Department for Science, Innovation & Technology

Republic of Estonia
Ministry of Economic Affairs and Communications

OECD
BETTER POLICIES FOR BETTER LIVES

United Kingdom - Estonia - OECD Expert Workshop on Privacy Enhancing Technologies and Artificial Intelligence

13 May 2024

Background Note

Drawing on relevant OECD work, this background note aims to support discussions at the United Kingdom - Estonia - OECD Expert Workshop on Privacy Enhancing Technologies (PETs) and Artificial Intelligence (AI) to be held on 13 May 2024. It serves as a preliminary reference throughout the expert workshop by:

- Introducing key concepts and terminologies associated with PETs and AI, ensuring a common vocabulary that supports dialogue among workshop participants.
- Contextualising these technologies within the broader framework of OECD's work, thereby aligning the discussion with international policy standards and practices.
- Presenting a small selection of use cases where PETs have been effectively implemented in the context of AI.

The content of this document is thus preliminary and subject to revision following discussions and feedback received during and after the workshop. It will serve as the foundation for developing an OECD paper on the topic, which will be further revised based on discussions at upcoming OECD expert workshop(s) on PETs, with the aim of finalisation by the fourth quarter of 2024.

Emerging PETs, their maturity, opportunities and challenges

PETs refer to a range of digital technologies and techniques that enable the collection, processing, analysis, and sharing of information while safeguarding data confidentiality and privacy. Although many emerging PETs are still in their early stages of development, they hold immense potential to advance privacy-by-design principles and foster trust in data sharing and re-use across organisations and sectors such as health and finance.

OECD (2023[1])[1] divides PETs into four categories: data obfuscation, encrypted data processing, federated and distributed analytics and data accountability tools.

- **Data obfuscation tools** include zero-knowledge proofs (ZKP), differential privacy, synthetic data, and anonymisation and pseudonymisation tools. These tools increase privacy protections by altering the data, by adding “noise” or by removing identifying details. Obfuscating data enables privacy-preserving machine learning and allows information verification (e.g., age verification) without requiring sensitive data disclosure. Data obfuscation tools can leak information if not implemented carefully however. Anonymised data for instance can be re-identified with the help of data analytics and complementary data sets.
- **Encrypted data processing tools** include homomorphic encryption, as well as trusted execution environments. Encrypted data processing PETs allow data to remain encrypted while in use (in-use encryption) and thus avoiding the need to decrypt the data before processing. For example, encrypted data processing tools were widely deployed in Covid tracing applications. These tools have limitations however. For instance, their computation costs tend to be high although tools are emerging that address this limitation.

[1] OECD (2023), “Emerging privacy-enhancing technologies: Current regulatory and policy approaches”, *OECD Digital Economy Papers* 351, http://dx.doi.org/10.1787/bf121be4-en.

- **Federated and distributed analytics** allows executing analytical tasks upon data that are not visible or accessible to those executing the tasks. In federated learning, for example, a technique gaining increased attention, data are pre-processed at the data source. In this way, only the summary statistics/results are transferred to those executing the tasks. Federated learning models are deployed at scale, for instance, in predictive text applications on mobile operating systems to avoid sending sensitive keystroke data back to the data controller. This also includes multi-party computation (including private set intersection) which allows multiple parties to compute a function over their inputs while keeping those inputs private. Federated and distributed analytics requires reliable connectivity to operate however.
- **Data accountability tools** include accountable systems, threshold secret sharing, and personal data stores. These tools do not primarily aim to protect the confidentiality of personal data at a technical level and are therefore often not considered as PETs in the strict sense. However, these tools seek to enhance privacy and data protection by enabling data subjects' control over their own data, and to set and enforce rules for when data can be accessed. Most tools are in their early stages of development, have narrow sets of use cases and lack stand-alone applications.

Table 1 provides the overview of the four major types of PETs and their opportunities as well as their challenges and limitations.

Table 1. Overview of major types of PETs, their opportunities and challenges

Types of PETs	Key technologies	Current and potential applications*	Challenges and limitations (unless addressed via other PETs)
Data obfuscation tools	Anonymisation / Pseudonymisation	Secure storage	- Ensuring that information does not leak (risk of re-identification) - Amplified bias in particular for synthetic data - Insufficient skills and competences
	Synthetic data	Privacy-preserving machine learning	
	Differential privacy	Expanding research opportunities	
	Zero-knowledge proofs	Verifying information without requiring disclosure (e.g. age verification)	- Applications are still in their early stages
Encrypted data processing tools	Homomorphic encryption	Computing on encrypted data within the same organisation	- Data cleaning challenges - Ensuring that information does not leak - Higher computation costs
	Trusted execution environments	Computing using models that need to remain private	- Higher computation costs - Digital security challenges
Federated and distributed analytics	Distributed analytics and federated learning	Privacy-preserving machine learning	- Reliable connectivity needed - Information on data models need to be made available to data processor
	Multi-party computation (including private set intersection)	Computing on private data that is too sensitive to disclose Contact tracing / discovery	
Data accountability tools	Accountable systems	Setting and enforcing rules regarding when data can be accessed Immutable tracking of data access by data controllers	- Narrow use cases and lack stand-alone applications - Configuration complexity - Privacy and data protection compliance risks where distributed ledger technologies are used - Digital security challenges - Not considered as PETs in the strict sense
	Threshold secret sharing		
	Personal data stores / Personal Information Management Systems	Providing data subjects control over their own data	

