PRIVACY ENHANCING TECHOLOGY - PROPOSED ADOPTION GUIDE

DRAFT FOR INDUSTRY CONSULT

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Contents

Overview

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

The PETs adoption guide, which the IMDA is proposing, is **intended for C-suite and senior decision-makers in organisations considering the adoption of PETs.** The guide outlines how leading organisations in different industries are leveraging PETs to:

- **Enable Artificial Intelligence and Machine Learning (AI/ML) development** by enabling safe access to more data for AI training; and
- **Forge multi-stakeholder collaborations (e.g., across teams in organisations or across sectors),** by allowing existing or more data to be used by different parties without compromising data confidentiality or privacy.

With more organisations now involved in real-world deployments of PETs with demonstrated value, there is also **growing momentum by the PETs industry** to address the varying demands of organisations. Besides enterprise-grade solutions being offered by technology companies, some **PETs are also increasingly becoming more cost-effective and straightforward to use** due to the availability of open-sourced software development tools and low or no code platforms.

The IMDA would like to better support organisations to adopt PETs by addressing challenges and barriers organisations face in the evaluation and implementation of PETs. The proposed guide would be accompanied by two practical resources to help organisations identify the PETs and processes that are relevant to their use case:

- **PETs Use Case Evaluation Tool –** Intended to help users match their use case to proven PET solutions, with real-world case studies to illustrate impact
- **Implementation Checklist –** Intended to include areas of consideration spanning Pre, During, and Post implementation across the PETs deployment journey to ensure secure and effective deployment

The guide and accompanying resources would have been developed based on insights derived from real world use cases in IMDA's PETs Sandbox. The IMDA intends for these resources to be constantly enhanced as new organisations participate in IMDA's PETs Sandbox, and we continue to incorporate feedback from industry and technology experts.

As organisational needs continue to grow, and technologies continue to evolve, PETs are no longer just about compliance—they are becoming a pathway to competitive advantage in a data-driven economy.

PETs have traditionally been used in the context of regulatory compliance

1. PETs allow the processing, analysis and extraction of insights from data without revealing the underlying personal or commercially sensitive data. Common PETs include technologies that obfuscate or hide the original data, for example homomorphic encryption, and technologies that facilitate the flow of insights without the transfer of data, such as federated learning.

2. PETs have been in existence for some time and have primarily been used to address privacy and data protection risks during data sharing and collaboration. For instance, in a recommendation on measures that supplement transfer tools in 2021, the European Data Protection Board has mentioned the use of PETs as a useful "supplementary tool" in the context of international data transfers to ensure compliance. Besides Europe, other data protection regulators around the world have also recognised the usefulness of some PETs such as the generation of synthetic data in addressing privacy and data protection risks[1].

PETs are increasingly being used to unlock value with the trusted use of data

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

   These are observed across two areas:

   a. **Access to data to enable AI/ML training.** Finetuning or tweaking 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 machine learning purposes, the emergence of PETs has gradually enabled privacy preserving machine learning, whilst ensuring sensitive data remains protected.

      Example of use cases include:

Training of AI model to predict customer's promotion-to-conversion likelihood

[1] For example, the Office of the Privacy Commissioner of Canada (OPC) had in a publication on synthetic data highlighted its benefits in protecting against traditional re-identification attacks and possibility of automation of the de-identification process. The Personal Information Protection Commission of South Korea (PIPC) and the Personal Data Protection Commissioner of Singapore (PDPC) have likewise published guidance on the use of synthetic data to address privacy and data protection risks.

Use case: Ant International[2] wanted **to train a machine learning model that could better predict each customer's promotion-to-conversion** likelihood by using historical data on common customers from its partner.
Solution: **Multiparty Computing Private Set Intersection (MPC-PSI) and Federated Learning (FL)** were used to privately identify common customers and ensure sensitive data could be trained in a decentralised manner and not shared with either party.

Training of diagnostic imaging AI models

Use case: Healthcare and tertiary learning institutions seek to **collaborate on training and improving of machine learning models** without revealing the identity of individual patients. However, their datasets typically contain sensitive personal information on patients.

