privacy to Partnership: The Role of Privacy Enhancing Technologies in Data Governance and Collaborative Analysis," January 2023, https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/From-Privacy-to-Partnership.pdf
4 Hes, R.; Borking, J.; *Privacy-Enhancing Technologies: The Path to Anonymity, Revised Edition*, The Hague, Netherlands, August 2000, https://www.researchgate.net/publication/243777645_Privacy-Enhancing_Technologies_The_Path_to_Anonymity
5 OECD, "Emerging Privacy Enhancing Technologies Current Regulatory and Policy Approaches," 23 March 2023, https://www.oecd-ilibrary.org/docserver/bf121be4-en.pdf?expires=1703180176&id=id&accname=guest&checksum=28CA591FCE3D9DAB72195AD8FEF0315D
6 Information Commissioner's Office UK, "Privacy Enhancing Technologies," 2023, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/
7 International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *ISO/IEC 29100:2024 Information technology — Security techniques — Privacy framework*, https://www.iso.org/obp/ui/#iso:std:iso-iec:29100:ed-2:v1:en

This white paper will adopt the European Union Agency for Cybersecurity (ENISA)[8] definition of PETs, which is "software and hardware solutions, i.e., systems encompassing technical processes, methods, or knowledge to achieve specific privacy or data protection functionality or to protect against risks to privacy of an individual or a group of natural persons."

PETs enable greater privacy and data utility within enterprises and promote collaborations externally in potentially competing organizations by reducing the risk associated with data sharing; hence, they have also been informally called partnership enhancing technologies[9] and trust technologies.[10]

Driving Growth in PET Markets

Processing sensitive personal data, such as health-related or financial information, and sharing it with third parties imposes legal liabilities and results in risk to enterprises. These challenges have stymied the full exploitation of data in the data ecosystem. PETs can extract value from underutilized data and enable potentially adversarial parties to perform data analysis without the need to trust each other.

PETs can extract value from underutilized data and enable potentially adversarial parties to perform data analysis without the need to trust each other.

In addition to being essential to safeguard privacy, they also provide significant economic advantages by unlocking the potential of data and enabling new business use cases. The following growth drivers will likely spur the expansion of the PET market over the next decade:

- **User expectations of privacy**—Customers expect enterprises to handle their data responsibly and ensure privacy. Nearly 68% percent of customers are concerned about online privacy.[11] This concern reflects how much they trust companies with their data, where loss of trust can result in revenue losses. A Cisco survey[12] reported that 76% of consumers discontinued using products and buying from organizations they did not trust with their data.
- **Evolving compliance mandates**—Laws governing data privacy exist in more than 130 nations,[13] resulting in a complex regulatory compliance landscape. Regulatory guidance on PETs to potentially reduce the compliance burden, such as the ENISA report on data protection engineering[14] and the UK ICO's PET guide,[15] is likely to accelerate PET implementation as a mechanism to support compliance efforts.
- **Innovation and new business opportunities**—As data supply chains extend, there is an increasing demand for collaboration and seamless data sharing among multiple parties. PETs enable multiparty collaborations while keeping data private and can support exploring new use cases previously deemed high-risk. Another example is the digital advertising industry, which is shifting to a cookieless future[16] with the need to evolve solutions for ad targeting, measurement, and attribution while still respecting user privacy. Some consortiums such as IAB Tech Lab are working on open-source solutions to evangelize PETs in the digital advertising industry.[17]

8 Hansen, M.; Hoepman J.; et al.; "Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies: Methodology, Pilot Assessment, and Continuity Plan," ENISA, December 2015, https://www.enisa.europa.eu/publications/pets/view/++widget++form.widgets.fullReport/@@download/Readiness+Analysis+for+the+Adoption+and+Evolution+of+Privacy+Enhancing+Technologies.pdf
9 Lundy-Bryan, L.; "Privacy Enhancing Technologies: Part 2—The Coming Age of Collaborative Computing," *Lunar Ventures: Insight Series*, 2021, https://www.kisacoresearch.com/sites/default/files/documents/pet_white_paper_part_2_-_the_coming_age_of_collaborative_computing10992.pdf
10 Infocomm Media Development Authority, "Singapore Grows Trust in the Digital Environment," June 2022, https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2022/singapore-grows-trust-in-the-digital-environment
11 *Op cit* Fazlioglu
12 Cisco, "Data Transparency's Essential Role in Building Customer Trust: Cisco 2022 Consumer Privacy Survey," 2022, https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-consumer-privacy-survey-2022.pdf
13 United Nations Conference on Trade and Development, "Data Protection and Privacy Legislation Worldwide," https://unctad.org/page/data-protection-and-privacy-legislation-worldwide
14 European Union Agency for Cybersecurity (ENISA), "Data Protection Engineering: From Theory to Practice," January 2022, https://www.enisa.europa.eu/publications/data-protection-engineering
15 *Op cit* Information Commissioner's Office UK, ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/
16 For more information on a cookieless future, see Defero, "How the Cookieless Future Impacts Digital Advertising," https://www.deferousa.com/the-cookieless-future-and-how-it-impacts-digital-advertising/
17 IAB Tech Lab, "Privacy Enhancing Technologies (PETs) Initiative," 14 February 2024, https://iabtechlab.com/pets/

- **Shift towards ethical data use**—Data ethics include the moral obligations of collecting, safeguarding, and responsibly using personal information. Advocacy groups and think tanks[18] are shifting the discussion of ethical data use beyond the realm of data scientists and chief data officers to board-level conversations. PETs can support data ethics and provide assurances while contributing to ongoing efforts such as the Canada CIO Strategy Council's AI Ethics Assurance Program[19] and the Institute of Electrical and Electronics Engineers (IEEE) Standard Model Process for Addressing Ethical Concerns During System Design.[20]
- **Emerging cryptocurrency market**—The cryptocurrency market size is estimated to be US $37.8 billion as of 2023[21] and likely to continue expanding with greater acceptance by institutions, increased awareness of decentralized finance platforms, and as a diversification tool to mitigate inflation fears. Privacy is critical for cryptocurrencies to safeguard transaction history. Hence, the growth of the cryptocurrency market is expected to fuel the growth of PETs such as zero-knowledge proofs (ZKPs), which enable verification of transactions without revealing sensitive financial data.

Classifying & Categorizing PETs

Several attempts have been made to classify and categorize PETs based on the underlying technology or the use cases to which they relate. These classes and categories can help enterprises determine which PETs may be best suited to their particular use case. The following examines a few examples of classification and categorization:

- The OECD taxonomy[22] classifies PETs into four categories: 1) data obfuscation, 2) encrypted data processing tools, 3) federated and distributed analytics, and 4) data accountability.
- The US Federal Reserve Bank of San Francisco PETs report[23] categorizes PETs by their function into three specific technologies: 1) altering data, 2) shielding data, and 3) systems and architecture.
- The UN PET Guide[24] adopts a concise categorization of PETs into 1) input privacy and 2) output privacy. Input privacy aims to allow multiple parties to submit data for computations without other parties accessing data in the clear, and output privacy aims to prevent the identification or re-identification of data from the disseminated output.
- The UK ICO's PETs guidance[25] classifies PETs that can help achieve data protection compliance, including data protection by design and default.
  1. PETs that derive or generate data that reduces or removes the identifiability of individuals to help fulfill the data minimization principle. Examples include differential privacy and synthetic data.
  2. PETs that "focus on hiding and shielding data [to] help achieve the requirements of the security principle." Examples include homomorphic encryption (HE) and ZKPs.
  3. PETs that "split or control access [to] personal data to help fulfill both data minimization and security principles depending on the nature of the processing." Examples include trusted execution environments (TEEs), secure multiparty computation (SMPC), and federated learning.
- The Center for Data Ethics and Innovation "PETs Adoption Guide"[26] categorizes PETs based on use cases. The two broad

18 Ethical Tech Project, "Commitment to the Ethical Use of Data," https://www.ethicaltechproject.com/initiatives
19 Digital Governance Council, "CIO Strategy Council Launches AI Ethics Assurance Program in Collaboration With KPMG in Canada," 16 November 2020, https://dgc-cgn.org/cio-strategy-council-launches-ai-ethics-assurance-program-in-collaboration-with-kpmg-in-canada/
20 Institute of Electrical and Electronics Engineers, "IEEE Launches New Standard to Address Ethical Concerns During Systems Design," 15 September 2021, https://standards.ieee.org/news/ieee-7000/
21 Statista, "Cryptocurrencies Worldwide," 2023, https://www.statista.com/outlook/dmo/fintech/digital-assets/cryptocurrencies/worldwide
22 *Op cit* OECD 2023, https://doi.org/10.1787/bf121be4-en
23 Asrow, K.; Samonas, S.; "Privacy Enhancing Technologies: Categories, Use Cases, and Considerations," Federal Reserve Bank of San Francisco, California, 1 June 2021, https://www.frbsf.org/banking/publications/fintech-edge/2021/june/privacy-enhancing-technologies/
24 United Nations Committee of Experts on Big Data and Data Science for Official Statistics, https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf
25 Information Commissioner's Office UK, "Chapter 5: Privacy-Enhancing Technologies ('PETs')," *Draft Anonymization, Pseudonymization, and Privacy Enhancing Technologies Guidance*, September 2022, https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf
26 Center for Data Ethics and Innovation, "Privacy Enhancing Technologies Adoption Guide," https://cdeiuk.github.io/pets-adoption-guide/what-are-pets

categories are 1) traditional PETs covering encryption in transit, encryption at rest, and de-identification techniques; and 2) emerging PETs that include homomorphic encryption, trusted execution environments, multiparty computation, differential privacy, and federated analytics.

For the purpose of simplification, this paper will use the Center for Data Ethics and Innovation's case-based categories and focus on emerging PETs. The following section examines common PETs.

Trusted Execution Environment

Trusted execution environment (TEE) is a dedicated area on a computer processor that is separated and secured from the operating system (OS). It stores data and runs code within its secured area. TEE assumes the OS is untrustworthy and does not allow the operating system to access data stored in the secure area. TEE can be used when sensitive data needs to be stored safely or there is a need to generate insights from data without revealing the dataset to the party running the analysis or hosting the TEE.

