הרשות להגנת הפרטיות
THE PRIVACY PROTECTION AUTHORITY
سلطة حماية الخصوصية

משרד המשפטים
MINISTRY OF JUSTICE | وزارة العدل

Guide on Implementing Privacy-Enhancing Technologies in Artificial Intelligence Systems

December 2025

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 ת.ד. קרית הממשלה,

Guide for Implementing Privacy Enhancing Technologies in Artificial Intelligence Systems

Introduction

An Artificial Intelligence (AI) system means "a machine-based system that infers from the input it is fed how to produce predictions, content, recommendations or decisions that can affect the individual or the activities of data controller or data processor, operating with varying levels of independence and adaptability[1]."

The use of AI systems introduces significant challenges to privacy protection[2]. AI systems rely on processing[3] large volumes of data, some of which may qualify as personal data[4]. This information is essential throughout the AI system's lifecycle – during the build phase of new systems (including development and training), and during the use phase, including system refinement and performance improvement over time[5].

Privacy-Enhancing Technologies (PETs) are a set of methods, processes and digital tools designed to support the protection of personal data[6]. These technologies enable a

[1] Definition of an Artificial Intelligence System in the draft directive of the Privacy Protection Authority on the subject of "Applicability of the provisions of the Privacy Protection Law to Artificial Intelligence Systems" (April 2025), which was published for public consultation (Hebrew). Compare to a definition by OECD: "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment", OECD Council Recommendation on Artificial Intelligence (amended 3 May 2024).

[2] See the Interim Report of the Inter-Ministerial Team on Artificial Intelligence in the Financial Sector (2024), pp. 72-93. See also: Michael Birnhack, Privacy and Artificial Intelligence, Social and Cultural Law, 8 (2024), Hebrew.

[3] Processing as defined in Section 3 of Amendment No. 13 to the Israeli Privacy Protection Law, 5741-1981 – ""Processing", "Use" – any action performed on personal information, including its receipt, collection, storage, copying, consultation, disclosure, disclosure, transfer, delivery or provision of access to it".

[4] The legal definition of "personal data", which is the cornerstone definition of Israeli Privacy Protection Law, 5781-1981, was entirely amended within Amendment No. 13 to the Privacy Law Protection, enacted in 2024. According to the new definition, personal data are defined as : "Personal data" – data relating to an identified or identifiable person; for the purposes of this definition, an "identifiable person" is one who can be identified with reasonable effort, directly or indirectly, including through an identifying detail such as name, ID number, biometric identifier, location data, online identifier, or one or more details relating to their physical, health, economic, social or cultural status."

[5] Explanatory Memorandum on The Updated OECD Definition of an AI System, OECD Artificial Intelligence Papers, No. 8 (March 2024).

[6] Further elaboration may be found in Guide for Privacy-Enhancing Technologies, Israel Privacy Protection Authority, 2025.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

balance between the powerful capabilities of AI systems and the protection of users' privacy. Implementing PETs can help mitigate privacy risks across all phases of the AI system lifecycle.

Objectives of This Document

1. To present approaches to mitigating privacy risks associated with the development and use of artificial intelligence systems through privacy-enhancing technologies.
2. To illustrate examples of practical application of these approaches across different sectors.

Target Audience

This document is intended for professionals responsible for assessing privacy risks and implementing appropriate solutions in projects involving the development of digital systems and services that incorporate AI components. It also can serve as a resource for developers in AI domains, offering guidance on embedding PETs from the early stages of project initiation through the system's entire lifecycle. In particular, this guide is relevant for:

- Data Protection Officers (DPOs)[7] and legal advisors engaging in privacy issues in AI-integrated projects.
- Product managers and project managers involved in the development, deployment, and operation of AI-based systems within digital products and services.

The use of the document does not require a technical background or technological expertise. Accordingly, the level of detail provided for each technology is intended to offer an understanding of its essence and to facilitate an initial evaluation of its suitability for specific fields or tasks. It should be noted that this document does not aim to exhaustively describe the full complexity of AI systems but focuses on the aspects necessary for understanding the application of PETs in this context.

[7] See the draft opinion statement (Hebrew) of the Privacy Protection Authority on the subject of "Appointment of a Privacy Protection Officer in an Organization According to the Requirements of Amendment 13 to the Privacy Protection Law," which was published for public comments.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

For practical implementation, including handling challenges and potential risks, users are encouraged to consult additional resources. Throughout the document, references and links to further information are provided. It should be noted that for the purpose of examining the appropriate architecture for a project or technological product, this document does not replace consultation with the relevant experts.

Presentation of Technologies and their Application

This guide describes Privacy-Enhancing Technologies and their applications within AI systems, explaining the operating principles that support privacy protection objectives. Each technology is accompanied, where relevant, by examples or diagrams that illustrate its core mechanisms.