Solution: The institutions leverage the NVIDIA Clara platform[3] which uses **Federated Learning (FL)** to enable collaborative AI development such as improving diagnostic imaging models. This approach safeguards patient data by allowing institutions to jointly use their data for machine learning without the different institutions being able to view one another's raw data.

b. **Extracting insights from data by multiple stakeholders (e.g., across teams in organisations or across sectors).** Different parties can come together to collaborate on data-driven projects. Parties previously find it impossible or impractical to forge such partnerships due to inherent confidentiality or privacy concerns. For instance, PETs allow organisations to collaborate on anonymised datasets with low risk of identifiability, without having to obtain consent from millions of existing customers to use their data for these collaborations.

Examples of use cases include:

Collaboration across multiple entities on financial crime intelligence

Use case: Mastercard wanted to share International Bank Account Numbers (i.e. IBAN) across multiple overseas banks to identify suspicious accounts for anti-money laundering purposes. However, this could constitute a disclosure of commercially sensitive information (i.e. information about Mastercard's customers).

Solution: For this POC, Mastercard used **Fully Homomorphic Encryption (FHE)**, to encrypt sensitive data in the query that was sent to a few banks.

[2] IMDA PETs Sandbox Use Case
[3] NVIDIA Clara Federated Learning to Deliver AI to Hospitals While Protecting Patient Data, Dec 2019

Individual banks would perform data matching on the encrypted query with their own customer information, but they would not "see" the sensitive data from Mastercard, and were prevented from learning which records in their data may have matched the query.

Partnership to share first-party data for targeted advertising

Use case: To enable advertisers to serve relevant and targeted advertisements to potential customers on its digital network, SPH Media explored the use of PETs to profile and amplify the existing customer base of their advertiser without the collection, sharing or learning of any personal data between the 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 in a highly secure environment while protecting the privacy of individuals.

PETs are increasingly becoming more cost-effective and easier to use

4. With increasing demand from organisations to leverage PETs, some PETs solutions are becoming increasingly productised. Large technology companies are now making available PETs such as Trusted Execution Environments[4], Differential Privacy[5] to their enterprise customers.

5. Furthermore, to enable broader access to PETs, some organisations have also open-sourced[6] their tools, making APIs[7], SDKs[8] and even no code or low code[9] options more readily available. These open-sourced tools[10] are typically more cost-effective and enable organisations to more easily explore the use of PETs

[4] Example of TEE integration is Azure's Confidential Computing, allowing secure data processing in sensitive environments.

[5] Example of Differential Privacy is Google's BigQuery DP which is an interface to apply Differential Privacy to datasets.

[6] Examples of these open-sourced tools include: Microsoft SEAL, an open-source Simple Encrypted Arithmetic Library (HE library); Google's Tensorflow Privacy which implements differential privacy for machine learning, as well as Private Join and Compute, which is an MPC tool, based on private set intersection (PSI); and PrivMRF, an open-source data synthesis tools which won the first place in the 2020 edition of NIST Differential Privacy Synthetic Data Challenge.

[7] API (Application Programming Interface) is a set of rules and specifications that allow different software systems to communicate with each other.

[8] SDK (Software Development Kit) is a collection of tools, libraries, documentation, or pre-built code examples, that help developers build applications for a specific platform or with specific functionalities.

[9] Examples of no / low code solutions include: Fortanix Armor which offers FHE enabled AI tools with no-code deployment interfaces; and 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 help organisations shorten time to evaluate and implement PETs

6. Despite the success of PETs adopters in implementing PETs for their specific use cases, **identifying and implementing the right PETs remains a complex process for most organisations.** This is because different PETs could be used to address a particular use case, and the optimal solution typically depends on multiple considerations including business needs, technology specifications, and various regulatory requirements. The scenarios outlined below illustrate the complexity of the process in identifying and implementing the right PETs just based on one consideration.