TEE is an environment that provides a level of assurance for three main properties:

- **Data Confidentiality**—No view access to data for unauthorized parties
- **Data Integrity**—No ability to add, remove, or modify data for unauthorized parties
- **Code Integrity**—No ability to add, remove, or modify code for unauthorized parties[27]

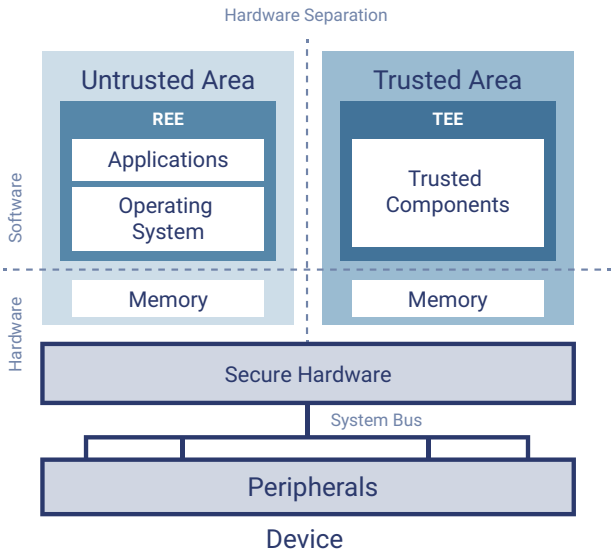
In addition to providing assurance that data is kept secured, the properties also help prove that the computations performed are correct, enabling trust in the computation results as well.[28]

As shown in **figure 1**, TEE is typically implemented partly in the hardware of a CPU and partly in associated software libraries.

Standards related to TEEs include:

- ISO/IEC 11889-4:2015 *Information Technology — Trusted Platform Module Library*[29]
- IETF *Trusted Execution Environment Provisioning (TEEP) Architecture*[30]
- IEEE 2830-2021 *IEEE Standard for Technical Framework and Requirements of Trusted Execution Environment based Shared Machine Learning*[31]
- GPD_SPE_055 *TEE Trusted User Interface Low-level API*[32]

FIGURE 1: Operating System Support for Run-Time Security with a Trusted Execution Environment



Source: Adapted from Gonzalez, J.; "Operating System Support for Run-Time Security with a Trusted Execution Environment (Doctoral Thesis)," ResearchGate, March 2015, https://www.researchgate.net/publication/297732884_Operating_System_Support_for_Run-Time_Security_with_a_Trusted_Execution_Environment

27 Confidential Computing Consortium, "A Technical Analysis of Confidential Computing," November 2022, https://confidentialcomputing.io/wp-content/uploads/sites/10/2023/03/CCC-A-Technical-Analysis-of-Confidential-Computing-v1.3_unlocked.pdf
28 *Ibid.*
29 ISO/IEC, ISO/IEC 11889-4:2015 *Information technology — Trusted Platform Module Library — Part 4: Supporting Routines*, Revised 2021, https://www.iso.org/standard/66513.html
30 Pei, M.; Tschofenig, H.; et al.; "Trusted Execution Environment Provisioning (TEEP) Architecture Draft," Internet Engineering Task Force (IETF), July 2023, https://datatracker.ietf.org/doc/rfc9397/
31 IEEE, IEEE 2830-2021 *IEEE Standard for Technical Framework and Requirements of Trusted Execution Environment based Shared Machine Learning*, 22 October 2021, https://standards.ieee.org/ieee/2830/10231/
32 Global Platform, *TEE Trusted User Interface Low-level API v1.0.1*, November 2018, https://globalplatform.org/specs-library/globalplatform-technology-tee-trusted-user-interface-low-level-api-v1-0-1/

Benefits

TEEs ensure data accuracy, privacy, and consistency by limiting access to unencrypted data. Data utility does not suffer as actual computation is done on unencrypted and noiseless data. They implement code assurance by authenticating it every time before loading data into memory.

When combined with other PETs, such as multiparty computation (MPC), TEEs can facilitate collaboration among distrusting parties, allowing code to be tested without direct export.

Limitations and Challenges

TEEs are vulnerable to side-channel attacks and timing attacks, which can leak cryptographic keys or infer information about the underlying operation of the TEE.

They also have higher acquisition and maintenance costs compared to software-based PETs. Commercial TEE solutions offer limited support for distributed computation on large data sets.

Example Applications

Apple's Secure Enclave is a dedicated secure subsystem in the latest versions of the iPhone, iPad, Mac, Apple Watch, etc. The Secure Enclave is isolated from the untrusted main processor and designed to keep user data safe even when the application processor kernel becomes compromised. It protects sensitive data such as user biometric data and encryption keys used by iOS and third-party applications.

Google's Trusty is a secure OS that provides a TEE for Android. Trusty and the Android OS run in parallel, with Trusty isolated from the rest of the system by both hardware and software. Trusty's isolation protects it from malicious applications that the user installs and potential vulnerabilities that may be discovered in Android.

Homomorphic Encryption

Homomorphic encryption (HE) is a cryptographic technique that directly computes encrypted data without ever decrypting. The computations are also encrypted, and only the party providing the data has the decryption key for the output. With HE, there is no need to mask or drop any features to preserve the privacy of data, enabling all features to be used in analysis without compromising privacy.

There are three types of HE, and the selection of the appropriate scheme will depend on use case, scale, types of mathematical operations, and data utility needs:

- **Fully Homomorphic Encryption (FHE)**—Supports all types of operations and has no limits on the number of operations
- **Somewhat Homomorphic Encryption (SHE)**—Supports addition and multiplication on encrypted data but places limits on the number of operations
- **Partial Homomorphic Encryption (PHE)**—Supports only addition or multiplication but not both

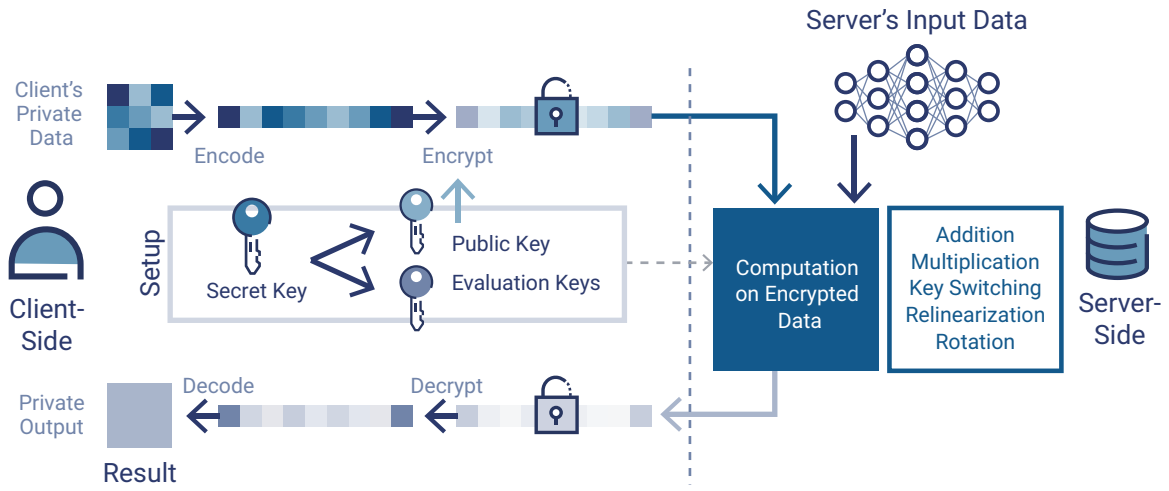
HE relies on a public key generation algorithm to generate a pair of private (or secret) and public keys and an evaluation key. As shown in **figure 2**, the client's public key is used to encrypt the data, and the evaluation key is used to perform computations on the encrypted data and is shared with another entity. The client, who retains the private key, decrypts the output and obtains the computation results. Because the entity possesses only the client's public key and the evaluation key, it cannot learn about the results. The data remains encrypted and requires the client's private key for decryption.

Standards relating to HE include:

- Homomorphic Encryption Standard 2018[33]
- ISO/IEC 18033-6:2019 *IT Security techniques — Encryption algorithms — Part 6: Homomorphic encryption*[34]
- ISO/IEC AWI 18033-8 *Information security — Encryption algorithms — Part 8: Fully Homomorphic Encryption*[35]

33 Albrecht, M.; Chase, M.; et al.; "Homomorphic Encryption Standard," Homomorphic Encryption Standardization, 21 November 2018, http://homomorphicencryption.org/wp-content/uploads/2018/11/HomomorphicEncryptionStandardv1.1.pdf
34 ISO/IEC, ISO/IEC 18033-6:2019 *IT Security techniques — Encryption algorithms — Part 6: Homomorphic encryption*, May 2019, https://www.iso.org/standard/67740.html
35 ISO, ISO/IEC WD 18033-8 *Information security — Encryption algorithms — Part 8: Fully Homomorphic Encryption*, https://www.iso.org/standard/83139.html. Note that the status of this standard is deleted.

FIGURE 2: Typical Computation Flow for FHE



Source: Adapted from Riazi, S.; "From Fully Homomorphic Encryption to Silicon — What is Microsoft's HEAX?," OpenMined, 7 July 2020, https://blog.openmined.org/from-fully-homomorphic-encryption-to-silicon/

Benefits

HE can be used to obtain insights from computation without revealing the contents of a dataset to those running the analysis. It reduces the risk of data breaches because personal data remains encrypted at rest, in transit, and during computation. It eliminates the need for trusted parties and can be used in public cloud environments and enable secure outsourcing.

HE can be used to obtain insights from computation without revealing the contents of a dataset to those running the analysis.

Additionally, HE can provide a level of guarantee that the result of the computation is comparable to that on the unencrypted data because the data is not altered (i.e., no noise is added). Data utility is preserved as there is no need to drop data features to preserve privacy.

Limitations and Challenges

Because only one secret (decryption) key exists, HE does not provide input privacy for more than a single party. HE typically has higher computation costs and needs cryptographic expertise for development. It is worth noting that HE does not guarantee security if the adversary obtains decryptions of selected cipher texts. HE also requires selection of the appropriate algorithm and key size to ensure the private keys remain secure.

Example Applications

IBM Research HE4Cloud[36] is a FHE service used to deploy privacy-preserving computing on the cloud. It allows clients to deploy their machine learning models and use encrypted data to train them or run inference requests in a cloud-native software as a service (SaaS) experience.

The Dana-Farber Cancer Institute and Duality Technologies, Inc. collaborated to leverage HE to drive insights from multisourced, encrypted data without ever decrypting it for secure, large-scale genomewide association studies.[37]

Secure Multiparty Computation

Secure multiparty computation (SMPC) enables multiple parties to analyze their combined data without disclosing its contents to one another. It uses a cryptographic technique called secret sharing, where each participating party's data is split into fragments and distributed to the other parties.