The document further outlines key considerations for practical implementation of each technology, alongside examples of AI-related applications from diverse domains and fields wherever possible. It also addresses challenges, limitations, and risks associated with each technology, highlighting specific areas requiring careful attention.

The examples included are designed to simplify and clarify the underlying principles of the technologies in AI contexts. They are intended for illustrative purposes only and do not constitute recommendations for specific courses of action, nor do they preclude other possible uses.

Practical deployment of PETs in AI domain requires a thorough evaluation of multiple factors, including case- and project-specific features, technological contexts, legal framework, and organizational considerations.

Information Sources for This Document

The content and examples provided are based, in part, on several key sources regarding the implementation of PETs in AI systems, including:

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

1. The OECD policy paper on Sharing trustworthy AI models with privacy-enhancing technologies[8], June 2025.
2. Guidance – Application of the Privacy Protection Law to Artificial Intelligence Systems (Hebrew), Draft for Public Comments, Israel Privacy Protection Authority, April 2025.
3. Guide to Risk Management and Responsible Use of Artificial Intelligence (AI) Tools in the Public Sector (Hebrew), Public Comments Version, Israel National Digital Authority, June 2025.
4. The UK Government's Repository of Privacy Enhancing Technologies (PETs) Use Cases[9].
5. Centre for Informational Policy Leadership's survey on Understanding the Role of PETs and PPTs in the Digital Age[10].
6. UK Centre for Data Ethics and Innovation's Repository of Use Cases[11].
7. UN Guide on Privacy-Enhancing Technologies for Official Statistics' Case Studies Repository[12].

References to additional sources are indicated in the footnotes and in the body of the document.

[8] OECD (2025), "Sharing trustworthy AI models with privacy-enhancing technologies", OECD Artificial Intelligence Papers, No. 38, OECD Publishing, Paris, https://doi.org/10.1787/a266160b-en.
[9] Repository of Privacy Enhancing Technologies (PETs) Use Cases, Updated: November 7, 2024.
[10] Privacy-Enhancing and Privacy Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, December 2023.
[11] Centre for Data Ethics and Innovation Repository of Use Cases, Updated June 2023.
[12] United Nations Global Working Group Task Team on Privacy Preserving Techniques: Case Study Repository, last modified on Apr 11, 2024.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

Table of Contents

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

Personal Data in Artificial Intelligence Systems

Introduction to Artificial Intelligence Systems

An artificial intelligence (AI) system is composed of algorithms and mathematical models that enable it to operate at varying levels of autonomy. The system is based, among other elements, on machine learning algorithms that process and analyze data inputs stored in databases, which the system then processes, analyzes, and uses to generate outputs[13].

The uniqueness of such a system lies in its ability to emulate human cognitive processes that were previously considered exclusive to human beings. Typically, the system operates in response to textual or voice commands (prompts) provided by human users. There are several types of artificial intelligence, each focusing on a different domain of application and delivering distinct outcomes. Among these types, generative artificial intelligence (Generative AI) stands out in recent years, enabling the creation of seemingly original outputs such as texts, videos, and images.

The Lifecycle of Artificial Intelligence Systems[14]

An artificial intelligence system typically comprises one or more models[15] developed based on input data or operator instructions. AI models may include statistical representations of inputs that facilitate the generation of desired outputs under appropriate conditions and scenarios. An AI model may enable the system to select a preferred action by evaluating the predicted consequences inferred from the input data. Models can be developed either manually by human programmers or automatically, for example, through machine learning algorithms and decision-making processes.

The lifecycle of an AI system includes several phases intended to build and optimize the use of the model. This document focuses on two key phases[16]:

[13] Ministry of Innovation, Science and Technology, Principles of policy, regulation, and ethics in the artificial intelligence domain, 2023 (Hebrew), pp. 21-22.
[14] The chapter is based on the OECD Explanatory Memorandum to the Updated Definition of Artificial Intelligence, above footnote #5.
[15] According to ISO/IEC 22989:2022: Model: physical, mathematical or otherwise logical representation of a system, entity, phenomenon, process or data (Information technology - Artificial intelligence - Artificial intelligence concepts and terminology).
[16] Based on the distinction made in above footnote #5.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

1. **Build Phase:** This is the phase in which the new system is developed and trained. During the build phase, the objectives and functions of the system are defined. The selection of mathematical models is made according to the types of tasks the system is intended to carry out, with each model adapted to the nature of the data and the required operations.

   The process includes building a comprehensive dataset that will serve as the foundation for model training. This dataset should encompass a wide range of scenarios in which the system is expected to operate, or the types of outputs it will be required to generate. The system is trained on the provided data, analyzing it and learning to recognize patterns, trends, and correlations among different factors, through learning and inference based on the input. The input may include task-relevant data, user requests, or search queries.