Note: (*) Only one application has been included for the sake of readability.
Source: (OECD, 2023[1])

PETs and the OECD Privacy Guidelines Basic Principles

PETs offer functionalities that can assist with the implementation of the basic principles of the OECD Privacy Guidelines presented in Box 1, in particular on the principles of collection limitation, use limitation and security safeguards.

Box 1. OECD Privacy Guidelines Basic Principles

Collection Limitation Principle: There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

Data Quality Principle: Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up to date.

Purpose Specification Principle: The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

Use Limitation Principle: Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [the Purpose Specification Principle] except:

a) with the consent of the data subject; or
b) by the authority of law.

Security Safeguards Principle: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.

Openness Principle: There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

Individual Participation Principle: Individuals should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to them;
b) to have communicated to them, data relating to them i. within a reasonable time; ii. at a charge, if any, that is not excessive; iii. in a reasonable manner; and iv. in a form that is readily intelligible to them;
c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and
d) to challenge data relating to them and, if the challenge is successful to have the data erased, rectified, completed or amended.

Accountability Principle: A data controller should be accountable for complying with measures which give effect to the principles stated above.

Source: OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188.

For instance:

- By adding random noise to datasets or queries on databases that include personal data, differential privacy allows to gain insights from data while making it difficult to identify individual entries. Consequently, the precision of the data collected and shared can be limited to what is necessary for the intended analysis. Differential privacy is frequently employed alongside other PETs such as federated learning and synthetic data as demonstrated in the use cases highlighted below.[2]
- Federated learning, as another example, allows AI models to be trained on devices or in localised environments without the need to centralize personal data. This approach significantly limits the amount of data collected centrally by processing data at its source and only sharing model improvements rather than raw data. For example:

> *Apple has leveraged federated learning to train the voice recognition software used by its AI Assistant, Siri. A local model is trained on an individual's iPhone, and the resulting model weights are periodically communicated back to a central server, which builds a global model by aggregating the weights from the local models. This global model is pushed out to users' iPhones, and the process repeats. Noise is injected during the training of the local model to ensure it is differentially private, so as to mitigate the risk of reidentification. Using this system, Siri can learn to recognise the voice of the iPhone owner, so that it only responds to them without Apple collecting any raw data relating to the users' voice.*

- The generation and use of synthetic data involves creating artificial datasets that mimic the statistical properties of original personal datasets without containing any actual personal data. Using synthetic data for testing and research purposes for instance minimises the need for collecting real personal data by reducing dependency on sensitive information. For example:

> *In Germany, insurance services company Provinzial collaborated with data privacy services firm Statice, using their synthetic data to train machine learning models which optimised their predictive analytics (in particular a "next best offer" recommender engine). A key outcome of this was saving over three months that would have otherwise been spent evaluating data privacy risks, thus addressing both expense of company time and data privacy in the process of optimising their systems.*

- MPC allows multiple parties to compute a function over their inputs while keeping those inputs private. This is particularly useful in contexts where data needs to be combined from multiple sources for analysis without revealing the actual data to other parties. For example:

> *Roseman Labs collaborates with the Dutch National Cyber Security Centre (NCSC) to increase digital resilience against cyber threats like hacking or ransomware. The NCSC collects cybersecurity intelligence from organisations, which report risks such as hacking or ransomware incidents. Organisations are not motivated to publish data on security breaches, which could compromise their reputation. The SMPC system enables the NCSC to anonymously and confidentially collect cybersecurity intelligence from various organizations. This approach allows the NCSC to identify trends without accessing the origin of the data directly, thereby preserving the anonymity and confidentiality of the contributors.*

To some extent, PETs can also support the individual participation and accountability principles. For example, the combined use of threshold secret sharing with personal data stores allows for a heightened degree of personal control over private information. By requiring a consensus among a predefined group to reconstruct the complete data set, threshold secret sharing ensures that no single entity can unilaterally access or alter personal information without permission.