36 IBM Research, "IBM FHE Cloud Service," https://he4cloud.com/public
37 Blatt, M.; Gusev, A.; et al.; "Secure Large-scale Genome-wide Association Studies Using Homomorphic Encryption," PNAS, 12 May 2020, https://www.pnas.org/doi/10.1073/pnas.1918257117

Another SMPC cryptographic technique is private set intersection (PSI), which enables two parties to compare their data and identify the common elements while maintaining the privacy of the remaining data.

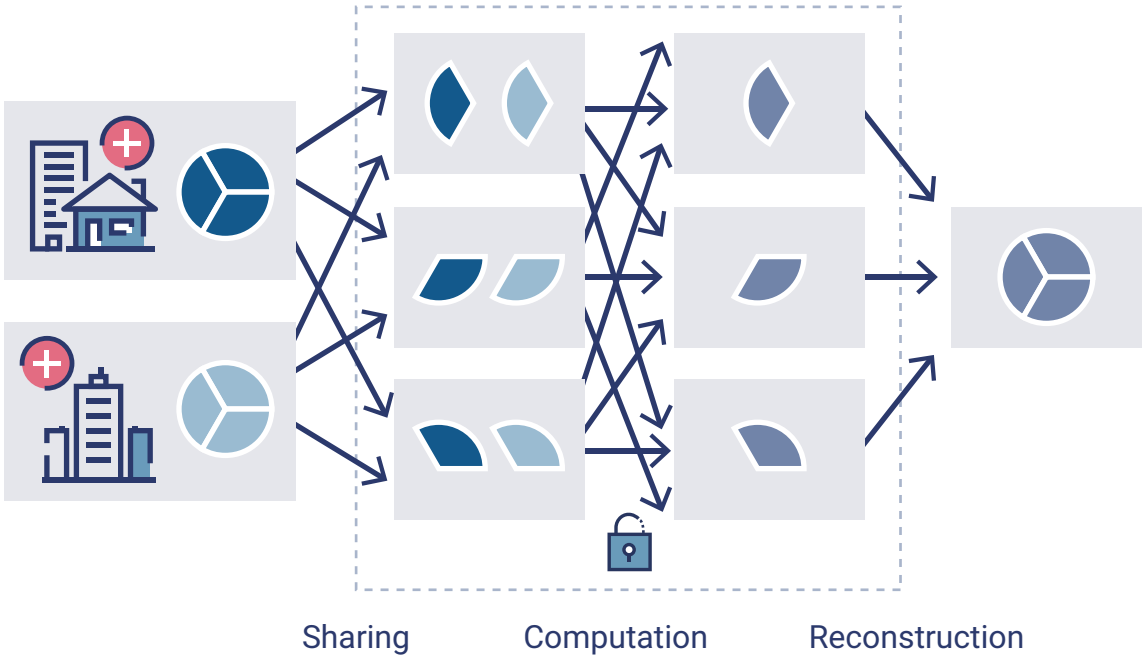
The risk of exposure through unintentional errors or malicious compromise is limited because each party has only a subset of the data. The data can only be revealed if fragments of each party's data are combined, which is unlikely because it would compromise the data security of multiple parties.

Figure 3 illustrates SMPC, where security values (denoted by dark and light blue pies) are split into any number of shares distributed among the computing nodes. No computation node can recover the original value or learn anything about the output (gray pie) during the computation. Any node can combine its shares to reconstruct the initial value.

Standards related to SMPC include:

- IEEE 2842-2021 *IEEE Recommended Practice for Secure Multi-Party Computation*[38]
- IETF Privacy Preserving Measurement (PPM) protocol standard[39]
- ISO/IEC 19592-2:2017 *Information technology — Security techniques — Secret sharing — Part 2: Fundamental mechanisms*[40]
- ISO/IEC CD 4922-1:2023 *Information security — Secure multiparty computation — Part 1: General*[41]
- ISO/IEC 4922-2 *Information security — Secure multiparty computation — Part 2: Mechanisms based on secret sharing (Draft)*[42]

FIGURE 3: Secure Multiparty Computation



Source: Adapted from Xu, J.; Glicksberg, B.; et al.; "Federated Learning for Healthcare Informatics," Journal of Healthcare Informatics Research, March 2021, https://www.researchgate.net/figure/Privacy-preserving-schemes-a-Secure-multi-party-computation-In-security-sharing_fig3_346526433

38 IEEE, IEEE 2842-2021 *IEEE Recommended Practice for Secure Multi-Party Computation*, 5 November 2021, https://standards.ieee.org/ieee/2842/7675/
39 IETF, *Privacy Preserving Measurement* (PPM) protocol standard, https://datatracker.ietf.org/doc/draft-gpew-priv-ppm/
40 ISO/IEC, ISO/IEC 19592-2:2017 *Information technology — Security techniques — Secret sharing — Part 2: Fundamental mechanisms*, October 2017, https://www.iso.org/standard/65425.html
41 ISO/IEC, *ISO/IEC CD 4922-1:2023 Information security — Secure multiparty computation — Part 1: General*, July 2023, https://www.iso.org/standard/80508.html
42 ISO/IEC, *ISO/IEC 4922-2 Information security — Secure multiparty computation — Part 2: Mechanisms based on secret sharing (Draft)*, https://www.iso.org/standard/80514.html

Benefits

SMPC can prevent data leakage by enabling computational inference to be made on encrypted data. Additionally, it eliminates the need for a trusted central authority that would have access to everyone's data. SMPC maintains data utility because the data is not masked. It allows multiple distrusting parties to collaborate because data remains safe from unwarranted interference. Furthermore, SMPC can also be secured from quantum attacks.[43]

Limitations and Challenges

With SMPC, the computational overhead of random number generation can slow down the run time. SMPC requires threat modeling and accurate predictions about the malicious parties participating in SMPC. Additional costs are associated with communication and connectivity among all parties as required for secret sharing, which can also lead to scalability issues. Deploying SMPC protocols correctly also requires significant technical expertise.

Example Applications

SMPC enabled researchers at a large healthcare provider to privately compute across organizational data sources to increase both sample size and patient attributes, leading to improved model performance and heart-disease prognosis.[44] Five agencies in a US county government leveraged SMPC to run sensitive queries involving incarceration status, usage of mental health facilities, and public housing benefits while keeping the input data (e.g., criminal records and mental health visit records) strictly confidential to each party that provided the data.[45]

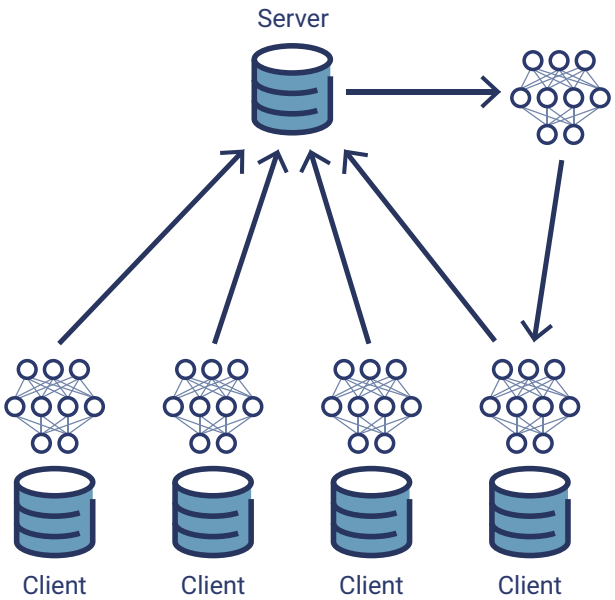
Federated Learning

Federated learning (FL) is an architectural PET that enables multiple parties to train models on their own data (i.e., local models). The parties then combine some of the patterns identified by those models into a single, more accurate global model without having to share any training data. FL localizes the control of data and even the control of models running on that data. There are two approaches to federated learning: centralized and decentralized.

> *In centralized FL, a coordination server creates a model or algorithm, and duplicate versions of that model are sent out to each distributed data source. The duplicate model trains itself on each local data source and sends back the analysis it generates. That analysis is synthesized with the analysis from other data sources and integrated into the centralized model by the coordination server. This process repeats itself to refine and improve the model [continually].*

Centralized FL is easier to manage, as there is a single point of control for the training phase, and it can accommodate a large number of clients (**figure 4**).

FIGURE 4: Centralized Federated Learning



Source: Information Commissioner's Office UK, "Chapter 5: Privacy-Enhancing Technologies ('PETs')," *Draft Anonymization, Pseudonymization, and Privacy Enhancing Technologies Guidance*, September 2022, https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf. Contains public-sector information licensed under the Open Government License v3.0.

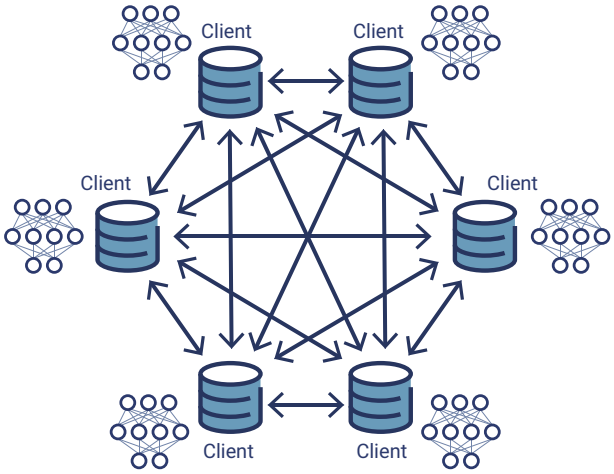
43 Chainlink, "Secure Multi-Party Computation," 30 November 2023, https://chain.link/education-hub/secure-multiparty-computation-mcp
44 Inpher, "Augmented Heart Disease Analysis," https://inpher.io/solutions/by-industry/healthcare/#heart-disease-analysis
45 Hart, N.; Archer, D.; et al.; "Privacy-Preserved Data Sharing for Evidence-Based Policy Decisions: A Demonstration Project Using Human Services Administrative Records for Evidence-Building Activities," SSRN, March 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3808054

With decentralized FL, as shown in **figure 5**, data remains on the user's devices/servers, and models are directly updated; no central coordination server is involved. Each participating entity communicates with the others, and they can all update the global model directly. This scheme is more resilient than central FL as there is no single point of failure.