   To ensure high-quality outcomes, the input must be accurate, balanced, representative, and relevant to the intended purpose. Throughout and following the build process, the model's performance is evaluated to verify that it produces results aligned with the defined objectives. Developers review the outputs generated by the model, assess their accuracy and reliability, and perform necessary adjustments as needed. This phase is essential to ensure that the model can perform its tasks optimally and efficiently.

2. **Use Phase**: At this phase, the system is made available for users, who input prompts and specific instructions to guide the system in producing personalized outputs (such as recommendations, forecasts, or decisions). The system receives new input data, for example, user information, texts, or images, and applies the pre-trained models to analyze, predict, recommend, decide, or generate new content. The system's performance at this phase is influenced both by the manner in which it was designed and developed, and by the quality and accuracy of the input it receives.

   Example: A clinical decision-support AI system might identify a suspicious mass in a medical imaging scan, recommend further diagnostic tests or a biopsy, or automatically alert the medical team.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 .ת.ד ,קרית הממשלה

The build and use phases may overlap: During the build phase, in some cases, it is necessary to reuse input data to improve performance and accuracy by adjusting and updating the model. In other cases, during the use phase, new input data (such as user queries or actual results) are used build (update and improve the model), either continuously or at defined intervals.

Personal Data in the Development and Use of Artificial Intelligence Systems

During both the development and use phases of an AI system, the input may include personal information. Such personal information can influence the model's behavior and may even become embedded within it, potentially leading to the exposure of personal data through the system's outputs. Accordingly, the protection of personal information in AI systems addresses the following key areas:

1. **Training Data**: Personal information used to train the system. Examples include medical records, user data, or financial transactions.
2. **Trained Model**: Personal information that becomes embedded within the model following training. Under certain circumstances, recovering personal information from the model may be possible by inferring typical input data based on the model's responses (Model Inversion) and other methods.
3. **User Input**: Personal information entered into the system during its use. Examples include free-text queries, service requests, personal images, or text. User inputs are processed by the model and may be used for model training or updating to enhance future performance.
4. **Model Output**: Personal information produced by the system, whether derived from the input, learned from prior data, or inadvertently exposed. Examples include personalized responses, recommendations based on personal profiles, or information unintentionally extracted from the model.

Figure 1 illustrates the domains of personal information protection across the lifecycle phases of an AI system.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 .ת.ד ,קרית הממשלה

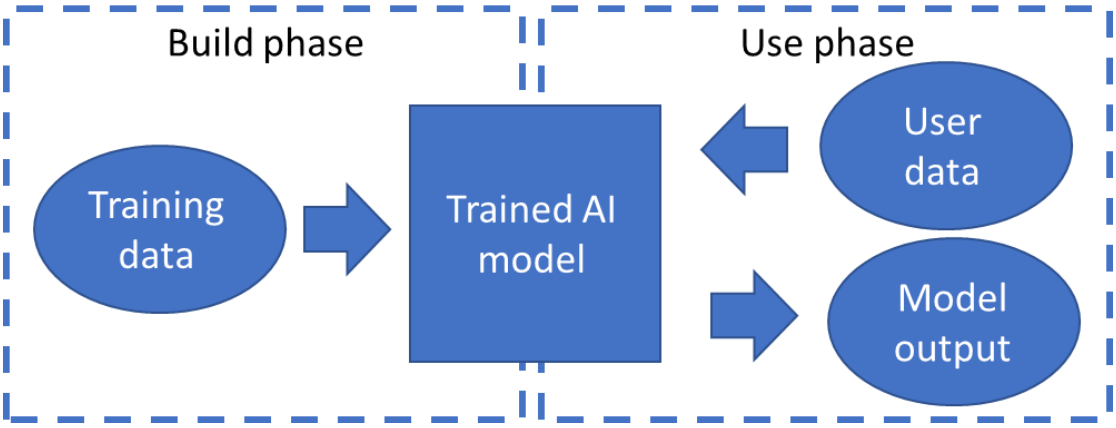

Figure 1: Domains of personal information protection across the lifecycle phases of an AI system

Protecting Personal Data in Artificial Intelligence Systems

Privacy-Enhancing Technologies

Privacy-Enhancing Technologies (PETs) are a set of methods, processes and digital tools designed to support the protection of personal data. PETs allow to obfuscate personal data and reduce its level of details, reduce the risk of exposure of personal data during processing, and enable greater control over how personal data are used[17].

The principles, practices, and detailed overview of PETs types were presented extensively in the Guide to Privacy-Enhancing Technologies published by the Privacy Protection Authority[18]. This document focuses on technologies that are particularly relevant and applicable to artificial intelligence systems, grouped into three main categories. Each category reflects a distinct operational principle aimed at mitigating privacy risks in the context of AI systems development and deployment.