But there are still significant limits in the ability of PETs to help implement the individual participation principle in the context of AI. For example, it is unclear to what extent PETs could help effectively control or even remove information already reflected in an AI model post-training, a technique referred to as

[2] See the repository of use cases of the Responsible Technology Adoption Unit, United Kingdom Department for Science, Innovation and Technology, https://cdeiuk.github.io/pets-adoption-guide/repository/.

"machine unlearning". Furthermore, PETs can also challenge the implementation of certain basic privacy principles. For example, data controllers using encrypted data processing tools may lose the ability to "see" data feeding into their models. This can contradict the need for personal data to be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, to be accurate, complete and kept updated ("data quality principle").

Therefore, and given their respective challenges and limitations highlighted above, PETs should not be regarded as "silver bullet" solutions. They cannot substitute legal frameworks but operate within them, so that their applications will need to be combined with legally binding and enforceable obligations to protect privacy and data protection rights.

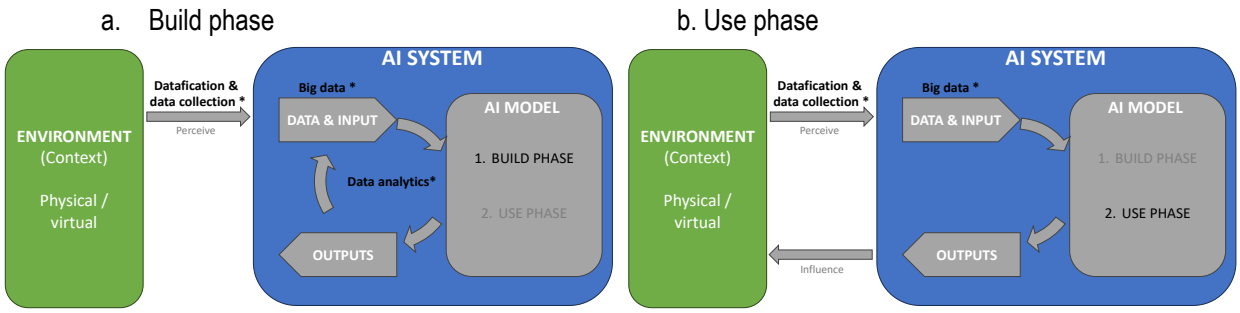
PETs for AI

In November 2023, OECD member countries approved a revised version of the Organisation's definition of an AI system. The updated definition reads as follows:

> *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, 2024[3])*[3]

This updated definition, combined with the data life cycle[4], can help identify the stages within the AI-data lifecycle where PETs can help mitigate data governance and privacy risks, including during its build (pre-deployment) and use (post-deployment) phases as illustrated in Figure 1. The focus here will be on models that are built via machine learning techniques, as other approaches such as symbolic or knowledge-based AI systems may rely less on big data collection and use, and thus on PETs.

Figure 1. Illustrative, simplified overview of an AI system emphasising the AI-data life cycle[2]



Note: This figure presents only one possible relationship between the development and deployment phases. In many cases, the design and training of the system may continue in downstream uses. For example, deployers of AI systems may fine-tune or continuously train models during operation, which can have significant impacts on the performance and behaviour of the system.
Source: Adopted based on (OECD, 2024[3])

[3] OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", in *OECD Artificial Intelligence Papers*, OECD Publishing, Paris, http://dx.doi.org/10.1787/623da898-en.

[4] The *data life cycle* describes a sequence of phases from *datafication and data collection* to a pool of data, sometimes referred to as *big data,* that can be processed via *data analytics* to generate insights or AI models. The use of the later creates value and help generate additional data. It was introduced as "data value cycle" in OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229358-en.

Documented use cases indicate that most PETs are used predominantly during the **build phase of AI development**, playing a critical role in the **'datafication and data collection' stage**[5]. In this context, PETs can help protect confidentiality and privacy before the data are used as input in AI systems.

This is for example the case for *synthetic data* (see the example of insurance services company Provinzial, p. 4), *homomorphic encryption* and *differential privacy*. The latter, differential privacy, is often used in combination with other PETs to significantly reduce the risk of re-identification of individual data points. It involves adding controlled noise to data before they are used as input for machine learning, federated learning (see the example of Apple's Siri, p. 4), or the generation of synthetic data. Differential privacy has also been successfully used to ensure the confidentiality and privacy of output results.

Other PETs are used to protect the actual data processing and calculation activities when building and using (local) AI models during the **data analytics stage**, often also in combination with the PETs highlighted above. *Trusted execution environments (TEEs*), for instance, ensure that data and AI models are processed in a secure, tamper-resistant environment, protecting it from unauthorized access, an issue that is particularly relevant for protecting intellectual property and proprietary AI models.