> **Scenario 1:** Two entities want to conduct an analysis of their private datasets and are **open to sharing** and pooling the datasets together.
>
> **Scenario 2:** Two entities want to conduct an analysis of their private datasets, and the entities would like to process their datasets **locally without sharing datasets with each other**.
>
> Examples of PETs they could explore include:
> - **Multi Party Computing (MPC),** a technique where at least two collaborating partners jointly analyse data, without any party needing to share all its information with one another; or
> - **Trusted Execution Environment (TEE),** where data from collaborating partners are uploaded into a secure environment for analysis.
>
> **For scenario 1,** Compared to the use of MPC, **TEE** may be more appropriate. Both entities can each upload their datasets in the secure environment and run their analysis on the TEE. Using MPC is likely to cost more given the greater computational or encryption overheads.
>
> **For scenario 2**, Compared to the use of TEE, **MPC** may be more appropriate. This is because MPC allows for each party to collaborate on the analysis of data without sharing their datasets with each other.
>
> In real world, there may be multiple other considerations including cost, time, regulatory jurisdiction or limitations to state of technology, adding further layers of complexity.

7. Therefore, to assist organisations in evaluating and determining the best PETs solution to address their problem statement, as well as in implementing the PETs solution, this guide offers the following two resources:

a. **PETs Use Case Evaluation Tool**

i. The evaluation tool, intended to be made available after industry consultations, aims to help users identify the suitable PETs for their use case through a decision-making flow whilst providing relevant implementation examples. It is intended to operate through a straightforward interface requiring just two user inputs: industry sector (e.g. finance) and specific use case (e.g. anti-money laundering).

ii. Based on these inputs, the tool provides relevant use case studies, including concise information about the problem statement, the PET solution implemented, its rationale, and effectiveness.

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

b. **Implementation Checklist**

The checklist provides guidance for organisations throughout their PETs deployment journey across the following three phases:

i. **Pre implementation:** Ensures proper assessment of business need, as well as evaluation of PETs solution providers and solution

ii. **During Implementation:** Ensures iterative testing and validation of the PETs solution, and effective data protection and compliance measures

iii. **Post Implementation:** Addresses ongoing performance monitoring, scalability considerations and ongoing regulatory compliance

Please see **Annex 2** for details of the implementation checklist.

Conclusion

8. Multiple real-world implementations have shown that some PETs have progressed beyond the experimental stage. No longer just compliance tools, organisations are adopting PETs to support practical, high-impact use cases.

9. Given the good potential for organisations to unlock value with the trusted use of data, PETs and the landscape of technology providers would continue to evolve and mature. It is hence a sensible time for organisations to begin seriously exploring where PETs might fit in their data strategies, as well as to experiment and implement them in their products or operations.

10. This proposed Adoption Guide is aimed at helping organisations kickstart their journey to make use of these technologies. Intended as a “living document”, industry and expert feedback is strongly encouraged for us to continue to refine and improve the guide, to ensure its relevance to real world use cases.

Annex 1 – Types of PETs and its applications[11]

Categories of PETs that	PETs	Examples of applications (non-exhaustive)
a) Technologies that obfuscate or hide the original data	Differential Privacy	• Data Release • Data Collection • Model Training
	Homomorphic Encryption	• Secure Analytics • AI/ML Training • Verifiable Secure Computations
	Synthetic Data	• AI/ML Training • Data Analysis and Collaboration • Software Testing
b) Technologies that facilitate the flow of insights	Federated Learning	• Healthcare and Medical Analysis • Fraud Detection • Edge Device Training
	Secure Multi-Party Computation	• Combining Multi-Source Data • Detecting and Matching with PSI • AI Model Training and Validation
	Trusted Execution Environments	• Identification and Authentication • Data Processing and Analytics • AI/ML Privacy
	Zero Knowledge Proofs	• Identity and Age Verification • Asset Ownership Verification • Cryptocurrency Transaction

a. Technologies that Obfuscate or Hide the Original Data

i. Differential Privacy

Definition and Key Concepts

Definition	Differential Privacy ("DP") is a mathematical framework that provides formal privacy guarantees by adding controlled random noise to data or computations. It ensures that the presence or absence of an individual's data in a dataset cannot be determined from the analysis results, while 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. A lower ε value provides stronger privacy guarantees but reduces accuracy. The privacy budget is cumulative, meaning it gets consumed across multiple queries or releases.	
Types of Implementations	Global DP	Involves a trusted aggregator who collects raw data and adds noise to the results before sharing them. This approach adds noise only once at the aggregate level.
	Local DP	Involves having individuals add noise to their data before sharing it with anyone else. This approach is suitable for large-scale data collection where individual accuracy is less critical than aggregated insights.

