# PRIVACY ENHANCING TECHNOLOGIES ADOPTION GUIDE

Developed by:

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Supported by:

pdpc PERSONAL DATA PROTECTION COMMISSION SINGAPORE

# CONTENTS

# EXECUTIVE SUMMARY

Privacy Enhancing Technologies (PETs) are rapidly evolving from niche technologies into strategic enablers of business growth and innovation. Traditionally, PETs were deployed to mitigate data protection risks and meet regulatory requirements. Today, however, PETs can **empower organisations to unlock new value through the trusted use and sharing of their data**.

This PETs adoption guide is **intended for senior decision-makers considering the adoption of PETs**.

This guide outlines how leading organisations across different industries are leveraging PETs to:

**Enable Artificial Intelligence (AI) and Machine Learning (ML) development** by providing safe access to more data for AI training; and

**Forge multi-stakeholder collaborations (e.g. across teams within organisations or across sectors)** by allowing more data to be used by different parties without compromising data confidentiality or privacy.

With more organisations now involved in real-world PETs deployments that have demonstrated clear value, there is **growing momentum in the PETs industry** to address the varying demands of organisations. In addition to enterprise-grade solutions offered by technology companies, some **PETs are also becoming more cost-effective and easier to use** due to the increasing availability of open-source software development tools and low- or no-code platforms.

**The Infocomm Media Development Authority (IMDA) aims to better support organisations in adopting PETs by addressing the challenges and barriers they face during evaluation and implementation.**

This proposed guide will be accompanied by two practical resources to help organisations identify PETs and processes that are relevant to their use case:

**PETs Discovery Tool** – helping users match their use case to proven PETs solutions, supported by real-world case studies that illustrate impact; and

**Implementation Checklist** – outlining key areas of consideration spanning pre-, during, and post-implementation stages to ensure secure and effective PETs deployment.

This guide and its accompanying resources have been developed using insights from real-world use cases in IMDA's PET Sandbox. These resources will continue to be refined as more organisations participate in IMDA's PET Sandbox, and as we incorporate feedback from industry and technology experts.

As organisational needs grow and technologies evolve, PETs are no longer just about meeting compliance requirements. **They are becoming strategic tools that offer organisations a competitive advantage in a data-driven economy.**

# READER'S GUIDE

| | PETs Adoption Guide | PETs Discovery Tool | Implementation Checklist |
|---|---|---|---|
| Objectives | To help potential adopters better understand how PETs can address business challenges related to data sharing, illustrated though real-world use cases. | To explore similar implementations and understand how different types of PETs have been deployed, including potential pitfalls. | To provide stakeholders with key tasks across three areas: business, technology, and data protection. |
| Intended Audience | Senior management seeking to understand the value of PETs for their organisation. | Business, product, and data protection stakeholders who need to understand how PETs operate and how they safeguard data while supporting business needs. | Teams responsible for evaluating, deploying, and ensuring the security of PETs solutions within their organisation, including in business, technology, and information security. |
| How to Use This Guide | Use the points laid out in this document to build or strengthen your company's internal business case by highlighting broad industry trends in PETs and explaining why they should be considered. | Use the reference implementations provided to offer tangible proof points that can help secure internal buy-in by demonstrating the benefits of PETs and how they enable data use and compliance. | Use this checklist as a "hygiene check" to ensure that key considerations have been addressed when companies are implementing PETs solutions. |

# BENEFITS OF PETs

# BENEFITS

## 1. PETs HAVE BEEN USEFUL IN THE CONTEXT OF REGULATORY COMPLIANCE

### 1.1 Enabling Privacy-preserving Data Use:

PETs enable the processing, analysis, and extraction of insights from data without revealing underlying personal or commercially sensitive information. PETs include technologies that obfuscate or conceal the original data, such as homomorphic encryption, and technologies that facilitate the flow of insights without transferring the data itself, for example, through federated learning.

Please see Annex 1 for a technology primer on the types of PETs and their common applications.

### 1.2 Established Use for Compliance:

PETs have existed for some time and have primarily been used to address privacy and data protection risks during data sharing and collaboration. For instance, in a 2021 recommendation on measures to supplement transfer tools, the European Data Protection Board cited PETs as a useful "supplementary tool" for ensuring compliance in international data transfers.

Beyond Europe, other data protection regulators around the world, such as the Office of Privacy Commissioner of Canada (OPC) and Korea's Personal Data Protection Commissioner (PIPC), have also recognised the value of certain PETs — for example, in the generation of synthetic data — in addressing privacy and data protection risks.[1]

### 1.3 Regulatory Guidance in Singapore:

The Personal Data Protection Commission (PDPC) of Singapore has issued regulatory guidance to organisations in the IMDA PET Sandbox, highlighting how the use of PETs can help them meet regulatory obligations.

[1] *For example, the OPC has highlighted in a publication on synthetic data its benefits in mitigating traditional re-identification attacks and enabling the automation of the de-identification process. Similarly, Korea's PIPC and Singapore's PDPC have issued guidance on the use of synthetic data to address privacy and data protection risks.*

**Example of use cases in the IMDA PET Sandbox include[2]:**

### 1.3.1 Using PETs to process personal data without requiring consent when exceptions apply.

**Use Case:**

- Kajima wanted to use personal data collected from building inhabitants for secondary purposes. Normally, individual consent would be required, but it may be unfeasible or impractical to seek consent from every single individual.
- Kajima, therefore, explored the use of synthetic data as a privacy-centric solution to enable the collected data to be used more freely.