The first category of privacy-enhancing technologies operates by the principle of **data transformation**. This group comprises a range of techniques designed to obscure, remove, or replace identifiable characteristics from raw data. Common examples include the removal of direct identifiers such as names or national ID numbers; the masking of indirect identifiers (e.g. precise geographic locations, exact dates of birth, or other unique attributes); value rounding; the addition of random noise; producing synthetic data that is similar in statistical structure to the original data but without

[17] Definition from Guide to Privacy-Enhancing Technologies, Israel Privacy Protection Authority, 2025.
[18] See the previous footnote for a link to the document.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

preserving real records; and, in some cases, the generalisation of values to reduce data granularity (for instance, aggregating specific ages into broader age ranges).

The core principle supporting this category is the direct transformation of raw data in a manner that prevents the original identification of individuals to whom the data relates, while preserving a certain level of utility. This reflects a delicate balance: on the one hand, there is a need to reduce privacy risks by minimising the likelihood of re-identification; on the other, it is important to retain sufficient data utility to enable effective training of AI systems.

Techniques within this category include:

1. **Anonymization**: Data anonymization is the removal of features or alteration of values in order to reduce or prevent identification of the data subject or the ability to recover personal information from the training input or from the trained model. Applying anonymization to training data hinders the ability to link specific data entries to individuals, thereby reducing the risk of re-identification from the trained model. One common approach to anonymization involves rounding values or grouping them into ranges, for example, representing age as a range (e.g. 40–49) instead of an exact value (e.g. 47), or using income range in place of precise monthly income figures. This reduces the risk of re-identification but also decreases the granularity of the data.
2. **Synthetic Data**: Synthetic data refers to artificially generated data designed to mimic "real" data, based on the statistical properties of genuine datasets, but is not supposed to preserve the original details of any person or event. In other words, it is fictional yet structurally accurate data intended to serve as a substitute for sensitive data, such as during the development, testing, or training of AI systems. Synthetic data allows artificial intelligence models to be trained while reducing the risk of exposure to personally identifiable information.
3. **Differential Privacy**: This technique limits the impact of any single data point on the learning process, preventing the identification or reconstruction of personal information even if an adversary has access to the model's outputs or external knowledge. By applying differential privacy, the model substantially

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

reduces its dependency on any single identifiable example, making it extremely difficult to reconstruct individual information.

The main advantage of this category lies in its relative simplicity and the ability to implement the techniques without requiring extensive computational resources or collaboration across multiple entities. However, data transformation may adversely affect model quality. When noise is added or generalization is applied, the data loses precision or granularity, which can, in turn, impact the performance of AI models. For example, a model trained using age ranges may be less accurate than one trained on precise age data.[19]

The second category of privacy-enhancing technologies operates on the principle of **distributed computation**. This approach is designed to enable systems to perform calculations on sensitive data without exposing the underlying information, leveraging advanced cryptographic methods and distributed processing techniques. Unlike the data transformation principle, where the data itself is altered or pre-processed to ensure protection, this category focuses on how data is used and processed without modifying the data itself, thereby preserving accuracy and utility.

The core principle of this approach is the distribution of data or computation across multiple entities or sources, ensuring that no single party holds the complete set of sensitive information. Rather than aggregating all data on a central server, the data remains in situ, for example, on the user's device, across multiple servers, or with organizational partners, and computations are performed in a distributed or encrypted manner. Often, only the final result of the computation is collected, without exposing or even transferring the underlying data itself.

Techniques within this category include:

[19] The opposite may also be true in certain cases, where the use of data that differs from precise personal information can actually enhance model performance. For example, training an AI system on a large volume of synthetic data, which does not represent the personal information of real data subjects, may lead to improved inference and generalization capabilities compared to training on a limited set of real personal data.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

1. **Federated Learning**: Training AI models without aggregating all input data in a central location. Instead, the model is distributed to local devices or servers, updated there based on local data, and only the model updates (not the raw data) are returned to the central server for aggregation into an improved version of the model.

   Federated learning, for example, enables local processing on users' devices, such as smartphones, laptops, or other smart devices. In this model, personal data remains on the device and is not exposed to the central server or other users. For instance, in an auto-completion system for search queries based on user input, rather than transmitting the typed text itself, the device sends only model updates to the central server, and not users' data. It should be noted that the data subject's raw personal information is not disclosed to third parties, but it is used to train the model and the model is updated in such a way that it can influence the prediction results and even produce the same personal information in the output.
2. **Secure Multiparty Computation (SMPC)**: Enabling multiple parties to collaboratively compute a result without disclosing their private or sensitive data to each other. Each party retains its own data and performs calculations without revealing them to any other party. Using SMPC techniques, organizations can collaboratively train AI models on sensitive datasets without sharing the underlying data.

   This method enables cooperation between organizations (e.g., corporations, financial institutions, or healthcare providers) to develop more accurate models without exposing sensitive customer or user information, or work on information that cannot be transferred directly from one place to another. For example, multiple hospitals can collaboratively compute a medical metric across their patient populations without sharing patients' sensitive medical data with one another.
3. **Private Set Intersection (PSI)**: Allowing comparison of datasets without exposing personal information unique to each party. Each party can determine

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

which entries are shared (e.g., lists of items, users, or data) without revealing non-shared items.