Common Applications

Domain	Example
Data Release	Publishing aggregate datasets while protecting individual privacy. E.g., national census data publication, public health statistics about disease prevalence, sharing research datasets with the academic community, and publishing geographic data about population movements or service usage patterns while maintaining individual anonymity.
Data Collection	Gathering insights about user behaviour and system performance E.g., operating systems can collect device performance metrics and crash reports, browsers can gather information about frequently visited websites to improve caching, and mobile applications can collect usage patterns to enhance user experience
Model Training	Developing AI systems that learn from sensitive personal information. E.g., training content safety classifiers using differentially private synthetic data, fine-tuning language model that don't memorise sensitive training examples

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	NIST SP 800-226 Guidelines for Evaluating Differential Privacy Guarantees
IMDA PET Sandbox Case Study	• Meta – Digital Advertising In A Paradigm Without 3rd Party Cookies • TikTok – Privacy Preserving Attribution Measurement

ii. **Homomorphic Encryption**

Definition and Key Concepts

Definition	Homomorphic encryption (HE) refers to a family of encryption schemes with a special algebraic structure that allows computations to be performed directly on encrypted data without ever decrypting.	
Types of HE	Fully HE (FHE)	FHE is the most comprehensive type of HE, allowing general computations to be performed on encrypted data. It also has no limits on the number of operations. However, its complexity means it can be computationally expensive.
	Somewhat HE (SHE)	SHE permits general computation up to a limit determined when the keys are generated.
	Partial HE (PHE)	PHE supports only addition or multiplication but not both

Common Applications

Domain	Example
Secure Analytics	Processing confidential information while keeping it encrypted throughout the analysis pipeline. E.g., tumor detection in MRI scans by third-party services
AI/ML Training	Training and running models on encrypted data without exposing the underlying information. E.g., cloud-based ML model training with encrypted data
Verifiable Secure Computations	Performing calculations that can be verified while maintaining data confidentiality. E.g., encrypted vote counting, ballot tracking and checking system

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	ISO/IEC 18033-6:2019 IT Security techniques - Encryption algorithms - Part 6: Homomorphic Encryption
	ISO/IEC CD 28033-1 Information security — Fully Homomorphic – Part 1: General <in progress>
Community Reference	homomorphicencryption.org – HE Standard
IMDA PET Sandbox Case Study	• Mastercard – Secure Sharing of Financial Crime Intelligence • Ant International - Enhancing Customer Engagement With Privacy Preserving AI

iii. **Synthetic Data**

Definition and Key Concepts

Definition	Synthetic Data (SD) refers to artificial data that has been generated using a purpose-built mathematical model (including artificial intelligence (AI)/machine learning (ML) models) or algorithm.	
Types of SD[12]	Fully Synthetic	Does not contain any connection to real-world data and data is completely artificial. This type offers the strongest privacy protection.
	Partial Synthetic	Retains all the information from the original real-world data except the sensitive data which is replaced with synthetic data. As not all original values are removed, the risk of re-identification is greater than that of a fully synthetic dataset.

Common Applications

Domain	Example
AI/ML Training	Generating large volumes of synthetically generated dataset for training and testing AI/ML models or augmenting training datasets can be cost-effective, especially when the source datasets are sparse. E.g., synthetic data has been employed for training AI for self-driving car projects like "Waymo"
Data analysis and collaboration	Making use of underlying trends or patterns of the data are useful for data analytics and it also allows data exploration as well as sharing of the analysis in industries and sectors like healthcare. E.g., sharing high-quality AI generated synthetic data for healthcare research
Software testing	Using of synthetic data for testing in software development can help organisations avoid data breaches in the event of the development environment being compromised. E.g., enterprise application system or product testing

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	IEEE SA Synthetic Data <in progress>
Community Reference	Dr Khaled El Emam Paper Privacy Disclosure Metrics for Synthetic Health Data
PDPC Guide	Proposed Guide to Synthetic Data Generation
IMDA PET Sandbox Case Study	Kajima – Generating Synthetic Data for Analysis and Research

[12] Types of SD 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

b. Technologies that Facilitate the Flow of Insights

iv. 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 (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.	
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	Uses same features across multiple parties
	Vertical FL	Uses different features about the same users across multiple parties

Common Applications[13]

Domain	Description and Example
Healthcare and Medical Analysis	Precision medicine research and medical image analysis where patient data is sensitive and must remain within institutional firewalls. E.g., cancer detection model training across multiple hospitals using pathology images while keeping the data secure within each institution.
Fraud Detection	Multiple organisations collaborate on fraud detection. E.g., life insurance organisations jointly training fraud detection models using synthetic data generated from their real-world data
Edge Device Training	Training of AI models on edge devices (like phones, laptops, digital assistants) while keeping user interaction data local. E.g., voice recognition, text prediction, and personalised device interactions.