FL schemes that use different features about the same set of users across multiple parties for training models without exposing the raw data or model parameters are referred to as vertical FL, while schemes that use the same features about the same set of users across multiple parties are referred to as horizontal FL.

IEEE 3652.1-2020 *IEEE Guide for Architectural Framework and Application of Federated Machine Learning*[46] is a standard related to FL.

FIGURE 5: Decentralized Federated Learning



Source: Information Commissioner's Office UK, "Chapter 5: Privacy-Enhancing Technologies ('PETs')," *Draft Anonymization, Pseudonymization, and Privacy Enhancing Technologies Guidance*, September 2022, https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf. Contains public-sector information licensed under the Open Government License v3.0.

Benefits

FL allows data owners to retain control and not share raw data. It (especially decentralized FL) is resilient to network failure and disruptions thanks to the decentralization of training. FL can support multiple parties in submitting model improvements while keeping sensitive data private. Additionally, FL does not require complex hardware.

Limitations and Challenges

FL requires reliable connectivity. It is susceptible to attacks where malicious servers can learn about specific data points from the training data or where data input (poisoning) can lead to performance degradation. Local data sets may have limitations around accuracy and labels. Additionally, data sets and parameters of local nodes must be interoperable with other nodes. A consideration that must be accounted for is that the characteristics of data sets may change over time.

Example Applications

Google uses FL to improve on-device machine learning models such as "Hey Google" in Google Assistant,[47] which allows users to issue voice commands. It is also used for next-word prediction[48] on virtual keyboards for smartphones.

> *FL has been evaluated in healthcare to support precision medicine research by gaining insights collaboratively… without moving patient data beyond the firewalls of the institutions in which they reside … Research has shown that models trained by FL can achieve performance levels comparable to ones trained on centrally hosted data sets and superior to models that only see isolated single-institutional data.*[49]

46 IEEE SA, *IEEE 3652.1-2020: IEEE Guide for Architectural Framework and Application of Federated Machine Learning*, 19 March 2021, https://standards.ieee.org/ieee/3652.1/7453/
47 Google, "Your Voice & Audio Data Stays Private While Google Assistant Improves," https://support.google.com/assistant/answer/10176224?hl=en
48 Hard, A.; Rao, K.; et al.; "Federated Learning for Mobile Keyboard Prediction," ARXIV, 28 February 2019, https://arxiv.org/abs/1811.03604
49 Rieke, N.; Hancox, J.; et al.; "The Future of Digital Health with Federated Learning," NPJ Digital Medicine, 14 September 2020, https://www.nature.com/articles/s41746-020-00323-1

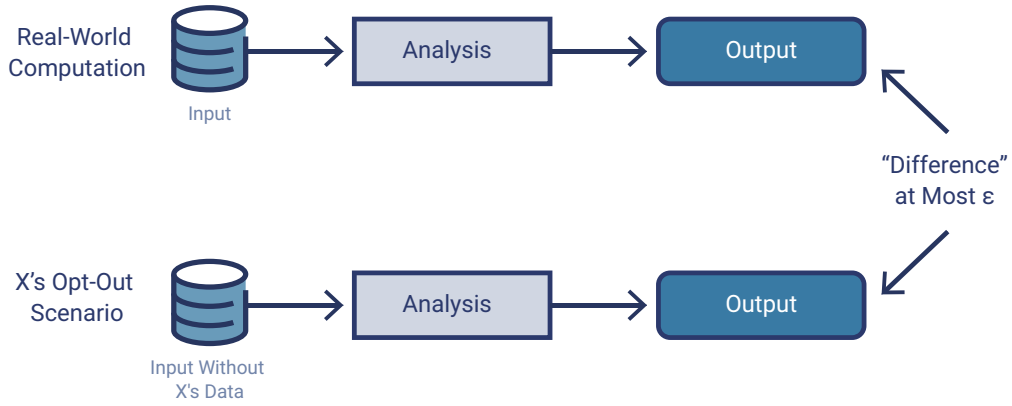
Differential Privacy

Differential privacy (DP) is a mathematical framework for ensuring the privacy of individuals in data sets. It achieves this by introducing a controlled amount of random noise into the data, effectively concealing the contribution of individual data points. This means that the results of any analysis remain largely unchanged, whether an individual's data is included or excluded from the dataset. It can provide a strong guarantee of privacy by allowing data to be analyzed without revealing sensitive information about any individual in the data set. Noise allows for plausible deniability, i.e., uncertainty about the actual value of private variables in a system,[50] for a data set containing a particular individual's data.

Differential privacy (DP) is a mathematical framework for ensuring the privacy of individuals in data sets. It achieves this by introducing a controlled amount of random noise into the data, effectively concealing the contribution of individual data points.

As shown in **figure 6**, DP introduces a privacy loss or privacy budget parameter to the data set, often denoted as epsilon (ε), which controls how much noise or randomness is added to the raw data set. ε0 completely protects privacy at the expense of accuracy because it introduces only noise. "X's opt out scenario" in the figure represents what would happen if an individual's information was not included in the dataset.

FIGURE 6: Differential Privacy



Source: Information Commissioner's Office UK, "Chapter 5: Privacy-Enhancing Technologies ('PETs')," *Draft Anonymization, Pseudonymization, and Privacy Enhancing Technologies Guidance*, September 2022, https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf. Contains public-sector information licensed under the Open Government License v3.0.

Two types of DP are available: "global differential privacy, which adds noise during aggregation; and local differential privacy, where each user adds noise to individual records before aggregation."[51]

Global DP is more accurate than local DP as it does not need to add a lot of noise for similar levels of privacy protection. The drawback of global DP is that it needs a trusted aggregator and has a single point of failure, which increases security risk.

Global DP is more accurate than local DP as it does not need to add a lot of noise for similar levels of privacy protection.

Currently, there are no standards available related to DP. However, in response to the US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 2023), the National Institute of Standards and Technology published draft "Guidelines for Evaluating Differential Privacy Guarantees."[52]

50 Monshizadeh, N.; Tabuada, P.; "Plausible Deniability as a Notion of Privacy," 2019 IEEE 58th Conference on Decision and Control, December 2019, https://ieeexplore.ieee.org/document/9030201
51 *Op cit* Information Commissioner's Office UK 2022
52 National Institute of Standards and Technology, *NIST SP 800-226 (Initial Public Draft): Guidelines for Evaluating Differential Privacy Guarantees*, 11 December 2023, https://csrc.nist.gov/pubs/sp/800/226/ipd

Benefits

DP provides measurable privacy guarantees that are easy to communicate. It allows for customizing the privacy afforded based on the context of the data and use case. DP is commercially ready and proven to scale for large data sets, offering protection from linkage attacks.

Additionally, it prevents attackers from accessing raw data. DP provides provable privacy guarantees with respect to the cumulative risk from successive data releases. It also enables transparency as computation and its parameters can be open.

DP provides provable privacy guarantees with respect to the cumulative risk from successive data releases.

Limitations and Challenges

It takes specialized skills and competencies to correctly implement DP. Depending on the use case, added noise may reduce the utility of the data. DP is also not suitable for low-count data sets and detecting anomalies within data. In the absence of industry-accepted guidelines and standards for DP, there is no consensus for setting and tuning privacy parameters.

Example Applications

The US Census Bureau started using DP with the 2020 Census to prevent the re-identification of US citizens who provided detailed demographic information[53] while still allowing the release of aggregate statistics about the population.

Microsoft uses DP to collect telemetry across millions of devices, employing the locally differentially private (LDP) mechanism[54] designed for the repeated collection of counter data. This mechanism provides formal privacy guarantees even after an extended period of execution.

Synthetic Data

Synthetic data transforms a sensitive data set into a new data set with similar statistical properties without revealing information on individuals from the original data set. It is generated from real data using a model trained to reproduce the characteristics and structure of that data.

Synthetic data transforms a sensitive data set into a new data set with similar statistical properties without revealing information on individuals from the original data set.

Two main types of synthetic data exist:

- Partially synthetic data synthesizes only some variables of the original data
- Fully synthetic data synthesizes all variables

Currently, no standards exist related to synthetic data.

Benefits

Synthetic data can reduce the risk of data breaches because it cannot be easily linked to individuals. Furthermore, it reduces data management costs associated with secure storage and maintenance.

Generating synthetic data is faster than data gathering and preparation. Additionally, synthetic data provides enterprises with greater control over the quality and format of data.

Limitations and Challenges

Bias may be amplified in synthetic data, especially if source data is not neutral. Applications of synthetic data are susceptible to data leaks and reconstruction attacks, particularly in the case of outliers in the data set.[55] Moreover, it may not be applicable for use cases that require high degrees of accuracy.

53 Census, "2020 Decennial Census: Processing the Count: Disclosure Avoidance Modernization," https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html
54 Ding, B.; Kulkarni, J.; et al.; "Collecting Telemetry Data Privately," December 2017, https://www.microsoft.com/en-us/research/publication/collecting-telemetry-data-privately/
55 Ganev, G.; De Cristofaro, E.; "On the Inadequacy of Similarity-based Privacy Metrics: Reconstruction Attacks against 'Truly Anonymous Synthetic Data,'" 8 December 2023, https://arxiv.org/pdf/2312.05114.pdf

Example Applications

American Express generates statistically accurate synthetic data from financial transactions to perform fraud detection and help train detection models.[56]

Waymo uses synthetic data[57] to develop realistic driving data sets to train its self-driving vehicle systems. This helps the enterprise generate varied situations for training data without spending significant time and resources to gather it from real cases.

Zero-Knowledge Proof

Zero-knowledge proof (ZKP) is a cryptographic method used to prove knowledge about data without revealing the data itself. There are two main types of zero-knowledge proofs:

- **Interactive zero-knowledge proofs**—The prover and the verifier interact several times. The verifier challenges the prover, who provides replies to these challenges until the verifier is convinced.
- **Noninteractive zero-knowledge proofs**—Proof delivered by the prover can be verified by the verifier only once. This is computationally more expensive than interactive ZKPs.

ZKPs can be used for decentralized identity and authentication management. For example, a ZKP-based identity solution can verify a person's citizenship without them having to provide their passport information, or perform age verification without them needing to disclose their date of birth.

ZKPs can be used for decentralized identity and authentication management.