**Assessment:**

- Personal data was used to generate synthetic data for research analysis. It was assessed that a Business Improvement Exception (BIE) could be relied upon if (a) the purpose for using the personal data cannot be reasonably achieved without using it in an individually identifiable form; and (b) that a reasonable person would consider such use of personal data appropriate.
- Since the PET in the Kajima use case is used to better understand customer profiles and preferences to enhance service offerings, BIE could be applied for that category of use.

### 1.3.2 Using PETs to sufficiently reduce the risk of re-identification such that data sharing does not constitute a disclosure of personal data.

**Use Case:**

- TikTok sought to measure advertisement attribution privately, without the advertiser and publisher accessing each other's raw inputs.

**Assessment:**

- A multi-phase process, including differential privacy, homomorphic encryption, and secure multi-party computing, was applied to the datasets to be shared between parties (i.e. the publisher and the advertiser). With these safeguards in place, it was assessed that there was no serious possibility of re-identification by either the data recipient or third parties.

[2] *These examples summarise key points from IMDA and PDPC's published case studies and practical guidance documents. For complete details and context, please refer to the original documents available on IMDA and PDPC's websites respectively.*

**1.3.3** **Using PETs can lower the risks of re-identification such that data is considered anonymised.**

**Use Case:**

- Grab wanted to make more data available for analysis.
- This was done by automating data tagging, labelling, and subsequent anonymisation via a large language model (LLM)-based solution, if the data was considered personal data.

**Assessment:**

- Grab had in place multi-layered data protection practices — including the removal or alteration of direct and indirect identifiers, the implementation of organisational, technological, and governance safeguards, and periodic reviews to assess adequacy of techniques. Where both direct and indirect identifiers in a data record are removed, the risk of re-identification is low, and the data record can be considered anonymised.

# 2. PETs ARE BEING USED TO UNLOCK BUSINESS OPPORTUNITIES

In recent years, the use of PETs has evolved beyond risk mitigation and compliance. Increasingly, they are being used to unlock new value in business and innovation as organisations adopt PETs that allow them to use and harness data more effectively within their organisations or with external partners.

These developments are observed across two areas:

### 2.1 Access to data to enable AI/ML training:

- Finetuning or refining a model to improve its performance requires more contextualised data, which is rarely available in the public domain.
- While organisations have been cautious about sharing their data for ML purposes, the emergence of PETs has gradually enabled privacy-preserving ML, ensuring that sensitive data remains protected.

**Examples of use cases include:**

### 2.1.1 Training AI models to predict customer's promotion-to-conversion likelihood.

**Use Case:**

- Ant International[3] wanted to **train a ML model that could better predict each customer's promotion-to-conversion** likelihood by using historical data on common customers from its partner.

**Solution:**

- **Multi-party Computing Private Set Intersection (MPC-PSI) and Federated Learning (FL)** were used to privately identify common customers and enable model training in a decentralised manner, ensuring sensitive data was not shared between either party.

### 2.1.2 Training diagnostic imaging AI models.

**Use Case:**

- Healthcare and tertiary learning institutions seek to **collaborate to train and improve ML models** without revealing the identity of individual patients. However, their datasets typically contain sensitive personal information.

**Solution:**

- The institutions leveraged the NVIDIA Clara platform[4] which uses **FL** to support collaborative AI development, including improvements to diagnostic imaging models. This approach safeguards patient data by enabling joint model training data without allowing institutions to access one another's raw data.

[3] *IMDA PET Sandbox Use Case.*

[4] *NVIDIA Clara Federated Learning to Deliver AI to Hospitals While Protecting Patient Data, Dec 2019.*

### 2.2 Extracting insights from data generated by multiple stakeholders (e.g. across teams within organisations or across sectors):

- Different parties can now collaborate on data-driven projects, including partnerships that were previously impossible or impractical due to confidentiality or privacy concerns.
- For instance, PETs allow organisations to work with anonymised datasets at a low risk of re-identifiability, without requiring consent from millions of existing customers for such collaborations.

**Examples of use cases include:**

#### 2.2.1 Collaboration across multiple entities on financial-crime intelligence.

**Use Case:**

- Mastercard sought to share International Bank Account Numbers (IBAN) with multiple overseas banks to identify suspicious accounts for anti-money laundering purposes. However, doing so could constitute a disclosure of commercially sensitive information (i.e. information about Mastercard's customers).

**Solution:**

- For this proof of concept (POC), Mastercard used **Fully Homomorphic Encryption (FHE)**, to encrypt sensitive data in the query sent to participating banks.
- Each bank performed data matching on the encrypted query using its own customer information, but could not view Mastercard's sensitive data and was unable to determine which of its records, if any, matched the query.

#### 2.2.2 Partnership to share first-party data for targeted advertising.

**Use Case:**

- To enable advertisers to serve relevant and targeted advertisements on its digital network, Singapore Press Holdings (SPH) Media explored the use of PETs to profile and amplify an advertiser's existing customer base without the collection, sharing or disclosure of any personal data between organisations.

**Solution:**

- **Trusted Execution Environments (TEEs)** were used to generate an audience list of SPH media users who were most similar to the advertising partner's existing customer base. This was done within a highly secure environment that protects the data of individuals.

# 3. PETs ARE BECOMING INCREASINGLY ACCESSIBLE AND EASIER TO USE

**3.1** With growing demand from organisations to leverage PETs, some solutions are becoming more productised. Large technology companies are now offering PETs to their enterprise customers. Some examples include:

- TEEs – Azure's Confidential Computing allows secure data processing in sensitive environments.
- Differential Privacy – Google's BigQuery has an interface that allows users to apply Differential Privacy to datasets.

**3.2** To enable broader access to PETs, some organisations have also open-sourced their tools, making Application Programming Interfaces (APIs),[5] Software Development Kits (SDKs)[6] and even no-code or low-code[7] options more readily available. These open-sourced tools are typically more cost-effective and enable organisations to explore the use of PETs with greater ease. Some examples of open-sourced tools include:

- Microsoft Simple Encrypted Arithmetic Library (SEAL), a homomoprhic encryption library.
- Google's Tensorflow Privacy, which implements differential privacy for ML, Privacy Join and Compute, a Multi-party Computing (MPC) tool based on Private Set Intersection (PSI) and PrivMRF, an open-sourced data synthesis tool.

[5] *The Application Programming Interface is a set of rules and specifications that allows for different software systems to communicate with each other.*

[6] *The Software Development Kit is a collection of tools, libraries, documentations, or pre-built code examples, that help developers build applications for a specific platform or with specific functionalities.*

[7] *Examples of no- or low- code solutions include: Fortanix Armor which offers FHE enabled AI tools with no-code deployment interfaces, and Amazon Web Services (AWS) Nitro Enclaves which offers isolated compute environments for multiple parties to collaborate on data processing without exposing their sensitive data to each other.*

# RESOURCES TO GUIDE PETs DISCOVERY AND IMPLEMENTATION

# RESOURCES TO GUIDE PETs DISCOVERY AND IMPLEMENTATION

1. Despite the success of early adopters in implementing PETs for their specific use cases, **identifying and implementing the right PETs can be a complex process for organisations**. This is because different PETs could be used to address a particular use case and the optimal solution typically depends on multiple considerations.

   These considerations include business needs, technology specifications, and various regulatory requirements.

   The scenarios outlined below illustrate how different data protection needs could influence the PETs used:

| | Scenario 1 | Scenario 2 |
|---|---|---|
| **Scenario** | Two entities want to conduct a combined analysis of their private datasets in a safe manner, **without either party gaining access to the other's raw data**. | Two entities want to analyse their private datasets while **processing their data locally, without sharing their datasets with each other**. |
| **Potentially Applicable PETs** | PETs are **applicable to both** use cases:<br>✓ TEE — Both entities can each upload their datasets in a secure environment for protected analysis.<br>✓ MPC — Two or more parties can jointly analyse data without any party having to reveal their full dataset. | |
| **MPC Consideration** | **MPC may not be as applicable**, considering greater computational or encryption overheads would likely cost more. | **MPC may be more appropriate** because it enables collaborative analysis without requiring either party to share or upload their datasets. |
| **TEE Consideration** | Compared to the use of MPC, **TEE may be more appropriate** because TEE can ensure that both parties can analyse data without seeing each other's raw data. | **TEE may not be as applicable** because the use of this PET typically includes the sharing of data. |

   In the real world, there may be other considerations to take note of such as cost, time, regulatory jurisdictions or limitations to the state of technology, which may add further layers of complexity.

2. To assist organisations in evaluating and determining the best PETs solution to address their problem statement, and to support them in implementing the PETs solution, this guide offers the following two resources.

| | |
|---|---|
| **PETs Discovery Tool**  | The discovery tool aims to help users identify relevant PETs for their use case, supported by relevant implementation examples.<br><br>Users can discover use cases through three inputs:<br><br>1. Data-sharing objectives (i.e. what you intend to achieve in sharing data)<br>2. The industry sector (e.g. finance)<br>3. Specific use cases (e.g. anti-money laundering)<br><br>Based on these inputs, the tool presents relevant case studies, including concise information about the scenario, PETs solution applied, its objectives, and outcomes. |
| **Implementation Checklist**  | The checklist provides guidance throughout an organisation's PETs deployment journey across three phases:<br><br>1. **Pre-implementation**: Ensures a proper assessment of business needs, as well as an evaluation of PETs solution providers and the proposed solution.<br>2. **During Implementation**: Ensures iterative testing and validation of the PETs solution, alongside effective data protection and compliance measures.<br>3. **Post-implementation**: Supports ongoing performance monitoring, scalability considerations, and ongoing regulatory compliance.<br><br>Please see Annex 2 for the detailed implementation checklist. |

# CONCLUSION

Real-world implementations have shown that many PETs have progressed beyond the experimental stage. No longer used solely as a tool for compliance, organisations are adopting PETs to support practical, high-impact use cases.

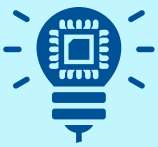
As organisations look to unlock greater value through the trusted use of data, PETs and the landscape of technology providers will continue to evolve and mature. This makes it an opportune time for organisations to begin seriously exploring how PETs might fit in their data strategies, as well as to begin experimenting with and implementing them in products or operations.

This Adoption Guide is designed to help organisations kick-start their journey in using these technologies. As a "living document", IMDA welcomes industry and expert feedback so that we can continually refine and improve this guide, ensuring its continued relevance to real-world use cases.

Finally, although PETs provide valuable supplementary safeguards, they are not intended to serve as a standalone solution. PETs should be integrated alongside existing processes to reinforce data protection and, in parallel, enable business value.