For example, let us assume that within the scope of its powers, the Ministry of Welfare wishes to contact the Ministry of Tourism to find out about periods of long stay abroad of people eligible for a pension, it can transfer the identity cards of those eligible for a pension to the Ministry of Tourism for review. In such a case, information about all pension beneficiaries would be transferred unnecessarily, while by private set intersection it is possible to reduce the transfer of information to only those of them who have actually been abroad for a long period of time.

The use of distributed computation offers significant advantages in privacy protection contexts, particularly when handling sensitive information, such as medical, biometric, or financial data, or when collaborating across multiple partners, including public institutions, regulatory bodies, or various service providers. This approach enables the extraction of value from data while minimizing exposure, reducing the risk of data breaches, and preventing centralization of information in a single location. However, these systems also present challenges: they typically require more complex coordination and incur higher computational costs. Additionally, advanced infrastructure and common standards are necessary to ensure effective collaboration among participating organizations.

The third category of privacy-enhancing technologies operates on the principle of **data separation and encryption**. This group employs advanced techniques designed to protect personal data during processing, not only before or after, by using strong encryption or physical/logical isolation of sensitive processes, enabling secure data handling.

While data transformation methods modify data prior to system intake, and distributed computation approaches keep different portions of data with separate entities, separation and encryption techniques allow data to be processed in a centralized manner under mechanisms that ensure the data remains inaccessible to unauthorized parties even during processing. These methods provide protection for data in use,

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ,ת.ד. קרית הממשלה

acknowledging that artificial intelligence often requires centralized computational power (such as cloud servers), where full control over the execution environment may not rest with the data owner.

Techniques within this category include:

1. **Homomorphic Encryption**: An encryption technique allowing computations on encrypted data without decrypting it. AI models can be trained on encrypted data, ensuring that the data remains protected throughout the training process. This way, both the trained model and its outputs can remain encrypted, accessible only to authorized users. For example, an insurance company can transmit encrypted medical data to a cloud service provider for the purpose of training a predictive model, without exposing the underlying information. The provider performs computations on the encrypted data, and the final result can be decrypted only by the company.
2. **Trusted Execution Environment (TEE)**: Processing data within an isolated and secure hardware environment designed to prevent unauthorized access or data tampering. Both the trained model and the input data can be maintained within the TEE after training is completed, ensuring continuous data protection. In such cases, the model remains protected from external access, and personal data is safeguarded against leakage or exposure. For example, a healthcare provider can train a disease detection model within a trusted execution environment, ensuring that sensitive data remains inaccessible outside this secure area. Even after training, the model can remain within the trusted environment and be accessible only for authorized uses, so preserving patient privacy.

The principal advantage of technologies in this category lies in their ability to enable full processing of personal data while protecting privacy. Specifically, these technologies facilitate the training and operation of sensitive artificial intelligence systems on confidential data, even when hosted on untrusted servers. However, technologies such as homomorphic encryption demand significantly greater processing time and memory resources, and they are not supported across all development

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 ת.ד. קרית הממשלה,

environments. Trusted execution environments require dedicated hardware and software components and do not offer absolute protection against breaches, given the capabilities of current or future attack methods.

The decision to implement privacy-enhancing technologies across training inputs, the trained model, user inputs, and system outputs depends on a range of considerations, including the sensitivity and volume of data, as well as the design and intended use of the specific system. Moreover, some technologies address broad challenges in the development and deployment of AI systems, while others are tailored to specific aspects of personal data protection within AI. The diverse operational principles of privacy-enhancing technologies have varying impacts on training data, the trained model, user inputs, and system outputs, and therefore the choice of a particular technology should take these factors into account.

To assist in aligning the operational principles of privacy-enhancing technologies with specific application domains, the following mapping is proposed. According to their categories and primary modes of operation within artificial intelligence systems, these are provided in Table 1.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 .ת.ד ,קרית הממשלה

Table 1: Mapping of PETs by groups and their core principles in the context of AI systems

Group	PETs	Training Input	Trained Model	User Input	System Output
Data Transformation	**Anonymisation**: Removing and blurring identifying features in data	Obfuscation of personal data and reduction of their level of detail	Derived from the protection of the training input	Limited applicability	Obfuscation of personal data and reduction of their level of detail
	Synthetic data: Using fictional data as a substitute for sensitive data				
	Differential privacy: introducing noise that obscures personal information				
Distributed Computation	**Multi-party computation**: Decentralizing personal data to multiple parties without exposing personal or sensitive data	Reducing the exposure of personal information during use	Limited applicability	Limited applicability	Limited applicability
	Private set intersection: comparing data sets without revealing information that is only available to one party.				
	Federated learning: local training of AI models without centralizing all input data				
Separation and Encryption	**Homomorphic encryption**: performing calculations on data while they are encrypted	Processing data while it is encrypted **at all stages** along the chain of building and using an AI system			
	Trusted execution environment: Data processing in an isolated and secure part of the computer system	Secure data processing in an encrypted environment **at one or more stages** along the chain of building and using an AI system			