Standards related to ZKP include:

- ISO/IEC 9798-5:2009 *Information technology — Security techniques — Entity authentication — Part 5: Mechanisms using zero-knowledge techniques*[58]
- ZKProof Community Reference[59]

Benefits

ZKPs enable secure transactions and verification schemes. Because data is not stored in a centralized location with ZKPs, they may be secure from cyberattacks.[60]

Limitations and Challenges

Implementing ZKPs requires significant technical expertise and knowledge of cryptographic protocols. Generating and verifying proofs can be computationally intensive. Additionally, scalability issues have made ZKPs impractical for use cases that require timely response. Hence, ZKPs have found applications in limited areas such as blockchain technology.

Implementing ZKPs requires significant technical expertise and knowledge of cryptographic protocols.

Example Applications

Cryptocurrency Zcash uses ZKPs to verify an individual's crypto wallet balance and transaction history without revealing the actual information to assure the blockchain network that the sender's balance will cover the transaction.[61]

ING Bank uses a zero-knowledge range proof solution that allows mortgage applicants to prove that their salary lies within a certain range, without revealing the exact figure.[62]

56 Vanian, J.; "Why American Express is Trying Technology that Makes Deepfake Videos Look Real," Fortune, 3 September 2020, https://fortune.com/2020/09/03/american-express-deepfake-artificial-intelligence/
57 Wiggers, K.; "The Challenges of Developing Autonomous Vehicles during a Pandemic," Venturebeat, 28 April 2020, https://venturebeat.com/ai/challenges-of-developing-autonomous-vehicles-during-coronavirus-covid-19-pandemic/
58 ISO/IEC, ISO/IEC 9798-5:2009 *Information technology — Security techniques — Entity authentication — Part 5: Mechanisms using zero-knowledge techniques*, December 2009, https://www.iso.org/standard/50456.html
59 ZKProof, "Community Reference," https://docs.zkproof.org/reference
60 Ray, A.; "Zero-Knowledge Proof: A Revolutionary Leap In Data Protection," Forbes, 25 July 2023, https://www.forbes.com/sites/forbesbusinesscouncil/2023/07/25/zero-knowledge-proof-a-revolutionary-leap-in-data-protection/?sh=f5a7d0772b76
61 Zcash, "Learn Zcash: What Are Zero-Knowledge Proofs?," https://z.cash/learn/what-are-zero-knowledge-proofs/
62 ING Bank, "ING launches Zero-Knowledge Range Proof solution, a major addition to blockchain technology," 16 November 2017, https://www.ingwb.com/en/insights/distributed-ledger-technology/ing-launches-major-addition-to-blockchain-technology

Key Factors in PET Selection

PET selection is highly dependent on context, specific business case, and privacy requirements. Several attempts have been made to develop PET selection decision trees,[63] covering both traditional and emerging PETs.

The Center of Data Ethics and Innovation PETs Adoption Guide[64] offers an interactive tool to assist technical architects and product owners in selecting PETs for sensitive data projects. The guide includes supporting technical references and a use case repository. However, these guides are not comprehensive. The pros and cons for each PET or combination of PETs being considered need to be weighed deliberately for the enterprise's specific requirements.

While there is no standard process for PET decision making, enterprises should consider the following factors for PET selection:

- **Performing a data protection impact assessment**—To assess whether the enterprise should consider a PET deployment, a good strategy is to perform a data protection impact assessment (DPIA)[65] to evaluate the risk for the specific data processing use case. The DPIA should consider the data processing context, scope, and purposes. It should assess necessity, proportionality, and compliance measures. Furthermore, it should identify the risk to impacted individuals and evaluate the measures to be taken, factoring in the cost of potential PET implementation for risk mitigation.
- **Developing the business case**—After deciding to mitigate risk through PET implementation, enterprises should consider developing a business case[66] with detailed requirements, including:
  - List of the various stakeholders involved, including the intended users of the technology and their goals
  - Type and volume of data to be processed
  - Data source(s) and destination(s)
  - Data output requirements
  - Access control requirements for the data
  - Computation of resource expectations
  - Privacy guarantees requirements
- **Reviewing data governance maturity**—Data governance is a prerequisite for PET deployment. It is essential to understand what personal data is in scope, where it is currently located, where it will be processed, and how it will be used. If data governance policies have been established, enterprises should understand which policy requirements would apply to the data under evaluation for PETs and assess how the policies will be enforced. Metadata, such as data labels, can be used to help understand the sensitivity of data and granular data protection requirements.

Data governance is a prerequisite for PET deployment. It is essential to understand what personal data is in scope, where it is currently located, where it will be processed, and how it will be used.

- **Building the evaluation criteria**—To ensure all requirements are factored in and assessed, enterprises should develop comprehensive evaluation criteria to aid decision making. The following comprise some common criteria that can be included in the evaluation process:
  - **Privacy goals**—Enterprises should assess the privacy goals for input and output privacy; SMPC and HE might be appropriate for input privacy, while output privacy may need DP.

63 Fekete, A.; "What are Privacy Enhancing Technologies? The 5 Best PETs for the Modern Tech Stack," 26 April 2022, Mostly.AI, https://mostly.ai/blog/what-are-privacy-enhancing-technologies; Thaine, P.; "Privacy Enhancing Technologies Decision Tree," Private-AI,18 October 2020, https://www.private-ai.com/2020/10/18/privacy-enhancing-technologies-decision-tree-v2/
64 *Op cit* Center for Data Ethics and Innovation
65 Information Commissioner's Office UK, "Data Protection Impact Assessments," https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/accountability-and-governance/data-protection-impact-assessments/; ISACA, "GDPR Data Protection Impact Assessments," 2017, https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004KoGtEAK
66 *Op cit* United Nations Committee of Experts on Big Data and Data Science for Official Statistics

- **Privacy guarantee requirements**—Enterprises should assess the privacy assurance requirements in collaboration with legal teams. DP and HE can provide provable privacy assurance for end users.
- **Number of parties involved**—Enterprises should assess the parties involved in the computation and whether they are trusted or nontrusted players. HE can support one data provider while TEE and SMPC can support multiple data providers.

Enterprises should assess the parties involved in the computation and whether they are trusted or nontrusted players.

- **Flexibility and scalability**—Scope creep and requirement changes are common in the dynamic environments in which most enterprises operate; therefore, enterprises should assess the adaptability of the PETs under consideration. Late changes in HE and SPMC implementations negatively impact time and cost.
- **Performance expectations**—Enterprises should assess performance needs because certain use cases may be sensitive to even minor degradation in computing speeds; for example, HE is typically considerably slower than plaintext operations.
- **Dependencies on other systems**—PETs typically must integrate with additional security and data tools, such as identity and access management solutions, data preparation tooling, and key management technologies. Integrations can introduce overheads and should be assessed early in the decision-making process.
- **Implementation expertise**—Specialized skills such as cryptography expertise can be hard to find, often making the development of PET solutions in-house challenging. Decision making should factor in the resource availability, skillset needs, and commercial readiness of the PETs being considered.
- **Configuration changes**—Privacy risk may evolve over time, and suitability decisions about PETs should include the ease of configuration changes to address the dynamic threat landscape.
- **Transparency objectives**—Enterprises should ensure that the PETs selected are auditable and should offer an honest representation about the PETs' capabilities to consumers and third parties. It is worth noting that the FTC has brought cases against organizations that failed to keep their privacy promises to consumers.[67]

- **Considering a single PET vs. combined PETs**—PETs are not guaranteed solutions for all privacy risk and business requirements. Enterprises should consider the combinations of PETs that can help, rather than making a decision based on the shortcomings of single PET solutions.

PETs are not guaranteed solutions for all privacy risk and business requirements.

HE and SMPC are frequently used to balance the speed and flexibility of operations. Similarly, DP and synthetic data are commonly combined to complement privacy protection features.

Regulatory Perspectives on PETs

Privacy regulation is technology-neutral; hence, most privacy laws do not explicitly reference PETs. PETs can support privacy by design and commonly accepted privacy principles.

However, the extent to which PETs can enable privacy regulatory compliance is unclear and requires careful analysis.

67 Fondrie-Teitler, S.; "Keeping Your Privacy Enhancing Technology (PET) Promises," Federal Trade Commission, 1 February 2024, https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/keeping-your-privacy-enhancing-technology-pet-promises

Leveraging PETs for Privacy by Design and Other Privacy Principles

Privacy by design is about integrating privacy into the entire engineering process in an enterprise. Underlying privacy by design is the notion that data protection in data processing procedures is best adhered to when it is already integrated into the technology when created. Enterprises should regularly evaluate technical advances in data protection technologies and deploy suitable technical controls, thereby building the case for PET evaluation and adoption.

Enterprises should regularly evaluate technical advances in data protection technologies and deploy suitable technical controls, thereby building the case for PET evaluation and adoption.

PETs can also support adherence to other foundational privacy principles, such as the OECD privacy principles,[68] which underpin several privacy laws and international privacy frameworks.

Figure 7 summarizes how PETs can be used to uphold data protection principles.[69]

FIGURE 7: PETs and Data Protection Principles

Data Protection Principle	Principle Description	Examples of Privacy Enhancing Technologies (PETs) Alignment
Purpose Limitation	Personal data collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes.	Trusted execution environments (TEEs) provide attestation mechanisms to remotely verify privacy request handling.
Data Minimization	Personal data should be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.	Federated learning (FL) eliminates centralized data collection and minimizes personal information processed during the model training phase. Zero-knowledge proof (ZKP) limits the amount of personal data required for processing. Secure multiparty computation (SMPC) enables collaboration without the need to share all the underlying raw data with all parties involved.
Storage Limitation	Personal data is kept in a form that permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.	Homomorphic encryption (HE), differential privacy (DP), and synthetic data may render the data "anonymous" such that the data is no longer subject to data protection compliance requirements and restrictions.
Accuracy	Personal data should be accurate and, where necessary, complete and kept up to date.	TEEs ensure data accuracy and consistency by limiting access to unencrypted data.
Security Safeguards (Integrity and Confidentiality)	Personal data should be protected by reasonable security safeguards against risk such as loss or unauthorized access, destruction, use, modification, or disclosure of data.	HE reduces the risk of data leakage by enabling computations on encrypted data without revealing the plaintext data. TEEs protect data against authorized access by storing data in a secured area. FL reduces the attack surface by eliminating the need for data transmission to centralized stores. ZKP enables secure processing while shielding the underlying data from the parties involved.
Accountability	The controller shall be responsible for, and be able to demonstrate compliance with, the data protection principles.	Depending on the implementation of PETs, they can be used as an instrument to demonstrate organizational accountability and supplement the privacy program in place. Organizations should ensure that any privacy claims made about the use of PETs must be accurate. For data governance, TEEs can provide evidence of the steps taken to mitigate risk[70] and enable the enterprise to demonstrate the accountability principle.

68 OECD Legal Instruments, *Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, Amended 7 October 2013, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188
69 Centre for Information Policy Leadership, "Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age," December 2023, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-understanding-pets-and-ppts-dec2023.pdf
70 *Op cit* Information Commissioner's Office UK 2022.

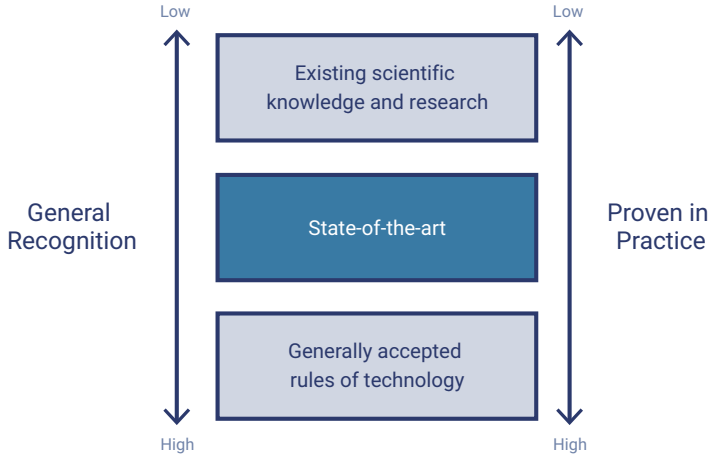
Legal Uncertainties Hindering PET Adoption and Remedial Strategies

The privacy and security capabilities offered by PETs align with the underlying principles of several regulatory frameworks. However, PETs are continually emerging and do not map explicitly to current laws. Also, PET is an umbrella term, and it is challenging to establish whether a particular PET or combination of them is legally acceptable for a specific use case, hindering adoption. The following are some challenges associated with PETs and regulatory requirements.

- A PET that may have received a favorable response from a particular regulator for a specific use case, such as the use of secure MPC for cross-border transfers,[71] may not be suitable for a similar use case where the data source or category of data is different. Each PET implementation needs to be evaluated for compliance on a case-by-case basis in the absence of prescriptive guidance.
- Where PET implementation involves multiple jurisdictions, different regulators may have varying opinions on the adequacy of a PET for a specific scenario. In addition to legal analysis for each representative jurisdiction, enterprises must evaluate PETs for cross-border data transfer law compliance including data sovereignty and localization issues.
- ENISA and TeleTrusT guidelines on state-of-the-art technical measures[72] define it as the "best performance available on the market to achieve...an IT security objective." **Figure 8** shows that a technical measure will start in the "existing scientific knowledge and research" stage at its inception. When it is introduced in the market and reaches market maturity, it shifts to the "state of the art" stage. Once the technical measure is recognized and adopted widely in the market, often described by corresponding standards, it is established as "generally accepted rules of technology." Measures that have been compromised or are no longer supported by manufactures lose recognition and should no longer be used in practice. PETs are yet to reach widespread adoption and prove themselves in practice for diverse scenarios. Additionally, standardization efforts for PETs are in progress, and maturity must be assessed independently. Therefore, there are uncertainties around establishing a particular PET as "state of the art" for a specific scenario.
- For collaborative PETs involving many participating enterprises, such as MPC or cross-silo federated learning, there may be instances where, depending on the level of involvement, all participants may classify as data controllers under regulations such as the GDPR[73] even if they never have access to personal data.
- The scope of data protection laws is limited to personal data as defined in the law; therefore, anonymized data is often out of scope for compliance. For example, GDPR does not apply to anonymized data.[74] However, the legal terms for anonymization and the degree of identifiability have not been normalized and uniformly defined, leading to uncertainty in leveraging PETs to render data anonymous. Under GDPR, there are arguments for HE being a technique of pseudonymization and anonymization.[75]

71 European Data Protection Board, *Recommendations 01/2020 on Measures that Supplement Transfer Tools to Ensure Compliance with the EU Level of Protection of Personal Data*, 10 November 2020, https://edpb.europa.eu/sites/default/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf
72 ENISA, "IT Security Act (Germany) and EU General Data Protection Regulation: Guideline 'State of the Art' Technical and Organizational Measures," TeleTrusT, 2021, https://www.teletrust.de/fileadmin/user_upload/2021-02_TeleTrusT-Guideline_State_of_the_art_in_IT_security_EN.pdf
73 Inverarity, C.; Kollnig, K.; "Modern PETs and Confidential Computing: No Way Out from GDPR Obligations," Open Data Institute, 8 September 2023, https://www.theodi.org/article/modern-pets-and-confidential-computing-no-way-out-from-gdpr-obligations/
74 Intersoft Consulting, "Recital 26: Not Applicable to Anonymous Data," https://gdpr-info.eu/recitals/no-26/
75 Koerner, K.; "Legal Perspectives on PETs: Homomorphic Encryption," Medium, 20 July 2021, https://medium.com/golden-data/legal-perspectives-on-pets-homomorphic-encryption-9ccfb9a334f

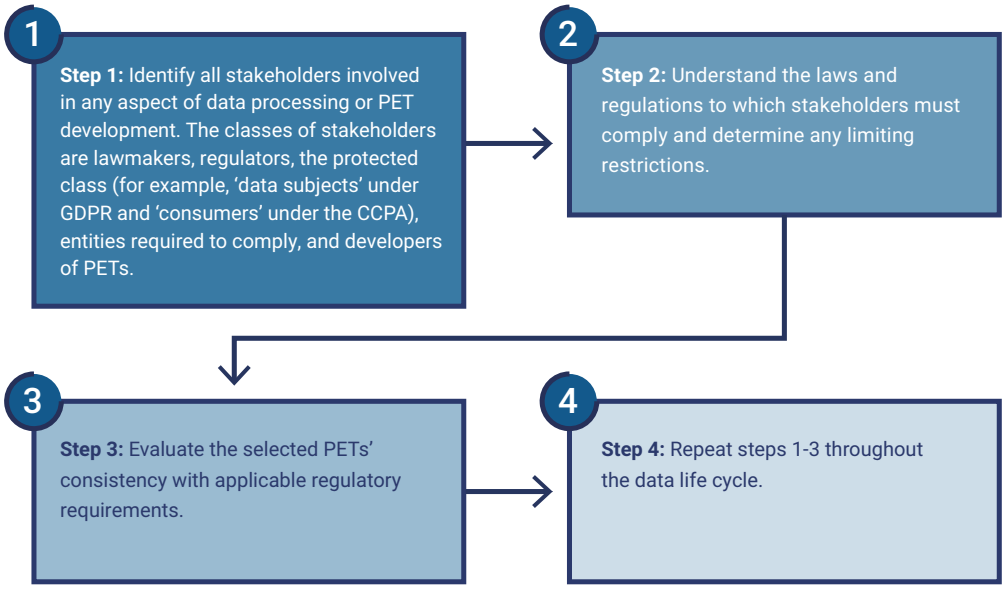
FIGURE 8: Three States of Technology Based on the Kalkar Decision



Source: ENISA, "IT Security Act (Germany) and EU General Data Protection Regulation: Guideline 'State of the Art' Technical and Organizational Measures," TeleTrusT, 2021, https://www.teletrust.de/fileadmin/user_upload/2021-09_TeleTrusT_Guideline_State_of_the_art_in_IT_security_EN.pdf

Enterprises must confirm that all stakeholders have fulfilled their legal obligations using a risk-based approach and assess whether implementing a PET in a specific use case complies with the regulation. The UN guide on PETs[76] recommends a four-step process for doing so, as shown in **figure 9**.

FIGURE 9: PET Selection and Compliance



Conclusion

PETs can play an essential role in a privacy-by-design approach to data governance when informed by appropriate legal guidance and privacy assurances. Beyond being leveraged as a compliance enabler, they support enterprises in forming new data partnerships and extracting maximum value from data

76 *Op cit* United Nations Committee of Experts on Big Data and Data Science for Official Statistics

while protecting individuals from privacy harms.

Commercial solutions and open-source libraries have helped reduce PET implementation costs, assisting in driving adoption for small and medium enterprises. PETs have been successfully used in production across various industries, including healthcare, finance, insurance, telecommunications, and law enforcement to mitigate privacy risk associated with data utilization.

With the rapid development and implementation of these technologies to solve real-world business problems, PETs are ushering in a new era of the ethical use of data.

The promise of widespread PET utilization is hindered by a lack of awareness and detailed guidance on using these technologies. Although several regulators and policymakers have initiated efforts to promote the innovation and adoption of PETs, it remains to be seen how the market will respond.

The promise of widespread PET utilization is hindered by a lack of awareness and detailed guidance on using these technologies.

Finally, PETs can enhance privacy and foster trust in the data economy but are not a substitute for a robust privacy program. Organizations considering the adoption of PETs still need to adhere to established data protection principles and view PETs as one part of the broader privacy framework, which may need to be supplemented by other technical and organizational measures.

Appendix: Case Studies

Privacy Enhancing Technologies as Business Enablers

The case studies in this section illustrate PETs' roles in real-world scenarios across industries. These case studies show that the purpose of PETs is not solely to safeguard individuals' privacy; instead, enterprises can also use them to enhance data partnerships and improve transparency.

the purpose of PETs is not solely to safeguard individuals' privacy; instead, enterprises can also use them to enhance data partnerships and improve transparency.

Context and business needs play a crucial role in determining the suitability and effectiveness of PETs. Therefore, the case studies listed should not be viewed as recommended solutions but as examples where PETs unlock otherwise prohibitive and high-risk data-use business opportunities.

Case Study 1: Insurance sector privacy-preserving predictive analytics using synthetic data

Source: Anonos[77]

AI and data science teams at a German public insurance enterprise encountered several data privacy challenges in leveraging customer data for predictive analytics. The sensitive personal nature of most customer data and restrictions on interdepartmental and external sharing posed considerable obstacles.

To initiate the use of this data, the data teams first underwent a privacy evaluation on a case-by-case basis, a process that often took several weeks. Anonymization methods, such as masking or k-anonymity, proved unsuitable because they compromised the usefulness of the data and failed to meet compliance requirements.

Additionally, navigating the complexities of corporate data sharing and usage systems was time-consuming, and the enterprise sought to expedite the time-to-data without making internal system modifications.