Federated learning is used predominantly during the build phase of AI systems, specifically in the data analytics stage for constructing local AI models. But it also contributes to enhancing confidentiality and privacy, where local models' parameters serve as inputs for a global AI model (without exchanging the underlying data). Differential privacy (see the example of Apple's Siri, p. 4) but also MPC (or similar protocols) are typically used to further enhance confidentiality and privacy in this context.

MPC can enhance confidentiality and privacy by enabling different parties to jointly compute functions over their inputs while keeping those inputs private from each other. Random number data can be used in addition to further mask input data in the MPC protocol. The models are processed locally and partial results which contain no information individually can be combined to reveal the final result thus enhancing output privacy (see example of Roseman Labs, p. 4).

Table 2 offers a tentative overview of the stages within the AI-data lifecycle where PETs can help address challenges related to data governance and privacy. The final version of the table may also help highlight which PETs are commonly combined and which deserve further attention by policy makers and regulators.

Table 2. Tentative overview of the AI-data lifecycle stages where the use of PETs can be beneficial

		Synthetic data	Homomorphic & other encryption	Differential privacy	TEEs	Federated learning	MPC
Build phase*	Datafication & data collection			x			
	Big data	x	x				
	Input	x	x	x		x	x
	Data analytics				x	x	x
	AI model		x				
	Output			x			x
Use phase	Input		x	x			
	AI model		x	x	x		
	Output		x	x			x

Note: This table provides a preliminary overview and will be further refined based on discussions and feedback received during the upcoming United Kingdom - Estonia - OECD Expert Workshop on Privacy Enhancing Technologies and Artificial Intelligence.
* On the build phase: in many cases, the design and training of the system may continue in downstream uses.

[5] Datafication and data collection "refer to the activity of data generation through the digitisation of content, and monitoring of activities, including real-world (offline) activities and phenomena, through sensors" (OECD, 2015[4]). This step also encompasses data preprocessing activities such as data cleaning and linkage.

Regulatory and policy approaches to foster the adoption of PETs

PETs are often addressed explicitly and/or implicitly in countries' privacy and data protection laws and regulations through: legal requirements for privacy and data protection *by design and by default*; requirements for *de-identification*, *digital security* and *accountability*; and/or regulatory mandates to privacy enforcement authorities (PEAs) to further promote adoption of PETs. (OECD, 2023[1]) To complement the above measures, governments or PEAs have issued guidance, which functions as supplemental material to help clarify rules. In June 2023, for instance, the United Kingdom Information Commissioner's Office issued guidance on how PETs can help organisations achieve compliance with UK data protection law.

In addition, countries have adopted a wide variety of complementary policy initiatives to promote innovation in and with PETs. Those that are relevant to AI, include: research and development (R&D) support, adoption of secure data processing platforms, certification of trusted PETs, innovation contests, and regulatory and other sandboxes. For example:

- **R&D support**: The 2023 United States' National Strategy for Privacy Preserving Data Sharing and Analytics aims to foster R&D to enable researchers, physicians, and others to gain better insights from sensitive data without the need for data access. This initiative will be complemented by the Privacy Enhancing Technology Research Act, a bill that was passed by the United States House of Representatives in 29 April 2024, and requires certain United States federal agencies, including in particular the National Science Foundation, to support R&D of PETs.
- **Innovation contest**: In July 2022, the United Kingdom and the United States governments launched a set of prize challenges to unleash the potential of PETs to combat global societal challenges. These challenges provided the opportunity to innovators from academia, industry, and the broader public to demonstrate the ability of PETs to address real-world financial crime and public health challenges via federated learning approaches that enabled multiple organisations to collaborate on data that they could not share between them. The challenges concluded in March 2023 with winners announced at the Summit for Democracy 2023.
- **Government procurement of PETs:** In 2022, the Estonia Ministry of Economic Affairs and Communications launched Bürokratt as "the world's first public service AI-based virtual assistant". For its ongoing development, its procurement process has focused on PETs such as federated learning, and synthetic data for publishing public sector data.
- **Regulatory sandboxes**: In Singapore, since its launch in July 2022, the IMDA and PDPC operated a regulatory sandbox on PETs that aims to provide a safe environment and testing ground to pilot PET projects.
- **Certification of trusted PETs**: In Japan, MIC and METI have formulated guidelines for the certification of personal data trust banks, which are used by private organisations such as the Information Technology Federation of Japan.

These complementary measures underline that governments are increasingly recognising PETs as a strategic priority within their national policies. Notably, some countries like the United States have acknowledged their importance at the highest political levels, as evidenced by the White House Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, which calls for "Strengthen[ing] privacy-preserving research and technologies". This strategic recognition underscores the need to further integrate PETs coherently across various policy and regulatory frameworks to ensure their effective adoption across different policy domains and sectors including within the context of AI.