Combinations of Privacy-Enhancing Technologies

The diverse operational principles of privacy-enhancing technologies (PETs), along with the range of technical implementations available for each principle, enable numerous combinations to achieve a high level of personal data protection and to establish a multi-layered, context-aware, and flexible data protection framework. The next section of the document provides examples of PETs implementation in artificial intelligence systems. In these examples several common combinations of complementary technologies can be identified. These combinations are applied as part of an integrated process tailored to the type of personal data and its specific processing characteristics. Such combinations are often developed in a modular fashion, taking

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

into account the nature of the data, its sensitivity level, applicable regulatory context, and the types of processing anticipated. Examples include:

1. **Differential Privacy + Trusted Execution Environment**: Differential privacy introduces random noise into data outputs to ensure that no individual's information can be inferred from the result. A trusted execution environment (TEE) creates an isolated and secure enclave within hardware or software that prevents unauthorized access, even by the operating system, to the data or the code running inside it. When combined, these technologies allow the noise injection process to occur entirely within the TEE. As a result, not only are the sensitive data protected, but the privacy-preserving computation itself is also secured from external interference. The computation is fully executed inside the isolated environment, and only a differentially private output is released from the enclave.

   This combination reduces the risk of data leakage during computation or training processes and offers dual protection, against both data exposure and manipulation or attacks during the noise injection phase.

   For example, suppose a Ministry of Health seeks to analyze trends in pharmaceutical consumption across the population. However, the data originates from multiple healthcare providers and cannot be shared in plain form. In this case, each provider performs local computation within a trusted execution environment, adds noise to the results (using differential privacy), and submits only the aggregated summary. The statistical information can then be securely combined and analyzed at a central level, enabling public health insights without exposing any individual patient's data.
2. **Synthetic Data + Private Set Intersection:** Synthetic data is artificially generated data that mimics the structure and statistical properties of real datasets without being tied to any identifiable individual. Record linkage techniques enable two or more parties to identify overlapping entries, such as shared national ID numbers, across their respective databases, without revealing any non-matching values. By combining these approaches, organizations can detect only the relevant or overlapping records between them and subsequently replace

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 ת.ד. ,קרית הממשלה

the real data with synthetic data for those records. This ensures that the AI model is trained only on pre-approved observations, while maintaining strong privacy protection by avoiding the use of identifiable personal data.

This combination allows for identity protection both at the record linkage stage and during subsequent data use. It provides significant flexibility in analyzing targeted subpopulations without exposing personal data, and facilitates collaboration between organizations without the need to transfer raw datasets.

Example: A financial supervisory authority seeks to determine whether certain high-risk groups are receiving loans from multiple banks. Each participating bank uses a privacy-preserving record linkage mechanism to identify only the shared customers. Synthetic data is then generated for those overlapping records to train a risk assessment model, without exposing or sharing real personal information.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

Examples of implementing PETs in artificial intelligence systems

Overview

Organization	Country	Data	Application	Technology	Area	Year	Source
Statistics Canada[20]	Canada	Consumer data	Developing a model for text classification	Homomorphic encryption	AA	2021	Link
Secretarium[20]	Denmark	Financial data	Sharing financial information without exposing it between participants	Multi party computation, trusted execution environment	AA	2023	Link
Statice[21]	Germany	Financial data	Using synthetic data to train a model	Synthetic data	TI	2023	Link
Statistics Netherlands[20]	The Netherlands	Healthcare	Developing a model for predicting cardiovascular risk from distributed data	Multi party computation, federated learning, homomorphic encryption	AA	2020	Link
Frontier Development Lab / Intel[21]	US	Healthcare	Developing a model for analyzing the relationship between space radiation exposure and cancer	Federated learning	TI	2021	Link
United Nations Economic Commission for Europe[20]	UN	Healthcare	Developing a model based on lifestyle information collected from mobile devices	Federated learning, differential privacy, homomorphic encryption	TI	2023	Link
Twitter and OpenMined[20]	US	Consumer data	Research on user data without exposing the data itself	Federated learning, differential privacy, multi-party computation	TI MO	2023	Link

Legend:	AA: All areas	MO: Model output	TM: Trained model	UI: User input	UI: User input

[20] Referred from UN site: UN GWG Task Team on Privacy Preserving Techniques - Case Study Repository
[21] Referred from ICO site: Repository of Privacy Enhancing Technologies (PETs) Use Cases

Details regarding the examples from the relevant information sources

It should be noted with respect to all examples that the details are based on the description that appears on the websites of the UN, ICO, and the relevant sources from which the examples are taken, and this description does not constitute an opinion or recommendation on how to protect personal information.