77 Anonos, "Predictive Analytics Help Insurers Make Most of Their Data," https://www.anonos.com/case-study/provinzial-synthetic-insurance-data

The optimal solution emerged in the form of synthetic data because it maintained the statistical value of the original data, thereby enhancing utility. The process of generating synthetic data completely severed the one-to-one relationships between original and synthetic records, minimizing the risk of re-identification.

The enterprise used synthetic data for a predictive analytics recommendation engine to identify the needs of more than a million customers, predicting their future service and product purchases. Establishing a data architecture with anonymized synthetic data eliminated the need for original data, accelerating the time-to-data.

The benefits and outcomes of leveraging synthetic data for predictive analytics were significant. They included a streamlined data usage approval process, achieving more than 80% usability of synthetic data while maintaining data anonymity; a 97% increase in performance effectiveness for machine learning models trained on synthetic data; and a reduction of four weeks in time-to-data without requiring adjustments to the internal data-sharing workflow.

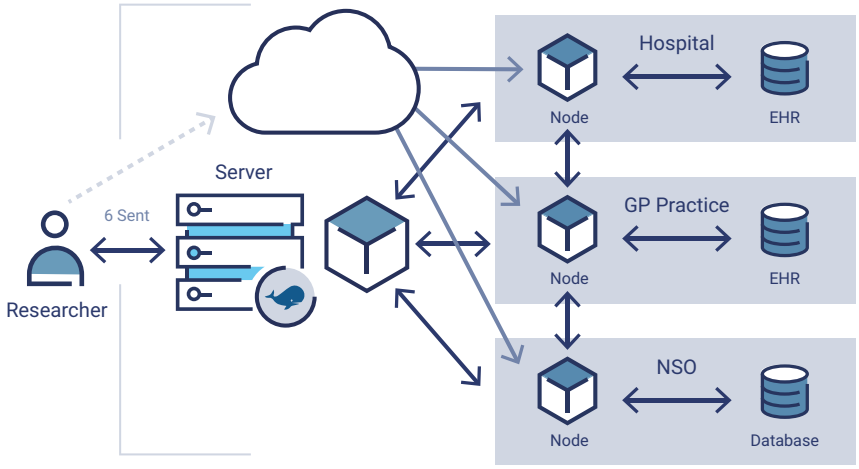
Case Study 2: Healthcare sector privacy-preserving cardiovascular risk prediction models

Source: UN Guide on Privacy-Enhancing Technologies for Official Statistics 2023[78]

The CARRIER (Coronary Artery Disease: Risk Estimations and Interventions for Prevention and Early Detection) project aims to detect and prevent coronary artery disease (CAD). Comprising clinicians, citizens, legal experts, and data scientists,[79] the project faces a significant challenge in linking datasets owned by different parties. The key concern in this process is the risk of re-identification of subjects, necessitating robust data security and privacy-preserving measures.

To address this challenge, the CARRIER solution employs a combination of several PETs, such as SMPC, HE, secret sharing, and FL. In this approach, the input and compute parties each run predefined code supplied via approved Docker images (refer to **figure 10**). Only images approved by the local party may be executed on the local data.

FIGURE 10: Architecture of Federated Learning Infrastructure



Source: Buckley, D.; "12. Statistics Netherlands: Developing Privacy-Preserving Cardiovascular Risk Prediction Models from Distributed Clinical and Socioeconomic Data," unstats.un, 9 February 2023, https://unstats.un.org/wiki/download/attachments/152797270/Screenshot%202023-02-09%20at%2013.32.29.png?version=1&modificationDate=1675949565217&api=v2

78 United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), "Statistics Netherlands: Developing privacy-preserving cardiovascular risk prediction models from distributed clinical and socioeconomic data," *UN Guide on Privacy-Enhancing Technologies for Official Statistics*, United Nations Statistics Division UN Statistics Wiki, 9 February 2023, https://unstats.un.org/wiki/display/UGTTOPPT/12.+Statistics+Netherlands%3A+Developing+privacy-preserving+cardiovascular+risk+prediction+models+from+distributed+clinical+and+socioeconomic+data

79 Scheenstra, B.; Bruninx, A.; et al.; "A Big Data-driven eHealth Approach to Prevent, Detect, and Reduce Atherosclerotic Cardiovascular Disease Burden," *European Journal of Preventive Cardiology*, Volume 29, Issue Supplement_1, May 2022, https://doi.org/10.1093/eurjpc/zwac056.305

This process is controlled via Vantage6, an open-source FL infrastructure. The participating parties can independently review the Docker images, and the executed transactions are kept in a central log. The final output is inspected manually for potential privacy leaks before release beyond the cooperating enterprises.

This project is in the proof-of-concept stage.

Case Study 3: Telecom sector secure collaboration to improve customer engagement using federated privacy-preserving analytics

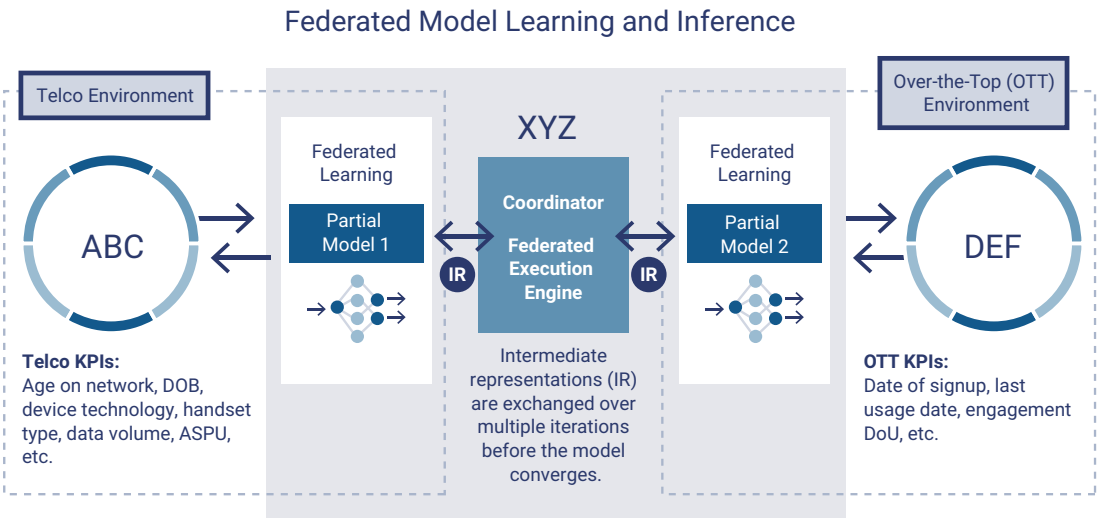
Source: Openmined[80]

A telecommunications enterprise in Africa with more than 272 million subscribers, MTN has a strategic partnership with Ayoba, a free messaging application with more than 5.5 million users. Ayoba aimed to understand its customer usage behavior to predict churn (defined as 30 days of continuous inactivity on the application) and mitigate inactivity by driving customer engagement. Global System for Mobile Communications (GSM) usage patterns of customers (available from MTN), when combined with Ayoba usage behavior, provide stronger indicators of Ayoba inactivity compared to either of these sources in isolation.

MTN and Ayoba leveraged federated privacy-preserving analytics (**figure 11**) to obtain customer insights while providing privacy protections for customer data. Vertical federated learning, implemented through split neural networks,[81] was employed to build the model on top of PySyft,[82] a privacy-preserving deep learning library. This solution enables training neural networks on vertically partitioned data features across multiple data owners, without requiring the movement of raw data from its owner's server. The identification of overlapping entities across different data owners is achieved through private set intersection (PSI), using encrypted IDs associated with the corresponding data points.

The privacy-preserving model predicted churn with a precision of 84.78% and recall of 82.64%, which was comparable to the predictive performance of the baseline shared-data mode with privacy protections.

FIGURE 11: Federated Privacy Preserving Framework Architecture



Source: OpenMined, "Case Study — Federated Privacy Preserving Analytics for Secure Collaboration Among Telco Partners to Improve Customer Engagement," 24 June 2022, https://blog.openmined.org/content/images/2022/06/image4.png

80 OpenMined, "Case Study — Federated Privacy Preserving Analytics for Secure Collaboration Among Telco Partners to Improve Customer Engagement," 24 June 2022, https://blog.openmined.org/federated-privacy-preserving-analytics-for-secure-collaboration-among-telco-and-partners-to-improve-customer-engagement/

81 Romanini, D.; Hall, A.; et al.; "PyVertical: A Vertical Federated Learning Framework for Multi-headed SplitNN," ARXIV, 14 April 2021, https://arxiv.org/abs/2104.00489

82 OpenMined/PySyft, https://github.com/OpenMined/PySyft

Case Study 4: Public sector and financial services confidential computing for cybercrime investigations

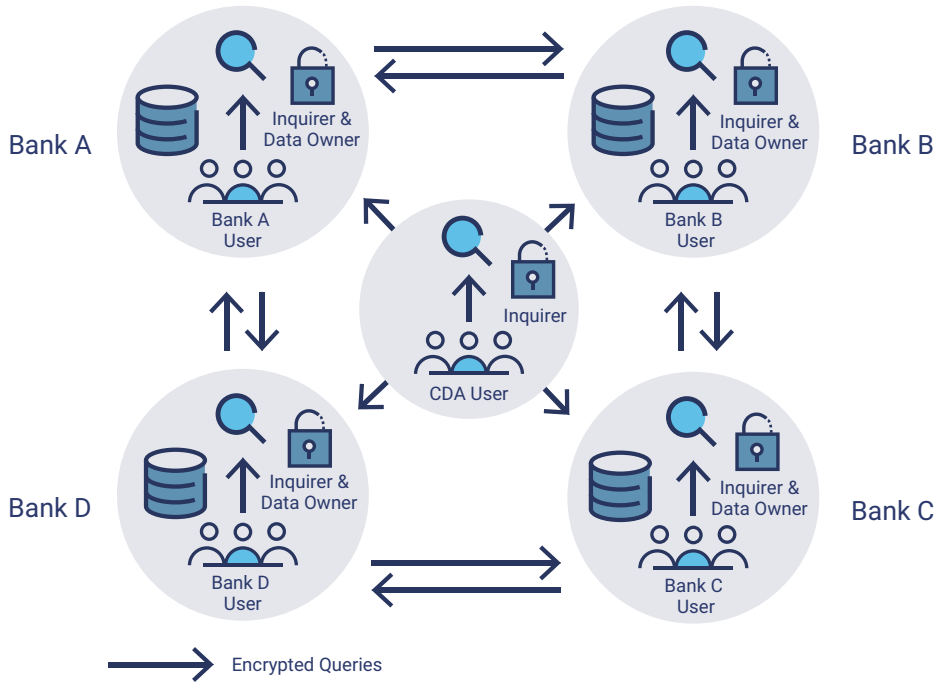
Source: World Economic Forum[83] *and Duality Technologies*

The Cyber Defence Alliance (CDA) is a UK-based non-profit public-private partnership that collaborates with the financial sector and law enforcement to share information to combat cybercrime. However, challenges emerge in collecting data essential for cybercrime investigation due to the inability to coordinate data sets with the correct requests among multiple parties in a timely manner. Criminals exploit these deficiencies to evade detection.