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	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

[13] FL is typically combined with other PETs. FL alone, while keeping raw data local, doesn't guarantee complete privacy as model parameters exchanged during training can leak sensitive information through various attacks.

v. **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 their combined data without revealing individual input to each other. Only the final agreed-upon computation result is revealed to the participating parties.	
Types of Method	Secure Aggregation Protocol	This often involves more than two parties. The computation happens on encrypted data, and only the final aggregate result is decrypted.
	Private Set Intersection (PSI)	This lets two parties find matching elements in their datasets without revealing their own data.

Common Applications

Domain	Example
Combining Multi-Source Data	Multiple parties combine and analyse their collective data. E.g., government agencies analysing citizens data across social services, or banks merging transaction and demographic data.
Detecting and Matching with PSI	Two parties identify matching elements between datasets. E.g., checking breached passwords across databases, or finding overlapping population segments between organizations.
AI Model Training and Validation	Organizations validate and test AI models without exposing proprietary data or training datasets. E.g., analysing medical images with machine learning while protecting both the patient data and the proprietary model.

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	IEEE 2842-2021 IEEE Recommended Practice for Secure Multi-Party Computation
	IETF, Privacy Preserving Measurement (PPM) protocol standard
	ISO/IEC 19592-2:2017 Information technology - Security techniques - Secret sharing - Part 2: Fundamental mechanisms
	ISO/IEC CD 4922-1:2023 Information security - Secure multiparty computation - Part 1: General
	ISO/IEC 4922-2 Information security - Secure multiparty computation - Part 2: Mechanisms based on secret sharing
IMDA PET Sandbox Case Study	• Ant International - Enhancing Customer Engagement With Privacy Preserving AI • Meta – Digital Advertising In A Paradigm Without 3rd Party Cookies • TikTok – Privacy Preserving Attribution Measurement

vi. **Trusted Execution Environment**

Definition and Key Concepts

Definition	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 in a protected way.	
Security Model	Hardware-based	Built directly into the CPU, it uses dedicated physical components and features like on-chip memory and hardware encryption to create a secure environment.
	Software-based	This implements security through hypervisor-level virtualization rather than physical hardware. It creates secure environments using software abstraction layers, virtual resource management, and controlled I/O channels.

Common Applications

Domain	Example
Identification & Authentication	Securing personal data in a protected environment during process. E.g., smartphone biometric unlock systems, protected PIN processing in payment terminals
Data Processing & Analytics	Enabling computation on sensitive data while maintaining confidentiality and integrity. E.g., market analysis across competing companies
AI/ML Privacy	Protecting AI models and user data during training and inference. E.g., LLMs processing user queries privately

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	ISO/IEC 11889-4:2015 Information Technology — Trusted Platform Module Library
	IETF Trusted Execution Environment Provisioning Architecture
	IEEE 2830-2021 IEEE Standard for Technical Framework and Requirements of TEE based Shared Machine Learning
	IEEE SA - IE 2952-2023 Standard for Secure Computing Based On TEE
	GPD_SPE_055 TEE Trusted User Interface Low-level API
	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 • Confidential Computing: Hardware-Based Trusted Execution for Applications and Data • A Technical Analysis of Confidential Computing v1.3
IMDA PET Sandbox Case Study	SPH Media – Collaboration on First Party Data to Enable Customer Activation

vii. **Zero-knowledge Proofs**

Definition and Key Concepts

Definition	A zero-knowledge proof (ZKP) is a cryptographic method that allows one party (the prover) to prove to another party (the verifier) that a specific claim or statement is true without revealing any additional information beyond the validity of the claim itself.	
Types of ZKP	Interactive	A multi-step process where the verifier challenges the prover repeatedly until convinced. 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. Though computationally intensive, it's faster overall and better suited for multiple verifiers.