Organization	Country	Data	Application	Technology	Area	Year	Source
Statistics Canada	Canada	Consumer data	Developing a model for text classification	Homomorphic encryption	**AA**	2021	Link

Overview

Statistics Canada conducted a proof of concept for training a machine learning model for text classification in the cloud while ensuring data privacy through the use of homomorphic encryption.

Process

a. The input data for the AI system training were encrypted using homomorphic encryption.
b. An AI system was developed on a remote server and trained on the encrypted data transmitted to it, such that the model parameters (neural network weights) remained encrypted throughout the process.
c. The AI system was used with user input which was also encrypted.

Outcomes

a. The input data and user data used by the AI system remained encrypted and were not exposed at any stage during processing.
b. The AI model's internal parameters (neural network weights) remained encrypted and were never exposed during processing.
c. The accuracy of the results achieved, despite the approximations introduced by encrypted computation, was comparable to that of training and use without encryption.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 .ת.ד ,קרית הממשלה

Organization	Country	Data	Application	Technology	Area	Year	Source
Secretarium	Denmark	Financial data	Sharing financial information without exposing it between participants	Multi party computation, trusted execution environment	**AA**	2023	Link

Overview

The DANIE consortium brings together banks and data providers collaborating on a shared platform to which banking data is uploaded for analysis. The consortium's primary objectives include: (1) improving customer data quality, (2) preventing money laundering, and (3) detecting fraud. Launched in 2020, the platform employs advanced encryption technologies and trusted execution environments (TEEs), ensuring that data undergoing processing remains confidential even from the users themselves.

DANIE relies on data privacy protection solutions developed by Secretarium. Both initiatives (Secretarium and DANIE) originated within the Société Générale innovation incubation program in London.

Participation in this initiative provides organizations with significant benefits, including: compliance with GDPR and avoidance of penalties; resource savings by reducing the need for data validation and correction efforts; and improved performance and efficiency of data analysis processes, enabled by a centralized and streamlined data processing system.

Process

1. The consortium employs secure computing and cryptographic technologies to enable collaboration between financial institutions without exposing sensitive data.
2. Since 2018, international banks have utilized technologies available through the consortium to process and reconcile millions of records.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

3. Within the consortium, trusted execution environments are used to process and reconcile sensitive data, ensuring privacy between participating organizations in a secure computing network.

Outcomes

1. Data sovereignty preserved: Each party maintains full control over its own data without disclosing it to other participants.
2. Full encryption and verifiable trust: All data remains encrypted at all times, with verifiable guarantees and full auditability.
3. Efficiency and scalability: The system supports billions of transactions, features a user-friendly interface, and includes robust error-handling mechanisms to enhance data quality.

Organization	Country	Data	Application	Technology	Area	Year	Source
Statice	Germany	Financial data	Using synthetic data to train a model	Synthetic data	**TI**	2023	Link

Overview

The German insurance services provider Provinzial partnered with data privacy solutions company Statice to use synthetic data for training machine learning models aimed at enhancing their predictive analytics capabilities, particularly for their "next-best-offer" recommendation engine. The initiative resulted in savings of over three months in data privacy risk assessments that were avoided thanks to the use of synthetic data.

Process

1. Generation of synthetic data derived from the insurance company's existing data repositories, preserving the statistical properties of the original datasets.
2. Training of a predictive analytics model using the generated synthetic data to identify patterns and make accurate forecasts.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

3. Benchmarking the performance of the model trained on synthetic data against a model trained on original data to evaluate effectiveness.

Outcomes

1. The synthetic data matched the statistical characteristics of the original data, enabling model training without exposing sensitive personal data.
2. The model trained on synthetic data demonstrated performance comparable to the one trained on real data, indicating the high quality of the synthetic dataset.
3. The use of synthetic data allowed Provinzial to enhance its predictive analytics processes while maintaining compliance with regulatory requirements.

Organization	Country	Data	Application	Technology	Area	Year	Source
Statistics Netherlands[20]	The Netherlands	Healthcare	Developing a model for predicting cardiovascular risk from distributed data	Multi party computation, federated learning, homomorphic encryption	**AA**	2020	Link

Overview

The statistics bureau of Netherlands (Statistics Netherlands – CBS) collaborated with several organizations on the CARRIER project (short for Coronary ARtery disease: Risk estimations and Interventions for prevention and EaRly detection). The project aimed to develop models for predicting cardiovascular disease risk while protecting the privacy of sensitive medical data used in the process.

Process

1. Data Collection: The project utilized diverse data sources, including primary care data (clinics), secondary care data (hospitals), and socio-economic information.
2. Privacy-Enhancing Technologies: To ensure the protection of personal data, advanced techniques were employed, such as multi-party computation (MPC), homomorphic encryption, and federated learning. These methods enabled data analysis without revealing identifiable personal information.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 .ת.ד ,קרית הממשלה

3. Model Development: Machine learning algorithms were applied to encrypted datasets to construct accurate cardiovascular risk prediction models.