To address these challenges, as shown in **figure 12**, a consortium of four banks within the CDA and the UK's Metropolitan Police implemented a PET-enabled collaborative platform to improve their ability to identify fraud by interrogating each other's systems for suspicious cybercrime activity. Intelligence requirements and data sources were pre-agreed by all participants. This enabled the automatic exchange of data across participants' systems, saving time and resources for investigative teams. It ensured that sensitive search parameters remained encrypted during the process, thus always safeguarding the subjects of investigations.

The results revealed that confidential querying mitigates the risk of disclosure and regulatory breaches while also preventing insider tip-offs. Timely responses from partner banks enabled more efficient detection and deterrence of malicious activity. This meant that law enforcement could take proactive and timely action, ensuring the prevention, for example, of the further transfer of funds through a money-laundering network.

FIGURE 12: PET and Joint Investigations



Source: Adapted from World Economic Forum (WEF), "Cyber Information Sharing: Building Collective Security," October 2020, https://www3.weforum.org/docs/WEF_Cyber_Information_Sharing_2020.pdf

83 World Economic Forum (WEF), "Cyber Information Sharing: Building Collective Security," October 2020, https://www3.weforum.org/docs/WEF_Cyber_Information_Sharing_2020.pdf

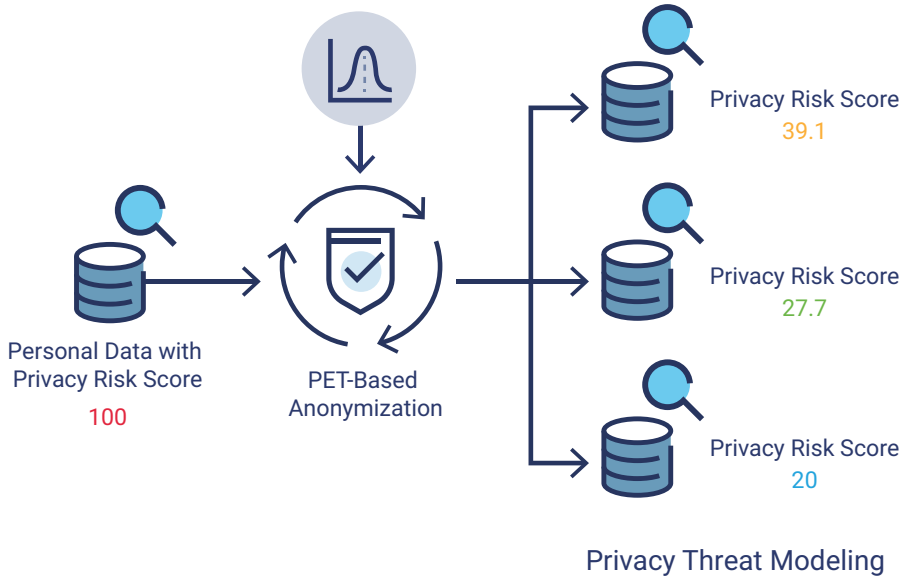
Case Study 5: Auditable data analytics based on privacy threat modeling for the automotive industry

Automotive enterprises have access to sensitive customer data[84] such as driving behavior, geolocation data, vehicle telematics, connected services data, inferred data, etc. This information must be protected in compliance with various privacy requirements. A large Asian automobile manufacturer aimed to conduct predictive analytics on vehicle resale value using sensitive data and wanted to share this data with partners collaboratively to gain data insights while preserving privacy. The enterprise also aimed to identify and quantify the privacy risk associated with using customer data in model training.

To address these challenges, the proposed solution, depicted in **figure 13**, involved a PET-based anonymization and risk scoring process. The approach simulated privacy attacks, such as singling-out attacks, linkage attacks, and outlier attacks, on a sampling of data sets from the source to perform attribute-level risk scoring for various data flows. Privacy threat modeling frameworks such as LINDDUN[85] and regulatory requirements served as the basis for employing a combination of PETs such as differential privacy, k-anonymity, and t-closeness. These technologies were used to mitigate the privacy threats identified. After the application of PETs, the residual risk scores were calculated, offering assurance and enabling the auditable application of mitigation techniques.

The solution helped unlock data usage while ensuring compliance with privacy regulatory requirements across Asia and the EU. It helped develop a mathematically grounded DPIA process that was repeatable and shareable. Data governance for customer data was automated and streamlined, making data pipelines privacy-aware and leveraging the predefined mapping between privacy risk scores and approved levels of data sharing within and outside the enterprise. The auditable nature of the PETs implementation provided assurance regarding appropriate configuration and supported fine-tuning the output by balancing privacy and utility parameters.

FIGURE 13: PET-Based Anonymization and Risk Scoring Process



Source: PrivaSapien. Reprinted with permission.

84 Caltrider, J.; Rykov, M.; et al.; "It's Official: Cars Are the Worst Product Category We Have Ever Reviewed for Privacy," Mozilla, 6 September 2023, https://foundation.mozilla.org/en/privacynotincluded/articles/its-official-cars-are-the-worst-product-category-we-have-ever-reviewed-for-privacy/
85 Linddun, "A Framework for Privacy Threat Modeling," https://linddun.org/

Acknowledgments

Lead Developer

Nandita Rao Narla

CISA, CISM, CRISC, CDPSE, CIPM, CIPP/US, CIPT, FIP

USA

Expert Reviewers

Chetan Anand

CDPSE, Agile Scrum Master, CCIO, CPEW, CPISI, FPT, ICBIS, ICCP, ICOSA, IRAM2, ISO 22301 LA, ISO 27001 LA, ISO 27701, ISO 31000, ISO 9001 LA, Lean Six Sigma Green Belt, SQAM

Profinch Solutions, India

Angus Ang

Citibank, Singapore

Timo Huebner

CISM, CGEIT, CDPSE

Germany

Katharina Koerner, Ph.D.

Tech Diplomacy Network, USA

Carol Lee

CDPSE, CISM, CRISC, C|CISO, CEH, CIPM, CSP, CSSLP

Hang Lung Group, Hong Kong

Anshu Singh

CIPT

Government Technology Agency (GovTech), Singapore

Tammy White

CISM, CRISC, CISSP

PricewaterhouseCoopers (PwC), USA

Board of Directors

John De Santis, Chair

Former Chairman and Chief Executive Officer, HyTrust, Inc., USA

Brennan P. Baybeck, Vice-Chair

CISA, CISM, CRISC, CISSP

Senior Vice President and Chief Information Security Officer for Customer Services, Oracle Corporation, USA

Stephen Gilfus

Managing Director, Oversight Ventures LLC, Chairman, Gilfus Education Group and Founder, Blackboard Inc., USA

Niel Harper

CISA, CRISC, CDPSE, CISSP, NACD.DC

Former Chief Information Security Officer, United Nations Office for Project Services (UNOPS), USA

Gabriela Hernandez-Cardoso

NACD.DC

Independent Board Member, Mexico

Jason Lau

CISA, CISM, CGEIT, CRISC, CDPSE, CIPM, CIPP/E, CIPT, CISSP, FIP, HCISPP

Chief Information Security Officer, Crypto.com, Singapore

Massimo Migliuolo

Independent Director, Former Chief Executive Officer and Executive Director, VADS Berhad Telekom, Malaysia

Maureen O'Connell

NACD.DC

Board Chair, Acacia Research (NASDAQ), Former Chief Financial Officer and Chief Administration Officer, Scholastic, Inc., USA

Erik Prusch

Chief Executive Officer, ISACA, USA

Asaf Weisberg

CISA, CISM, CGEIT, CRISC, CDPSE, CSX-P

Chief Executive Officer, introSight Ltd., Israel

Pamela Nigro

ISACA Board Chair 2022-2023

CISA, CGEIT, CRISC, CDPSE, CRMA

Vice President, Security, Medecision, USA

Gregory Touhill

ISACA Board Chair 2021-2022

CISM, CISSP

Director of the CERT Division at Carnegie Mellon University's Software Engineering Institute, USA

Tracey Dedrick

ISACA Board Chair, 2020-2021

Former Chief Risk Officer, Hudson City Bancorp, USA

About ISACA

ISACA® (www.isaca.org) is a global community advancing individuals and organizations in their pursuit of digital trust. For more than 50 years, ISACA has equipped individuals and enterprises with the knowledge, credentials, education, training and community to progress their careers, transform their organizations, and build a more trusted and ethical digital world. ISACA is a global professional association and learning organization that leverages the expertise of its 180,000+ members who work in digital trust fields such as information security, governance, assurance, risk, privacy and quality. It has a presence in 188 countries, including 225 chapters worldwide. Through the ISACA Foundation, ISACA supports IT education and career pathways for underresourced and underrepresented populations.

DISCLAIMER

ISACA has designed and created *Exploring Practical Considerations and Applications for Privacy Enhancing Technologies* (the "Work") primarily as an educational resource for professionals. ISACA makes no claim that use of any of the Work will assure a successful outcome. The Work should not be considered inclusive of all proper information, procedures and tests or exclusive of other information, procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific information, procedure or test, professionals should apply their own professional judgment to the specific circumstances presented by the particular systems or information technology environment.

RESERVATION OF RIGHTS

ISACA
1700 E. Golf Road, Suite 400
Schaumburg, IL 60173, USA

Phone: +1.847.660.5505

Fax: +1.847.253.1755

Support: support.isaca.org

Website: www.isaca.org

Participate in the ISACA Online Forums:
https://engage.isaca.org/onlineforums

X: www.x.com/ISACANews

LinkedIn:
www.linkedin.com/company/isaca

Facebook:
www.facebook.com/ISACAGlobal

Instagram:
www.instagram.com/isacanews/