Common Applications

Domain	Example
Identity and Age Verification	People prove they meet specific age or identity requirements without revealing their actual personal details. E.g., a person can prove they are old enough to drive or to access certain services without showing their birth certificate or ID card. When using facial recognition for mobile banking or accessing corporate systems, users can prove their identity with ZKP while biometric information remains encrypted and never leaves their device.
Asset Ownership Verification	Individuals demonstrate their financial standing or asset ownership without exposing sensitive financial information. E.g., property owners can prove their ownership status without exposing their purchase history, while investors can demonstrate they meet wealth requirements without disclosing their exact net worth.
Cryptocurrency Transaction	To maintain network integrity during blockchain transactions. Users can prove they have sufficient funds for transactions without revealing their wallet balance, verify their trading eligibility without exposing transaction history, and execute smart contracts while keeping specific terms confidential.

Further Reading and Resources

Source Type	Document Name
Standards Development Organisation (SDO)	ISO/IEC 9798-5:2009 Information technology - Security techniques - Entity Authentication Part 5: Mechanisms using zero-knowledge techniques
Community Reference	Zkproof.org - ZKProof Community Reference (Version 0.3)

Annex 2 – Proposed Implementation Checklist

This checklist provides guidance for organisations throughout their PETs deployment journey across three phases – 1) pre-implementation, 2) during implementation, and 3) post-implementation.

1) <u>Pre-implementation:</u>

Considerations in this phase include defining the problem and assessing the need for PETs with internal stakeholders within an organisation. It also includes assessments on the organisation's readiness to support the PETs solution, as well as assessments of the PETs solution provider, and solution to meet the needs of the organisation.

a) Defining the problem and assessing the need and readiness of the organisation to adopt PETs with internal stakeholders:

	Business-related
1	**Use case definition** • Clarify business use case where a PET may be needed to address a data sharing or collaboration challenge and how it could be used to address the challenge
2	**Data flows and roles** • To determine and map existing data flows and roles within the company (e.g. how data is ingested, processed, stored and accessed) to identify where risks and where the PET may be applied
	Technology-related
3	**Infrastructure and deployment** • Analyse various deployment options, such as cloud, on-premise, or hybrid solution to assess the best fit • Assess whether current infrastructure set up and architecture would be able to handle compute requirements
4	**Interoperability with data tools** • Determine if APIs or SDKs are supported for the PETs solution • To check internally if the PETs solution could work within machine learning workflows (e.g. TensorFlow or PyTorch)
5	**Data preparation and transformation:** <u>Data Quality and Structure</u> • To check if required datasets need to be cleaned, completed and formatted for the use of a specific PET technique and perform the required data pre-processing <u>Data Classification</u>

	• To check if datasets and data types have assigned classification or sensitivity levels to ensure appropriate handling Data Minimisation • Consider reducing data scope to essential elements only and removing unnecessary identifiers • Consider whether certain data fields can be eliminated without compromising the solution's effectiveness Synthetic data for Testing • Consider the use of synthetic data for testing before using real data
6	**PETs Architecture design considerations** Integration Points • Identify where integration points for PETs implementation within the IT infrastructure may be • Provision for any additional hardware, software and middleware if deploying in-house. (e.g. some PETs like TEE may require specialised hardware or software in its implementation that the existing infrastructure may not have in place) Security protocols and Safeguards • Determine encryption, key management, access controls and external communications requirements needed by the PETs (e.g. HE which requires proper key management) and • Assess implications to the existing organisation's security policies on how encryption and cryptographic keys should be handled and securely used Audit and Monitoring • To ensure there are existing tools within the company or PETs solution provider to monitor and track security issues, privacy metrics expected by the PETs functionality, and performance
	Legal/Data Protection related
8	**Legal and compliance assessments** • To ensure data protection and / or legal teams are kept in the loop throughout the implementation cycle and have assessed and reviewed solution for potential regulatory implications • To ensure Data Protection Impact Assessments (DPIA) were conducted to assess whether the data protection risks can be mitigated, and whether residual risks are acceptable

b) Assessing the PETs solution provider and the solution:

	Business related
1	**Expected performance guarantees** • Assess the impact of PETs solution on computational cost, operational latency and overall performance to ensure alignment with business needs
	Technology related
2	**Technical standards and regulatory requirements** <u>Solution Reliability</u> • PETs solution providers should provide evidence or attestation of their adherence to industry standards and best practices, for e.g. industry best practices such as those established by ISO, IEEE, NIST, to ensure functional reliability. <u>Infrastructure and Hosting Resilience</u> • Check if PET solution adheres to baseline cybersecurity and data protection best practices as outlined in industry guidelines such as PDPC's ICT guide, or recognised certifications including DPTM, CEM, CTM, ISO 27001, SOC 2 <u>Assurance of PETs solution</u> • Show evidence that regular audits are conducted on PETs solutions to ensure compliance to regulations, standards or best practices <u>Failure Recovery and Redundancy</u> • To put in place PETs solution contingency plans, for e.g. in the application of federated learning or multi-party computing, there may be a need to address the event a node fails
	Legal/Data Protection related
3	**Expected privacy guarantees** • Assess the PETs solution's expected effectiveness in addressing privacy-related risks, such as re-identification • Evaluate effectiveness of additional governance and data protection features in place, including access controls and audit logging

2) <u>During Implementation:</u>

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

	Business related
1	<u>Project Governance, Scope and Change Management</u>

	• Ensure proper project management practices are implemented to manage scope and effectively track timelines and milestones • Provide regular updates on project milestones to stakeholders • Manage project risks through continuous monitoring of key metrics (e.g. PETs performance, quality of the data, any data drift over time) • Ensure proper change control process is in place to account and document change requests, budget and project timelines
	Technology related
2	<u>Staging Environment</u> • Consider performing software testing and validation activities in a staging or pre-production environment to allow safer isolated testing <u>Authentication and Access</u> • Decide between using either federated or distributed Identity and Authentication Management (IAM) to integrate proper access control into the PETs solution. <u>Quality testing and Validation</u> • Ensure systems and software quality by performing unit testing of software functions and integration testing how the functions properly integrated into the software • Ensure PETs API, SDK functionality are tested prior to integration into existing data pipelines • Ensure data quality/privacy of the PETs by verifying that the output data corresponds to desired PETs application of the test input data <u>Deployment Checks</u> • Confirm and validate that hosting model (for e.g. cloud, on-prem or hybrid) aligns with the organisation's ICT policies • Ensure the PETs solution or integration meets the organsiation's cybersecurity baselines through assurance testing such as Vulnerability Assessment and Penetration Testing • Ensure the deployed PET meets its intended expectations by fulfilling the User-Acceptance-Test (UAT) conducted with users
	Legal/Data Protection Related
3	**Data protection measures** • Ensure only necessary data is processed and minimise unnecessary exposure of personal or sensitive data • Verify that adequate safeguards are in place to prevent re-identification of personal data
4	**Outlier detection and sample size (for model training use cases)** • Ensure that outliers are removed before model training

	• Ensure that there is sufficiently large sample size to reduce re-identification risk

5	**Assess if PET goals meant to remove personal data are achieved**
	
	<u>Remove direct identifiers</u>
	• Identify all direct identifiers in the input dataset and verify if direct identifiers are still present in the output dataset
	• Understand the processing taken to remove direct identifiers from the input data. Removal of direct identifiers may include other techniques rather than relying entirely on a PET's based technique
	• Check that direct identifiers from the input data is no longer present in the output data
	
	<u>Hide identifiers within datasets</u>
	• When identifiers are still present, they should be hidden and obscure any relationship between direct and indirect identifiers in the dataset
	• Ensure direct or indirect identifiers present and needed in the dataset should be hidden through 1) obfuscation, or 2) splitting of data, to disassociate the linkages between them
	• Ensure identifiers are properly encrypted and keys used are appropriately secured, rotated and managed
	• When identifiers are hashed, consider adding random values known as "salts" to the identifiers prior to hashing to minimize risk of re-identification through matching of hashes to precomputed hash tables
	• Ensure hidden identifiers do not retain a relationship that can map back to the original identifier
	  • If relationship mapping is retained, there must be valid purpose on its need. These mappings must be securely stored and preferably out-of-band and away from these hidden identifiers.
	• Ensure hidden identifiers remain hidden in their encrypted or split form for the duration of the intended purpose
	  • Should identifiers ever be decrypted (when they are encrypted) or joint back (when they are split), they must undergo a conversion process that protects its hidden form into other hidden forms or computed as an aggregated output (that can no longer distinguish which individual contributed into the aggregation output)
	  • Ensure Identifiers that are meant to be hidden should never be allowed to be assessed in the clear
	• Ensure pseudonymised identifiers do not reveal its relationship back to the original direct identifier or use it to link back to the individual
	
	<u>Prevent linkages across other datasets</u>
	• Ensure present identifiers in the dataset should not have any relationship with identifiers of other datasets

	• When indirect identifiers are still present for each data record in the dataset, consider further PET techniques are often used to prevent linkages to other datasets especially when there is a need to release or disclose the data. They may include Differential Privacy or Synthetic Data Generation • When use of further PETs may be impractical, consider applying other techniques to reduce identifiable linkages to other datasets, e.g. generalisation and data aggregation on indirect identifiers to make the results indistinguishable from individual records from the dataset • Subsequently evaluate the risk of re-identification through methods such as k-anonymity. • Ensure direct identifiers are removed or hidden and remain hidden for their entire lifecycle. Its hidden form (e.g. encrypted, hashed or split) should not be associated with or linked to any of its associated indirect identifiers • Review the need to persistently store any direct identifiers within the dataset and only retain identifiers in the dataset for its intended purpose
6	**Addressing residual risks if resultant data still contains personal data** <u>Purpose of Processing</u> • Ensure compliance with purpose limitation obligation (of the PDPA), and check on applicability of any exceptions which may apply <u>Role of Organisation</u> • Depending on its role, e.g. data controller versus intermediary with respect to the data processing, to clarify which PDPA obligations would apply <u>Cross Border Data Transfers</u> • The Transfer limitation Obligation would apply if personal data were to be transferred to a third party. Consider putting in place relevant transfer mechanisms such as binding corporate rules, contractual clauses, etc

3) <u>Post Implementation:</u>

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.

	Business related
1	**Evaluating trade-offs:** <u>Performance versus Privacy:</u> • Quantify the trade-offs between performance and privacy protection. Assess if the level of encryption or data protection is causing any trade-offs in the

	accuracy or speed of results. E.g., compare performance of models using encrypted versus raw data <u>Scalability:</u> • Assess if the PETs solution can scale to meet growing business needs without impacting user experience • Establish acceptable latency and compute thresholds for production environments
	Technology related
2	**Continuous Monitoring and Optimization:** <u>System Performance:</u> • Track performance metrics such as speed, latency, and response time to ensure the solution does not over burden company infrastructure • Implement automated monitoring to catch any issues early <u>Feedback Loops:</u> • Consider a feedback mechanism for users to optimise and improve the effectiveness of the PETs solution and to identify potential errors <u>Incident Reporting Mechanisms:</u> • Ensure proper incident reporting mechanisms and protocols are in place <u>Third-Party Checks:</u> • Consider engaging independent third parties to audit the PETs implementation periodically, ensuring that compliance standards are continuously met <u>Re-identification Reassessment:</u> • Periodically reassess the risk of re-identification in PET-processed data outputs, considering advances in technology and the availability of new auxiliary data sources. • Reassess risk of re-identification upon major system changes, when there are changes to external data releases, or updates in regulatory guidance.
3	**Documentation and Reporting:** <u>Audit Logs:</u> • Ensure that audit logs are consistently maintained, capturing key information on who accessed what data, when and why. This would be crucial for maintaining compliance and identifying potential issues quickly <u>Documentation for Compliance:</u>

	• Maintain up-to-date documentation on the implementation process, privacy safeguards, and compliance measures taken throughout the deployment lifecycle of the PET. • Record and justify all key threshold settings and configuration choices made for privacy (e.g. epsilon in differential privacy, k-value in k-anonymity,) made during the PET implementation.
	Legal / Data Protection related
4	**Data Privacy checks** <u>Data Privacy Compliance:</u> • Regularly audit (e.g. annually) deployed PET to ensure compliance with the applicable data protection regulations <u>Regular Privacy Protection Assessments:</u> • Periodically assess the dataset for any potential exposure of personal data (e.g. annual), especially after system updates or infrastructure changes