# ANNEX 1 – TYPES OF PETs AND ITS APPLICATIONS

# ANNEX 1 – TYPES OF PETs AND ITS APPLICATIONS

| Categories of Privacy Enhancing Technologies (PETs) | PETs | Examples of Applications (Non-exhaustive) |
|---|---|---|
| **Technologies that Obfuscate or Hide the Original Data**  | **Differential Privacy** | ▶ Data Release<br>▶ Data Collection<br>▶ Model Training, Validation and Deployment |
| | **Homomorphic Encryption** | ▶ Secure Analytics<br>▶ Model Training, Validation and Deployment<br>▶ Verifiable Secure Computations |
| | **Synthetic Data** | ▶ Secure Analytics<br>▶ Model Training, Validation and Deployment<br>▶ Software Testing |
| **Technologies that Facilitate the Flow of Insights**  | **Federated Learning** | ▶ Cross-silo Learning<br>▶ Cross-device Learning |
| | **Secure Multi-party Computing** | ▶ Combining Multi-source Data<br>▶ Detecting and Matching with Private Set Intersection<br>▶ Model Training, Validation and Deployment |
| | **Trusted Execution Environments** | ▶ Identification and Authentication<br>▶ Secure Analytics<br>▶ Model Training, Validation and Deployment |
| | **Zero-knowledge Proofs** | ▶ Identity and Age Verification<br>▶ Asset Ownership Verification<br>▶ Cryptocurrency Transaction |

# TECHNOLOGIES THAT OBFUSCATE OR HIDE THE ORIGINAL DATA

## i. Differential Privacy

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | | ▶ Differential privacy is a mathematical framework that provides formal privacy guarantees by adding controlled random noise to data or computations.<br>▶ It ensures that the presence or absence of an individual's data in a dataset cannot be inferred from the analysis results, while still maintaining useful statistical accuracy. |
| **Privacy Budget (Epsilon ε)** | | ▶ The privacy budget, denoted as epsilon (**ε**), controls the trade-off between privacy protection and data utility.<br>▶ A lower **ε** value offers stronger privacy guarantees but reduces accuracy.<br>▶ The privacy budget is cumulative, meaning it gets consumed across multiple queries or releases. |
| **Types of Implementations** | Global Differential Privacy | ▶ A trusted aggregator collects raw data and adds noise to the results before sharing them.<br>▶ Noise is added only once at the aggregate level. |
| | Local Differential Privacy | ▶ Individuals add noise to their data before sharing it with anyone else.<br>▶ This approach is suited for large-scale data collection where individual accuracy is less critical than aggregated insights. |

Common Applications

| Use Case | Description |
|---|---|
| **Data Release** | Publishing aggregate datasets while protecting individual privacy, such as:<br>▶ National census data;<br>▶ Public health statistics on disease prevalence;<br>▶ Research datasets with the academic community; and<br>▶ Geographic data on population movements or service usage patterns while preserving anonymity. |
| **Data Collection** | Gathering insights about user behaviour and system performance, such as:<br>▶ Device performance metrics and crash reports collected by operating systems;<br>▶ Frequently visited website information used by browsers to improve caching; and<br>▶ Usage patterns gathered by mobile applications to enhance user experience. |
| **Model Training** | Developing Artificial Intelligence (AI) systems that learn from sensitive personal information such as:<br>▶ Training content-safety classifiers using differentially private synthetic data; and<br>▶ Fine-tuning language models in ways that prevent memorisation of sensitive training examples. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation (SDO)** | National Institute of Standards and Technology (NIST) Special Publication (SP) 800-226 Guidelines for Evaluating Differential Privacy Guarantees |
| **Infocomm Media Development Authority (IMDA) PET Sandbox Case Study** | ▶ Meta – Digital Advertising in a Paradigm Without 3rd Party Cookies<br>▶ TikTok – Privacy Preserving Attribution Measurement |

## ii. Homomorphic Encryption

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | Homomorphic Encryption (HE) is a family of encryption schemes that enable computations to be performed directly on encrypted data without ever decrypting it. | |
| **Types of Homomorphic Encryption** | Fully HE (FHE) | FHE:<br>▶ is the most comprehensive type of HE;<br>▶ supports arbitrary computations on encrypted data; and<br>▶ is highly flexible but computationally expensive for complex operations. |
| | Somewhat HE (SHE) | SHE:<br>▶ supports general computation but only up to a defined limit. |
| | Partial HE (PHE) | PHE:<br>▶ supports only one operation (addition or multiplication). |

Common Applications

| Use Case | Description |
|---|---|
| **Secure Analytics** | Processing confidential information while keeping it encrypted throughout the analysis pipeline, such as:<br>▶ Tumour detection in Magnetic Resonance Imaging scans performed by third-party service providers. |
| **Model Training, Validation and Deployment** | Training, validating and running models on encrypted data without revealing the underlying information, such as:<br>▶ Cloud-based Machine Learning (ML) model training using encrypted datasets. |
| **Verifiable Secure Computations** | Performing computations on independently verifiable results while maintaining data confidentiality, such as:<br>▶ Encrypted vote counting, ballot tracking and checking systems. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | The International Organisation for Standardisation (ISO) and the International Electrotechnical Commission (IEC) 18033-6:2019 IT Security Techniques – Encryption Algorithms – Part 6: Homomorphic Encryption<br>ISO/IEC Committee Draft (CD) 28033-1 Information Security – Fully Homomorphic – Part 1: General <In Progress> |
| **IMDA PET Sandbox Case Study** | ▶ Mastercard – Secure Sharing of Financial Crime Intelligence<br>▶ Ant International – Enhancing Customer Engagement with Privacy Preserving AI |

## iii. Synthetic Data

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | Synthetic data refers to artificially generated data generated using a purpose-built mathematical model, including AI/ML models or algorithms. | |
| **Types of Synthetic Data[8]** | Fully Synthetic | ▶ Entirely artificial with no direct link to real-world data.<br>▶ Offers the strongest privacy protection. |
| | Partially Synthetic | ▶ Retains most original information but replaces sensitive attributes with synthetic values.<br>▶ As not all original values are removed, the risk of re-identification is greater than that of a fully synthetic dataset. |

[8] *Types of synthetic data are referenced from Centre for Information Policy Leadership (CIPL)'s Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, December 2023.*

Common Applications

| Use Case | Description |
|---|---|
| **Model Training, Validation and Deployment** | Generating large synthetic datasets for training, testing and augmenting AI/ML models, especially when real data is limited, such as:<br>▶ Synthetic data employed for training AI for self-driving car systems like 'Waymo'. |
| **Secure Analytics** | Analysing trends and patterns without exposing sensitive real-world records, such as:<br>▶ Sharing high-quality AI generated synthetic datasets for healthcare research. |
| **Software Testing** | Testing systems without using real personal data, reducing the risk of data breaches in development environments, such as:<br>▶ Using synthetic data for enterprise application systems or product testing. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | ISO/IEC AWI TR 42103 – Informational Technology – Artificial Intelligence – Overview of Synthetic Data in the Context of AI Systems (Under Development) |
| **Personal Data Protection Commission of Singapore (PDPC) Guide** | Proposed Guide to Synthetic Data Generation |
| **IMDA PET Sandbox Case Study** | Kajima – Generating Synthetic Data for Analysis and Research |

# TECHNOLOGIES THAT FACILITATE THE FLOW OF INSIGHTS

## i. Federated Learning

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | ▶ Federated learning (FL) is an architectural PET that enables multiple parties to train models on their own data via local models.<br>▶ 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.<br>▶ FL localises control of both data and the models operating on that data. | |
| **Architectural Approaches** | Centralised FL | Uses a coordination server to manage model distribution and updates. |
| | Decentralised FL | Enables direct communication between participating entities without a central server. |
| **Types of Data Distribution** | Horizontal FL | Parties share similar features across different user groups. |
| | Vertical FL | Parties hold different features about the same users. |

Common Applications[9]

| Use Case | Description |
|---|---|
| **Cross-silo Learning** | Enables multiple organisations to collaboratively improve model performance while retaining control of their own data, such as:<br>▶ Life insurance organisations jointly training fraud-detection models using synthetic data generated from real-world data. |
| **Cross-device Learning** | Supports the training of AI models directly on user devices, while keeping user interaction data local, such as:<br>▶ Voice recognition;<br>▶ Text prediction; and<br>▶ Personalised on-device AI interactions. |

[9] *FL is typically used together with other PETs. While FL keeps raw data local, it does not guarantee complete privacy on its own, as model parameters exchanged during training can still leak sensitive information through certain types of attacks.*

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | IEEE 3652.1-2020 Guide for Architectural Framework and Application of Federated Machine Learning |
| **IMDA PET Sandbox Case Study** | Ant International – Enhancing Customer Engagement with Privacy Preserving AI |

## ii. Secure Multi-party Computing

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | | ▶ Secure Multi-party Computing (SMPC) is a cryptographic technique that allows multiple parties to jointly compute and analyse combined data without revealing their individual inputs.<br>▶ Only the final agreed-upon output is disclosed to participants. |
| **Types of Method** | Secure Aggregation Protocol | ▶ Supports computation across data from more than two parties.<br>▶ The computation happens on encrypted data, and only the final aggregate result is decrypted. |
| | Private Set Intersection (PSI) | Enables two parties to identify matching elements in their datasets without revealing any non-matching data. |

Common Applications

| Use Case | Description |
|---|---|
| **Combining Multi-source Data** | Allows multiple parties to combine and analyse their collective data, such as:<br>▶ Government agencies analysing citizen data across social services; and<br>▶ Banks merging transaction and demographic data. |
| **Detecting and Matching with PSI** | Helps two parties identify matching elements between datasets, such as<br>▶ Checking breached passwords across databases; and<br>▶ Finding overlapping population segments between organisations. |
| **Model Training, Validation and Deployment** | Supports the testing and validating of AI models without exposing proprietary data or training datasets, such as:<br>▶ Analysing medical images with ML while protecting both the patient data and the proprietary model. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | IEEE 2842-2021 IEEE Recommended Practice for Secure MPC<br>Internet Engineering Task Force (IETF), Privacy Preserving Measurement (PPM) Protocol Standard in Draft<br>ISO/IEC 19592-2:2017 Information Technology – Security Techniques – Secret Sharing – Part 2: Fundamental Mechanisms<br>ISO/IEC CD 4922-1:2023 Information Security – Secure MPC – Part 1: General<br>ISO/IEC 4922-2 Information Security – Secure MPC – Part 2: Mechanisms Based on Secret Sharing |
| **IMDA PET Sandbox Case Study** | ▶ Ant International – Enhancing Customer Engagement with Privacy Preserving AI<br>▶ Meta – Digital Advertising in a Paradigm Without 3rd Party Cookies<br>▶ TikTok – Privacy Preserving Attribution Measurement |

## iii. Trusted Execution Environment

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | | ▶ A Trusted Execution Environment (TEE) is a dedicated area on a computer processor that is isolated and secured from the operating system.<br>▶ It stores data and runs code within its secured area to ensure confidentiality. |
| **Types of Method** | Hardware-based | ▶ Built directly into the central processing unit, it uses dedicated physical components such as an on-chip memory and hardware encryption to create a secure environment. |
| | Software-based | ▶ This implements security through hypervisor-level virtualisation rather than physical hardware.<br>▶ It creates secure environments using software abstraction layers, virtual resource management, and controlled input/output (I/O) channels. |

Common Applications

| Use Case | Description |
|---|---|
| **Identification and Authentication** | Secures personal data in a protected environment during processing, such as:<br>▶ Smartphone biometric unlock systems; and<br>▶ Protected PIN processing in payment terminals. |
| **Secure Analytics** | Enables computation on sensitive data while maintaining confidentiality and integrity, such as:<br>▶ Performing market analysis across competing companies. |
| **Model Training, Validation and Deployment** | Protects AI models and user data during training and inference, such as:<br>▶ Processing user queries on large language models (LLMs) privately. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | ISO/IEC 11889-4:2015 Information Technology – Trusted Platform Module Library<br>IETF TEE Provisioning Architecture<br>IEEE 2830-2021 IEEE Standard for Technical Framework and Requirements of TEE based Shared Machine Learning<br>IEEE SA – IE 2952-2023 Standard for Secure Computing Based On TEE<br>GPD_SPE_055 TEE Trusted User Interface Low-level Application Programming Interfaces (APIs)<br>NISTIR 8320 (May'22) Hardware-Enabled Security: Enabling a Layered Approach to Platform Security for Cloud and Edge Computing Use Cases |
| **Community Reference** | CONFIDENTIAL COMPUTING CONSORTIUM[10]<br>▶ Confidential Computing: Hardware-based Trusted Execution for Applications and Data<br>▶ A Technical Analysis of Confidential Computing v1.3 |
| **IMDA PET Sandbox Case Study** | Singapore Press Holdings (SPH) Media – Collaboration on First-party Data to Enable Customer Activation |

## iv. Zero-knowledge Proofs

Definition and Key Concepts

| | | |
|---|---|---|
| **Definition** | A zero-knowledge proof (ZKP) is a cryptographic method that allows one party (the prover) to demonstrate to another party (the verifier) that a statement is true, without revealing any information beyond the validity of the claim. | |
| **Types of ZKP** | Interactive | ▶ A multi-step process where the verifier repeatedly challenges the prover until convinced.<br>▶ While computationally efficient, it requires ongoing communication and works best with few verifiers. |
| | Non-interactive | ▶ A one-step process where the proof is self-contained and can be verified independently without further communication.<br>▶ Though computationally intensive, it is faster overall and better suited for multiple verifiers. |

[10] *The Confidential Computing Consortium is a project community at the Linux Foundation, which brings together hardware vendors, cloud providers, and software developers to accelerate the adoption of TEE technologies and standards.*

Common Applications

| Use Case | Description |
|---|---|
| **Identity and Age Verification** | Allows individuals to prove specific age or identity requirements without revealing personal details, such as:<br>▶ Mobile banking users proving their identity via face recognition while biometric information remains encrypted on the device. |
| **Asset Ownership Verification** | Enables demonstration of financial standing or asset ownership without disclosing sensitive financial information, such as:<br>▶ Investors proving they meet wealth requirements without disclosing their exact net worth. |
| **Cryptocurrency Transaction** | Maintains network integrity during blockchain transactions while preserving confidentiality, such as:<br>▶ Users proving they have sufficient funds for transactions without revealing their wallet balance;<br>▶ Verifying trading eligibility without exposing transaction history; and<br>▶ Executing smart contracts while keeping specific terms confidential. |

Further Reading and Resources

| Source Type | Document Name |
|---|---|
| **Standards Development Organisation** | ISO/IEC 9798-5:2009 Information Technology – Security Techniques – Entity Authentication Part 5: Mechanisms Using Zero-knowledge Techniques |

# ANNEX 2 – PROPOSED IMPLEMENTATION CHECKLIST

# ANNEX 2 – PROPOSED IMPLEMENTATION CHECKLIST

This checklist provides guidance for organisations throughout their PETs deployment journey across three phases:

1. **Pre-implementation**, which includes assessing internal technology readiness and external solution provider suitability;
2. **During implementation**, which includes initial evaluations on deployed PETs solution; and
3. **Post-implementation**, which includes ensuring that the deployed solution continues to meet business requirements.

## 1. Pre-implementation: Assessing Readiness

Considerations in this phase include defining the problem, determining whether PETs are needed, and engaging internal stakeholders. This involves assessing the organisation's readiness to support the PETs solution, evaluating the PETs solution provider, and determining whether the solution is able to meet the organisation's needs.

| Steps | Stakeholders |
|---|---|
| A **Use Case Definition** | Business<br>Technology<br>Data Protection |

This section outlines key considerations around business use case fundamentals, to assess if and how a PET could address data sharing needs while protecting sensitive data.

**To Do:**

- ☐ Identify the business challenge requiring PETs to enable data sharing (e.g. legislation, internal policies, contractual limits, technical constraints).
- ☐ Analyse data requirements:
  - ○ Identify the data needed and whether it includes personal data; or if less granular or aggregated data would be sufficient.
  - ○ Determine sources of data (internal, external partners or across jurisdictions).
  - ○ Note any constraints on data sharing.
- ☐ Build a data flow diagram:
  - ○ Map and analyse specific data flows (ingestion, processing, storage and access).
  - ○ Specify key data points and how they move between systems to pinpoint where and how PETs would add value.
- ☐ Use **IMDA's PETs Discovery Tool** to find suitable PETs according to your use case.
- ☐ Define stakeholder roles (e.g. data controller, processor, intermediary) and their responsibilities.
- ☐ Establish applicable data protection obligations.

| Steps | Stakeholders |
|---|---|

## B Partners and Internal Stakeholder Alignment

The use of PETs typically involves data sharing or collaboration with internal or external partners.

This section outlines key action items to ensure all stakeholders are aligned on key project parameters, including the project scope, chosen solution, success metrics and expected outcomes.

**To Do:**

- ☐ Align with collaborating partners on data sharing requirements, project scope and expected outcomes.
- ☐ Establish key business metrics (e.g. compute requirements, latency, etc.) and privacy metrics (e.g. ε in differential privacy, k-value in k-anonymity).
- ☐ Form a cross-functional team (business, technology, data protection) to address regulatory requirements, including laws that restrict data use or sharing, and assess how PETs can support compliance.

**Stakeholders:**

- Business
- Technology
- Data Protection

## C Internal Technology and Data Readiness

Before implementing a PETs solution, organisations should assess whether existing datasets and the technical operating environment can support integration, and whether additional capabilities are required.

This section outlines key action items to evaluate technical readiness and prepare data accordingly.

**To Do:**

- ☐ Assess the current operating environment for compatibility with the PETs solution:
  - ○ Determine whether additional computing or networking resources are required.
  - ○ Review integration needs (e.g. APIs compatibility).
- ☐ Standardise and document data in the data dictionary:
  - ○ Data formats (e.g. casing, data types)
  - ○ Constraints (e.g. min-max, non-null values)
  - ○ Data categories and classifications
  - ○ Sensitivity levels
- ☐ Cleanse and prepare data:
  - ○ Remove outliers and unnecessary identifiers, (e.g. before model training or post-training).
  - ○ Reduce only essential data fields.
- ☐ Format data structures for PETs input.

**Stakeholders:**

- Technology – Infrastructure
- Technology – Data

| Steps | Stakeholders |
|---|---|

D

## PETs Solution Assessment

Selecting the right PETs solution requires a systematic review of both business and technical capabilities.

This section outlines key action items to assess solution suitability and validate technical capabilities.

**To Do:**

### Solution suitability

- ☐ Evaluate key performance metrics to ensure alignment with business requirements (e.g. compute cost efficiency, latency, output accuracy).
- ☐ Assess how well the PETs solution addresses business and operational needs (e.g. specific data sharing challenges, operational constraints, scalability, integration complexity with existing systems).
- ☐ Analyse data protection risks mitigation while ensuring the level of data utility required by the business.
- ☐ Verify the solution provider's credibility.
  - ○ Review business references from comparable implementations.

### Technical Capabilities

- ☐ Confirm adherence to recognised technical standards or best practices (e.g. ISO, IEEE, NIST).
- ☐ Request independent audit reports detailing the PETs solution's capabilities, features and limitations.
- ☐ Assess the effectiveness of cybersecurity controls and governance frameworks in mitigating data protection risks (e.g. access management, data handling procedures, audit logging, incident response mechanisms).
- ☐ (Good to have) Validate recognised certifications, such as the Data Protection Trustmark (DPTM), Cyber Essentials Mark (CEM), Cyber Trust Mark (CTM), ISO 27001, and System and Organisation Controls (SOC) 2.

Business

Technology

## 2. During Implementation: Deployment and Compliance Validation

Considerations in this phase concern the actual deployment of the PETs solution, including systems and integration checks, the implementation of data protection measures, and addressing regulatory obligations. The process is iterative to ensure that measures are properly implemented, and obligations are addressed.

| Steps | Stakeholders |
|---|---|
| **E**  **Project Governance and Change Management**<br><br>After initial implementation, effective project management is essential to ensure timely delivery whilst mitigating risks, and to accommodate necessary changes.<br><br>This section outlines key action items to ensure proper project governance processes are in place.<br><br>**To Do:**<br>☐ Ensure robust project management practices:<br>○ Conduct milestones reviews<br>○ Track timelines<br>○ Ensure deliverable quality<br>☐ Manage project risks through continuous monitoring of key metrics:<br>○ Review PETs performance<br>○ Assess data quality<br>○ Detect data drift over time<br>☐ Establish proper change control processes to account and document change requests, budget, and project timelines. |  Business |
| **F**  **Technical Implementation and Data Protection Assurance**<br><br>This section outlines technical and data protection related action items following the initial implementation of the PETs solution, to ensure that data is processed securely and protected adequately.<br><br>**To Do:**<br>**Environment and Access Management**<br>☐ Set up a staging or pre-production environment for isolated testing of PETs solution.<br>☐ Implement access control mechanisms to assign proper rights and privileges for PETs solution users.<br>☐ Ensure production and staging data remain separate, with role-based access restrictions. |  Technology – Infrastructure<br>  Technology – Security<br>  Data Protection |

| Steps | Stakeholders |
| --- | --- |
| **Integration and Testing**<br>☐ Fully test PETs API and Software Development Kits' (SDKs) functionality before the integration into existing data pipelines.<br>☐ Verify that the PETs solution and integration meet project requirements and cybersecurity baselines through assurance testing, such as unit-testing, integration testing, user acceptance testing, vulnerability assessment and penetration testing.<br>**Data Protection and Minimisation**<br>☐ Apply data minimisation principles to use only the data necessary for the use case.<br>☐ Use synthetic data or data augmentation techniques for testing, instead of using production data.<br>☐ Verify dataset sample sizes are sufficiently large enough to reduce re-identification risk.<br>☐ Conduct reviews to assess the presence of personal data.<br>☐ Ensure the PETs solution sufficiently addresses data protection objectives, e.g. remove or hide identifiers to lower the re-identification risk – see below for **Data Protection Best Practices on Handling Identifiers.**<br>*Note: This process may be iterative.* | |

## Data Protection Best Practices

### Practical Guidance on Handling Identifiers

| | |
|---|---|
| **Remove Direct Identifiers**  | ▶ Identify all direct identifiers in the input dataset and verify that they are no longer present in the output dataset.<br>▶ Review how direct identifiers are removed from the input data, including any techniques beyond the PETs solution.<br>▶ Check that direct identifiers from the input data are fully removed from the output data. |
| **Hide Identifiers Within Datasets**  | ▶ Hide identifiers that remain in the dataset and obscure any relationships between direct and indirect identifiers.<br>▶ Ensure direct or indirect identifiers are hidden, e.g. through obfuscation, or data splitting to disassociate linkages.<br>▶ Encrypt identifiers and securely manage encryption keys, ensuring proper and secure rotation.<br>▶ If hashing is used to "hide" the identifiers, consider adding random values known as "salts" to the identifiers prior to hashing to minimise re-identification risk. This is achieved through the matching of hashes to precomputed hash tables.<br>▶ Disassociate any mapping between hidden identifiers and raw identifiers, retain mappings only if needed for a specific purpose (e.g. identifying a single individual).<br>▶ If relationship mapping is retained, these mappings must be securely stored, preferably out-of-band and away from the hidden identifiers. |
| **Prevent Linkages Across Other Datasets**  | ▶ When indirect identifiers remain in the dataset, consider applying additional PET techniques to prevent linkage with other datasets especially when there is a need to release or disclose the data. Examples include differential privacy or synthetic data generation.<br>▶ When further PETs are impractical, apply alternative techniques to reduce identifiable linkage across datasets, such as generalisation or data aggregation on indirect identifiers.<br>▶ Subsequently evaluate the risk of re-identification using methods such as k-anonymity.<br>▶ Ensure direct identifiers are removed or hidden and remain hidden for their entire lifecycle. Its hidden form (e.g. encrypted, hashed or split) should never be linkable to associated indirect identifiers.<br>▶ Review whether any direct identifiers need to be stored, and retain them only when necessary for the intended purpose. |

## 3. Post-implementation: Delivering Utility and Privacy Outcomes

Considerations in this phase include monitoring, maintaining, and refining the PETs solution deployed to ensure it continues to meet performance expectations and complies with data protection regulations.

| Steps | Stakeholders |
|---|---|
| **G**  **Evaluating PETs Performance for Business and Privacy Requirements**<br>The relationship between utility and privacy typically involves trade-offs, where increased security or privacy may reduce system performance.<br>This section outlines action items to help companies balance requirements and achieve acceptable outcomes for all stakeholders.<br>**To Do:**<br>☐ Establish internal baseline thresholds for acceptable performance vs. the required privacy level (e.g. determine whether the increased processing time from homomorphic encryption is tolerable for enhanced data protection).<br>☐ Monitor and refine performance metrics through collaborative efforts to identify the optimal balance between data protection and utility.<br>*Note: This process may be iterative.* |  Business<br>  Technology<br>  Data Protection |

| Steps | Stakeholders |
| --- | --- |

## H Monitoring and Optimisation

This section outlines action items for continuous monitoring, audit logging, ongoing assessments, and optimisation of the PETs solution, to be considered as part of the organisation's enterprise risk framework.

**To Do:**

- ☐ Track performance metrics such as processing speed, latency, and response time, to ensure optimal PETs operations.
- ☐ Maintain consistent audit logs of all access activities, including user identities, data accessed, timestamps, and purpose of access.
- ☐ Implement continuous monitoring to address threats, data breaches and policy violations in a timely manner.
- ☐ Ensure proper incident reporting mechanisms and protocols are in place.
- ☐ Implement user feedback processes to improve PETs effectiveness and identify potential errors.
- ☐ Document implementation processes, privacy safeguards, and compliance measures taken throughout the deployment lifecycle.
- ☐ Conduct periodic data protection assessments, including:
  - ○ Regularly review (e.g. annually) the deployed PETs solution to ensure alignment with business and privacy expectations.
  - ○ Periodically evaluate datasets for potential exposure of personal data, especially after system updates or infrastructure changes.
  - ○ Assess re-identification risks in data outputs, considering technological advancements, new auxiliary data sources, major system changes, external data releases, or updated regulatory guidance.
  - ○ Ensure proper disposal of records containing personal data.

**Stakeholders**

- Business
- Technology – Infrastructure
- Technology – Security

# ACKNOWLEDGEMENTS

The IMDA and PDPC team would like to acknowledge industry partners from the following organisations, whose feedback have been instrumental in helping to refine the PETs Adoption Guide, Discovery Tool and Implementation Checklist.

Companies (In alphabetical order):

- Advance.AI
- Ant International
- AsiaDPO
- Agency for Science, Technology and Research Institute for Infocomm Research (A*STAR $I^2R$)
- Betterdata
- Digital Trust Centre
- Fortanix
- Google
- Government Technology Agency (GovTech)
- Grab
- IBM
- Kajima Corporation
- Mastercard
- NHG Health
- National University of Singapore (NUS) AI Institute
- Oblivious.AI
- OCBC
- Partisia
- Prudential
- Silence Labs
- Singapore Press Holdings (SPH) Media
- TikTok

# PRIVACY ENHANCING TECHNOLOGIES ADOPTION GUIDE

2026