Outcomes

1. Federated Data Integration: The project successfully combined data from multiple organizations without exposing personal information, demonstrating the feasibility of inter-organizational collaboration with enhanced privacy protection.
2. Predictive Accuracy: The developed models achieved high predictive performance comparable to models trained on unencrypted data, showing that privacy-enhancing technologies can yield valuable insights without compromising data confidentiality.

Organization	Country	Data	Application	Technology	Area	Year	Source
Frontier Development Lab / Intel[21]	US	Healthcare	Developing a model for analyzing the relationship between space radiation exposure and cancer	Federated learning,	**TI**	2021	Link

Overview

Researchers from the Frontier Development Lab (FDL), in collaboration with mentors from Intel, conducted an innovative study aimed at better understanding the physiological effects of radiation exposure on astronauts. The study explored the link between space radiation exposure and the development of cancer, using a federated learning approach to access sensitive, protected astronaut data without exposing the data itself.

The key advantage of this approach was significant cost reduction. Traditionally, accessing such medical data requires major investment in data security infrastructure and complex bureaucratic processes. The use of federated learning enabled resource savings without compromising on high standards of data privacy.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב, 33503 ת.ד. קרית הממשלה

Process

1. Use of Federated Learning: The project employed Intel's OpenFL framework, which allowed models to be trained on local data without transferring sensitive data across institutions.
2. Data Integration from Multiple Sources: The data originated from NASA, the Mayo Clinic, and NASA's GeneLab, with all data remaining on-site and identifiable information protected.
3. Model Training and Aggregation: Models were trained and aggregated in a manner that ensured personal data remained confidential throughout all stages of learning and processing.

Outcomes

1. Participant Privacy Protection: Medical data never left the institutions that held it, preserving the privacy of astronauts and patients.
2. Reduced Legal and Ethical Barriers: The approach helped overcome major legal and ethical challenges, enabling broader collaboration between public and private institutions.
3. Successful Training of Accurate Models: The project demonstrated that advanced medical research can be carried out using privacy-enhancing technologies, without compromising data protection standards.

Organization	Country	Data	Application	Technology	Area	Year	Source
United Nations Economic Commission for Europe[20]	UN	Healthcare	Developing a model based on lifestyle information collected from mobile devices	Federated learning, differential privacy, homomorphic encryption	**TI**	2023	Link

Overview:

The United Nations Economic Commission for Europe (UNECE) initiated a pilot project to explore the application of federated learning, aiming to enable collaboration between national statistical offices in different countries without exposing personal

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

data. The project was conducted using a public dataset of human activity, consisting of accelerometer and gyroscope readings from smart devices. The data was divided into four subsets, one for each statistical office participating in the experiment.

Process:

1. An AI system was trained using federated learning across the four data repositories held by the participating statistical offices.
2. A simulated environment was built using open-source tools and libraries to detect and classify human activities into multiple categories based on accelerometer data collected from smart and wearable devices.
3. The simulation environment was used to evaluate both the level of privacy protection and the training results.

Outcomes:

1. Privacy Protection: The project successfully demonstrated that accurate AI models can be trained without transferring personal data.
2. Technical Feasibility: The project illustrated the feasibility of using federated learning technologies within the context of official statistics.
3. Coordination and Approvals: The project highlighted that real-world scenarios would require coordination and agreements among participating organizations.

Organization	Country	Data	Application	Technology	Area	Year	Source
Twitter and OpenMined[20]	US	Consumer data	Research on user data without exposing the data itself	Federated learning, differential privacy, multi party computation	**TI** **MO**	2023	Link

Overview:

Twitter and OpenMined collaborated to enable external researchers to work with internal data, even when the data itself could not be released. The project focused on building a secure environment that allows researchers to analyze unreleased digital

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

assets while protecting user privacy. The primary goal of the initiative was to enable high-quality scientific research without exposing sensitive personal data.

Process:

1. Building a secure and isolated environment: Twitter provided an execution infrastructure based on a secure technology environment that ensured no party (including cloud providers) could view the data being analyzed.
2. Use of open-source code: Researchers developed their data analysis code outside the environment. After Twitter reviewed and approved the code, it was executed within the secure environment.
3. Output of processed results: Only processed and statistical results were allowed to leave the secure environment, following a filtering process to ensure no personal data was disclosed.

Outcomes:

1. Protection of user privacy: There was no need to release the original data or expose it to the researchers: the access was tightly controlled and protected by technologies that prevent data leakage.
2. Ethical reproducibility of research: The project enabled researchers to replicate studies using the same original data under scientific transparency principles, without violating user privacy.
3. Demonstration of external auditability: The setup allows for external auditing of AI systems, even when regulatory or contractual constraints prevent the data from being publicly released.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה