Layered Privacy

Privacy-Enhancing Technologies (PETs) and Their Role in AI

reglab
center for strategy & regulation

About reglab

We are a private research center specializing in the media and technology sector, supporting companies, associations, and policymakers in making strategic decisions based on data and evidence.

Learn more at www.reglab.com.br

About the policy briefs series

The **Policy Briefs Series** brings together studies that evaluate trends, existing public policies, or legislative proposals, using both qualitative and quantitative data to inform and guide decision-making. The goal is to present complex topics in an accessible way, highlighting key points of analysis, impacts, and possible recommendations.

Acknowledgements

Executive Director: Pedro Henrique Ramos

Research Coordinator: Marina Garrote

Authors: Pedro Henrique Ramos and Daniela Naomi Shimabukuro Nomura

Researchers: Marina Garrote, Stephanie Souza, Giulia Brombine and Vinícius Pimenta

Final layout: Eliza Natsuko Shiroma, Camyla Romão

Suggested citation: RAMOS, P.H.; NOMURA, D. N. S.Layered Privacy: Privacy-Enhancing Technologies (PETs) and Their Role in AI. Policy Briefs Reglab n.3. São Paulo: Reglab, 2025.

Executive Summary

The adoption of Generative Artificial Intelligence and systems based on personal data has added new layers of complexity to the privacy debate. **In this first-of-its-**kind study in Brazil, Reglab investigated how Privacy-Enhancing Technologies **(PETs)** can be applied to mitigate risks in the training and operation of these systems, drawing on the perspectives of experts in AI, data protection, and cybersecurity.

The qualitative interviews revealed **consensus on the central role of PETs** in the reduction of risks of re-identification, misuse, or unauthorized exposure. While they do not completely eliminate vulnerabilities, these technologies provide additional layers of protection that make data processing more reliable and capable of giving organizations that operate solutions with AI greater confidence.

Among the key findings are:

- **Risk mitigation:** PETs enhance the protection of personal data, but currently they cannot fully eliminate existing privacy risks;
- **Project enablement:** technologies such as federated learning and trusted execution environments make it possible to pursue initiatives with high social values, such as health research, that would otherwise be unfeasible due to privacy risks;
- **Persistent limitations:** experts emphasize that there is no “zero risk” in cybersecurity and that many solutions still demand large-scale testing, which limits market confidence and delays investments;
- **Fragmentation in Brazil:** the practical use of PETs in the country remains sporadic and unsystematized, hindering scalability and integration with broader organizational strategies.

Building on these elements, the study proposes the **Layered Privacy Model**, a pioneering framework that organizes data protection into five interconnected levels:

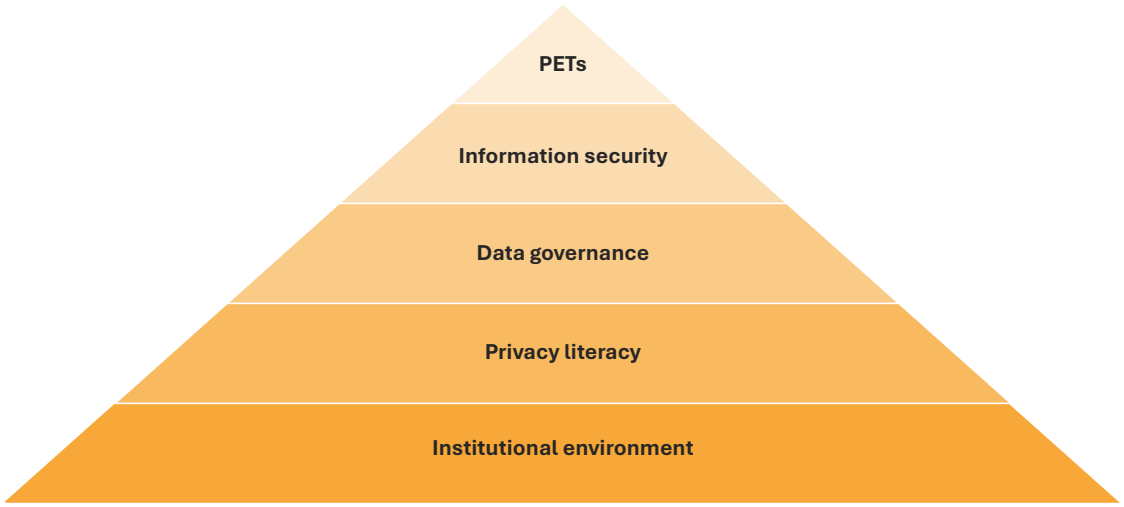
i. **Institutional environment** (regulation and external pressures);

ii. **Literacy** (education and training in privacy);

iii. **Governance** (internal processes for strategic alignment);

iv. **Information security** (robust cybersecurity infrastructure);

v. **PETs** (technical tools for risk mitigation).



Source: author's own elaboration

This model enables organizations to visualize different degrees of maturity and helps guide both companies in self-assessment and investment prioritization, and policymakers in designing regulatory environments that encourage the responsible adoption of privacy-preserving technologies.

The main contribution of this study is to demonstrate that PETs only realize their potential when integrated into a broader strategy of governance, literacy, and security. On their own, these technologies are not sufficient to guarantee full protection, but together they can help consolidate a stronger regulatory and business ecosystem, aligned with the responsible use of data.

TABLE OF CONTENTS

1. Introduction

The importance of Artificial Intelligence (AI) in society is no longer up for debate; it has become an established fact. It is a transformative element across multiple dimensions: in the economy, with the potential to boost global GDP by up to 7 trillion dollars and increase productivity growth by 1.5% over ten years (Goldman Sachs, 2023); and in organizational management, where it can reduce the time dedicated to people management by 10% (Edin et al., 2025), to name just a few examples.

The current challenge seems not only to lie in recognizing the benefits of AI, but in **developing governance models capable of keeping pace with its rapid spread and the disruptive processes it triggers.** One of the most complex issues is the relationship between this technology and privacy and personal data protection.

As different applications become integrated into everyday life and critical societal decisions, **questions about how to safeguard personal data and comply with data protection laws** are increasingly present in headlines, court rulings, and legal forums. On the other hand, there seems to be equally significant technological progress in the field of information security solutions. Between these domains, however, it is not uncommon to perceive a gap in knowledge.

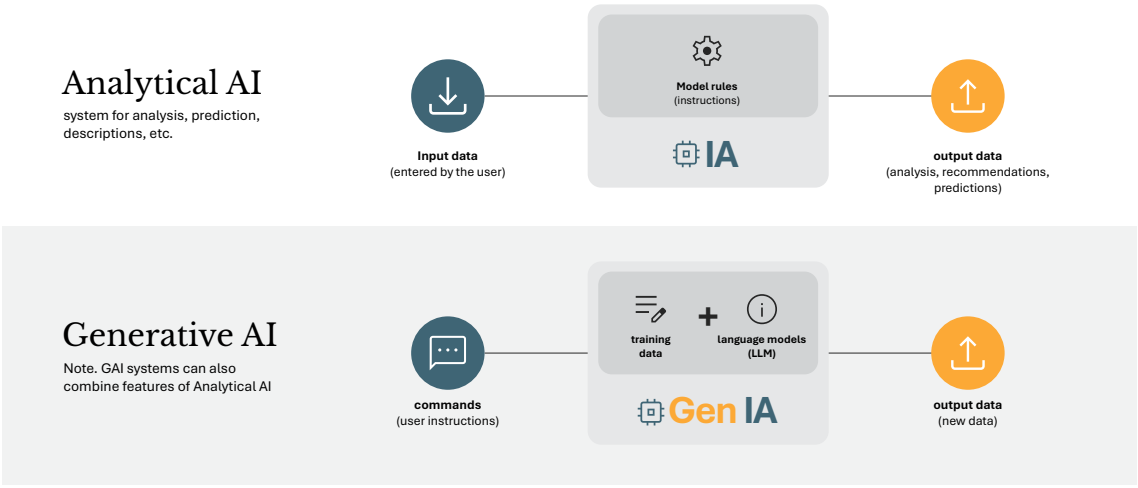
This study seeks to serve as **a bridge between the field of data protection and advances in Privacy-Enhancing Technologies (PETs).** Our aim is to explore how these technologies can redefine the contours of data governance in AI systems, especially in a context where regulatory frameworks strive to keep pace with the speed of technological development.

1.1. What is AI and why does it matter?

As with any research work, it is essential to clearly define our object of analysis. In this study, we adopt the following definitions:

- **AI systems:** software systems that process and analyze data through algorithms and mathematical models, using statistical and computational techniques to identify patterns, make predictions, and generate specific outputs. For simplified explanation, they can be classified into two main types:
  - Analytical AI systems: systems designed to solve specific problems using structured datasets, operating within predefined parameters. They function as “sophisticated calculators” that perform deterministic tasks such as classification, prediction, or recognition of existing patterns (Barr, 2023). Examples include banking fraud detection systems and personalized recommendations by virtual assistants.
  - Generative AI systems: systems that employ statistical and machine learning techniques to generate new content such as text, images, and videos (Barr, 2023).

They are based on large language models (LLMs), which transform training data into mathematical representations (vectors) that capture patterns and statistical correlations. Currently, the most popular applications are chatbot models (Stryker & Scapicchio, 2024).



Source: Reglab's elaboration, based on Ramos (2023)

Personal data refers to information related to an identified or identifiable natural person, which can be processed in various ways within AI systems—whether as input data directly provided by a user or as part of the datasets analyzed during training. We will explore this relationship further in section 2.1 below.

1.2. The methodological approach of this research

Between 2016 and 2024, **several countries developed or updated their personal data protection laws.** Most of these regulations, such as Brazil's General Data Protection Law ("Lei Geral de Proteção de Dados" - LGPD), enacted in 2018, were established before the widespread adoption of generative AI applications and, consequently, do not always provide clear guidance for emerging challenges.

In this context, **it is important to investigate the practical knowledge of how the market itself,** operating with a dynamism and pace different from that of legislators—has developed technological solutions to protect rights in the field of AI, adding layers of security to reduce risks and fostering trust in its adoption.

> **This research examines the relationship between data protection and PETs in AI systems in Brazil. Our objective is to identify pathways that support the development of best practices and privacy-by-design guidelines, facilitating the incorporation of privacy from the earliest stages of development and throughout the use of AI systems by Brazilian companies.**

The study is based on one of Reglab's main methodological premises: the **policy translation approach,** still little explored in the field of digital governance in Brazil. This methodology emphasizes the **active process of interpreting and adapting complex research findings into formats that are understandable, relevant, and applicable for public policymakers** (Ingold & Monaghan, 2016).

When we use colored boxes, graphs, examples, and anecdotes, we do so intentionally. We recognize the risk of possible technical imprecisions, but we understand that translating complex evidence into applied knowledge requires making the content clearer and more accessible. This is a necessary methodological choice—and a position we adopt with full transparency.

For our data collection methodology, we opted for a different approach from conventional literature reviews and documentary research: **in-depth qualitative interviews.** Inspired by reception study methods, we sought to understand how **professionals who face technical cybersecurity challenges on a daily basis** perceive the processing of personal data in AI systems, which PETs are available, how they are effectively implemented in practice, and in what ways they contribute to better protection of individual privacy.

Over the course of one month, we conducted **11 interviews with experts**, focusing on senior-level professionals with practical experience in cybersecurity and data protection compliance. The interviews followed pre-defined scripts and confidentiality protocols, and their transcripts and records were analyzed using Atlas.ti software through thematic analysis techniques.

Interviewee	Description
A	Man, business sector, machine learning engineer
B	Man, academic sector, cybersecurity professor
C	Man, academic sector, cybersecurity professor
D	Man, business sector, systems engineering consultant
E	Man, business sector, cybersecurity consultant
F	Man, business sector, cybersecurity researcher
G	Woman, business sector, privacy and security management consultant
H	Woman, business sector, cybersecurity consultant
I	Woman, business sector, information security and privacy consultant
J	Woman, business sector, privacy consultant
K	Woman, academic sector, data scientist

The complete methodology, with details about the adopted procedures, is available in the end of the study.

2. Results

2.1. How is personal data used in AI models?

FUNDAMENTAL CONCEPTS

In Analytical AI, personal data is generally processed in a structured manner, within defined database maintained by organizations. This format increases protection obligations but facilitates the application of principles such as minimization, purpose limitation, and legal basis, since the contexts of use are predefined.

In Generative AI, Large Language Models (LLMs) are trained on massive volumes of data collected from the internet. These data are broken down into units called tokens and converted into mathematical representations (vectors), which capture statistical relationships between words and phrases.

As a rule, models do not store personal data directly. However, since they operate based on statistical patterns, information that is repeated frequently during training can be "remembered" and reproduced in outputs, since vectors function as representations of knowledge.

Models can also generate personal data that does not exist or was not present in the training set. In these cases, it involves statistical inference rather than memory. For example, this occurs when a model generates combinations of numbers with the format of a valid CPF (Brazilian taxpayer ID) without having stored that specific data.

The interviews with experts helped us understand how personal data is used and how it is anonymized during processing. However, while there appears to be a uniform understanding when interviewees discuss **Analytical AI systems**, we observed divergences and knowledge gaps when delving into the workings of **Generative AI systems** and, in particular, **LLMs**, which received greater emphasis in the interviews.

Use of data – Analytical AI

In so-called **Analytical AI**, interviewees emphasized that personal data is processed in a more **structured and delimited way, and directly linked to the purposes of the project.**

Strucured Data: at some stage of processing, the information is organized into standardized formats that allow for classification or prediction, such as tables in databases or sets of numerical variables.

Controlled Database: when personal data is used, these records are associated with individuals either explicitly (e.g., taxpayer ID, medical record, account number) or through techniques such as pseudonymization, which replaces direct identifiers with codes.

Defined Purpose: the data is processed to solve specific problems, such as credit risk analysis, fraud detection in financial transactions, or supporting assisted medical diagnoses.

> *"In practice, these models are mathematical functions that are trying to find patterns in the data. A classic example is having structured data in table format, which you use as input, and then you apply a specific function to interpret that data." [Interviewee K]*[1]

In general, interviewees highlighted that Analytical AI offers smaller scale and greater precision: the volume of data may be lower, but the quality and relevance of individual information tend to be more decisive for model performance.

However, some interviewees emphasized that these models carry greater re-identification risks, and, compared to LLMs, *may even be more dangerous to privacy* ***[Interviewee B].***

This is because data remain in structured databases and, even when pseudonymized, can be easily reconnected to individuals when combined with other sources. This operation differs from LLMs, which tend to dilute patterns into large-scale statistical representations, as we will see later.

Data Use - Generative AI and the role of LLMs

Whether explicitly or implicitly, different interviewees explained the use of personal data in generative AI models through two phases:

- **Training:** the stage in which large volumes of data — potentially containing personal information — are collected and transformed into **tokens** to generate **statistical representations.** The risk lies in the potential **memorization of recurring excerpts,** which may resurface in the outputs.

- **Inference:** the stage in which the model generates outputs from new user prompts. At this point, there is **no direct access to the original dataset,** but rather to **statistical recombinations.** Nevertheless, personal information may still emerge — either through **memorization** or through **plausible inferences** produced by the model's patterns.

DIDACTIC EXPLANATION

Training: it's like when someone studies for a test by reading dozens of books and "learns" patterns from the acquired knowledge. If a piece of information appears repeatedly in this material (e.g., the periodic table), the model will end up memorizing it and reproducing that information later.

Inference: it's like using that knowledge in a closed-book exam: the model does not look back at the original books, but rather combines statistics to create the answer. In doing so, it may either repeat something memorized or invent something new that seems true.

1 In order to preserve the anonymity and confidentiality of the research participants, specific modifications were made to the quotes presented in this study. In certain circumstances, linguistic adaptations were made to ensure the interviewees' original intent in the transcripts. The original discursive register was preserved whenever possible, in accordance with the established methodological principles.

During training, interviewees explained that LLMs operate through massive data analysis, converting texts into numerical representations. These systems do not function as repositories of complete sentences nor as databases of raw personal data: what they internalize are **statistical language patterns,** that is, the frequencies and relationships between words and expressions (Stryker, 2025).

In practice, this means that **the model prioritizes statistical correlations rather than individual records.** Just like someone who reads a thousand résumés does not memorize all the names but instead notices career patterns (for example, that people with a degree in Business Administration often work in private companies), an LLM absorbs tendencies in word usage. However, there are situations in which specific information may be reproduced, especially when it is very frequent in the training material (Kandpal; Wallace; Raffel, 2022):

> *"The model only knows who Harry Potter is because there are tens of thousands of pages on the internet mentioning the character. This is related to what we call* **statistical relevance.** *Whether or not the model was trained on the original Harry Potter books is practically irrelevant, since those passages have already been replicated across thousands of other websites used during training."* **[Interviewee D]**

There are also points of concern during the inference stage. During inference, personal data may be processed transiently, especially when users input personal information into prompts or share documents. This information is processed temporarily by the model to generate responses, but **it does not become part of the permanent base model,** although it may remain in short-term memory or be used for personalization of the user's own profile.

> *"People often feel like AI is learning all the time, right? And that's not true. In practice, it's not learning anything, especially in more controlled environments, where practically every interaction is disposable from the perspective of anything beyond the user — it remains confined within that user's space."* **[Interviewee H]**

Another recurring point in the interviews was the ability of Generative AI to produce personal data in its outputs, e**ven when such records did not appear literally in training.** Interviewees described this process as an "invention" by the model, the result of its statistical capacity to combine patterns in a plausible way. Examples include generating valid CPF sequences (following the 11-digit rule with a verifier) or common name combinations such as "Ana Maria."

This point is relevant because it shows that, even without functioning as personal data repositories, generative models can produce identifiable information, which raises practical challenges for the interpretation and application of data protection laws.

Data Anonymization in Generative AI

The most debated issue among the interviewees was **the anonymization of data in Generative AI models.** The main emerging point is the difficulty of ensuring the complete loss of identifiability of personal data of personal data, especially in the most advanced models.

There is a convergence among specialists: **data is not stored as raw records but transformed into statistical vectors.** In this process, names, numbers, and phrases are converted into tokens and mathematical weights that come to represent language patterns. This feature significantly changes the debate on data protection, shifting the focus from literal collection and storage to the **risks of reidentification and the statistical use of information.**

At this point, many of the specialists, senior professionals and with experience in their field, **were cautious in their responses and, in some cases, even admitted not knowing exactly how the process works.** Their statements and explanations did not demonstrate the same level of confidence as when explaining other aspects of the technology's functioning.

> *"It's not that I turn it into numbers, it's not a 'pseudonymization.' It simply dissolves into tokens [...] it doesn't have a database; what it has are pieces of words."* **[Interviewee D]**
>
> *"I read an article, I was reading an article about this, that when I extract data from one AI model to another, I can reidentify people. And then someone takes that and puts it into another model, and another model, and then I can get to people. [...] Yes, there is a risk of reidentification, even if I apply encryption, anonymization, pseudonymization, there is. If I start cross-referencing there, I can do it."* **[Interviewee I]**
>
> *"I don't believe it loses the identifiable nature. Eventually it will identify you. It's very strange because I can't explain. It says it has no memory, but it does. I can't explain."* **[Interviewee G]**

In any case, the debate revealed a common conclusion: **there is a need to increase public literacy and knowledge about anonymization in Generative AI.** The divergences among specialists indicate that this is not merely a technical issue, but also one of interpretation and risk management.

2.2. What are PETs

FUNDAMENTAL CONCEPTS

- **Definition:** PETs is the acronym for Privacy Enhancing Technologies, which translates as Tecnologias de Aprimoramento da Privacidade. These are software solutions developed to reduce privacy risks and improve cybersecurity.
- **Little-Known and Underused Term:** Although it is popular within the data protection legal community, most interviewees agree that the term "PETs" is not well known or widely used in Brazil. However, standardization would help consolidate the field and create clear references regarding which mechanisms are effective and in which scenarios.
- **Importance of Raising Awareness:** Interviewees consistently pointed to the lack of knowledge and low technical literacy about PETs as one of the main challenges to their adoption in Brazil. Without awareness, companies are unlikely to invest in PETs, especially given the high financial and computational costs of these technologies.

PETs function as an umbrella term encompassing various technologies and techniques aimed at reducing privacy and cybersecurity risks. Their objective is to enable the responsible use of data, making projects viable that, without these tools, could be considered unfeasible or excessively risky from a regulatory and personal data protection standpoint.

Although the term is internationally recognized and known among legal privacy specialists, interviewees noted that it still has limited penetration in Brazil. Many argued that the term should receive greater visibility in the national debate, aligning with the importance it already holds in international forums — highlighting a disparity between the global discussion and the application of the concept in the Brazilian context.

WHY DO PETS APPEAR MORE IN THE LEGAL FIELD THAN AMONG CYBERSECURITY SPECIALISTS?

One possible explanation is that the term Privacy Enhancing Technologies initially emerged from reports by data protection authorities in Canada and the Netherlands in the 1990s, followed by a well-known European Commission report in 2007 (NicFab, 2023). From this, the acronym PETs became consolidated in guidelines and standards as an umbrella category to describe practices such as anonymization, pseudonymization, and privacy by design, serving the function of translating different techniques into a single normative language.

In the technical field, however, it seems that professionals prefer to name specific methods, often considering umbrella labels too vague, as they do not clarify which measures were applied, under what conditions, or with what guarantees.

Despite the low recognition of the term PETs, interviewees demonstrated familiarity with specific technologies such as differential privacy, federated learning, and homomorphic encryption when mentioned concretely. Several accounts indicated that professionals tend to refer more to the manufacturer's name or the commercial brand of privacy solutions than to the technical term for the technology itself.

> ***This appears to show that the debate about PETs* in Brazil is more market- and product-oriented, *rather than based on a standardized or conceptual language*.**

Interviewees repeatedly pointed out that **knowledge, consistency, and standardization around the term PETs are crucial for its adoption in Brazil.** They emphasized that low technical literacy and lack of awareness about the topic compromise both investment decisions and the consolidation of the field, and that the absence of standardization also prevents the creation of clear references regarding the effectiveness of each mechanism in different scenarios.

The high financial and computational costs of these technologies, combined with a lack of familiarity, make it difficult to justify investments. This gap affects both public and private funding: as several interviewees highlighted, Brazil lacks specific research support lines for PETs, which compromises the development and feasibility of these solutions.

> ***"To be honest: I know the term, but I've never seen it being used, either in academia or in industry (...)*** *I think the term PETs might be appropriate because we need to start building this knowledge; there is still a lot of work to be done in the field, and the formalization of a terminology would be an interesting step."* ***[Interviewee B]***
>
> ***"People will talk more about a tool, a solution, or a technique,*** *but they say that for me, there is still a lack of literacy in this regard."* ***[Interviewee I]***

In other words, to evaluate risks and potential more concretely, it is necessary to go beyond the generic label and examine the specific technologies that make up this universe. It is these tools, with their limits and practical applications, that allow the actual utility of PETs in AI projects to be assessed.

The following technologies are organized according to the frequency of mention in the interviews and the level of knowledge demonstrated by the interviewees regarding each of them.

Differential Privacy

- **Definition:** A technique that introduces statistical noise into data or results, making the presence or absence of an individual practically indistinguishable, while preserving aggregated utility.
- **Didactic Example:** It's like adding a little noise to the voices in a crowd: you can understand what the group is saying overall, but you cannot identify any single person.
- **Practical Example:** Smartphone manufacturers apply differential privacy in operating systems to collect usage statistics without exposing individual users.
- **Relevance:** Provides formal mathematical privacy guarantees, allowing data to be analyzed or used for training without revealing raw records.
- **Challenge:** Calibrating the noise can reduce model accuracy and make training more costly in terms of time and resources.

Differential privacy is a technique that protects against re-identification risks in large datasets (RTT, 2024). It works by adding controlled random noise to the information, significantly reducing the possibility of associating data with specific individuals (CIPL, 2025). This way, it is still possible to extract patterns and make useful inferences from the data while maintaining the privacy of the people involved.

In the interviews, differential privacy was characterized as a complex yet well-established anonymization method capable of strengthening protection in contexts of intensive data use. As Interviewee H summarized:

> *"With differential privacy, we don't lose the AI's ability to learn from the data, but we make it harder to re-identify that individual, because I am manipulating the original data without it losing its essence."* ***[Interviewee H]***

In this sense, **differential privacy** has gained prominence in sectors that heavily rely on data, such as healthcare (Feretzakis et al., 2024). In a concrete project shared by Interviewee I, the introduction of controlled noise into sensitive datasets, such as medical records, allowed researchers and developers to train AI models capable of identifying relevant clinical patterns without exposing individual patient information.

Despite its high anonymization potential, interviewees also warned about the trade-off between privacy protection and utility: the technique can reduce AI model accuracy and decrease the analytical value of data, a concern confirmed by recent research (CIPL, 2025).

Experts indicated that differential privacy **works best in Analytical AI with tabular data** (age, medical records), where controlled noise reduces re-identification risks without significantly compromising data utility. **In Generative AI, adding noise can be interpreted as a statistical signal**, causing the model to learn this "artificial pattern" instead of the original data, reducing training effectiveness and compromising practical utility.

Trusted Execution Environment (TEE)

- **Definition:** Creates an isolated environment (enclave) within the hardware, where data is processed securely, protected from unauthorized access, including by the infrastructure provider itself.
- **Didactic Example:** It's like a "digital safe" where data can be used for computations, but inside it, no one can peek at what happens.
- **Practical Example:** Cloud services that allow processing health data in TEEs, so that even the engineers of the cloud company cannot access the raw information.
- **Relevance:** Differentiates itself from techniques such as anonymization because it does not alter the data, but ensures that they are only processed in a controlled and highly secure environment. It is essential for sectors such as healthcare or finance.
- **Challenge:** Most TEE solutions are provided by large foreign technology companies, raising concerns about data sovereignty and technological dependence.

Trusted Execution Environments (TEEs) are isolated areas within a computing system that allows data to be processed with a high degree of security (CIPL, 2025). Unlike other PETs that operate directly on the data through techniques such as anonymization or encryption, TEEs protect the environment in which the data is processed. As Interviewee B explained: *"Other PETs work at the data layer [...]. The secure execution environment does not operate on the data itself, but on the environment in which it will be processed." [Interviewee B]*

This architecture is especially useful in applications involving sensitive data. By processing sensitive information within a TEE, it ensures that data remains protected even in contexts where the rest of the system may be vulnerable.

Despite its advantages, interviewees also expressed concerns regarding data sovereignty. As most TEEs available today are owned by foreign companies, they highlighted the importance of developing national data centers capable of providing trusted execution environments without allowing the infrastructure provider access to the processed information. Interviewee D explained that even if a cloud instance is located in Brazil, if the infrastructure belongs to a U.S. company, the Cloud Act may permit remote access to Brazilian users' data by other jurisdictions (Teofilo, Rocillo, 2018).

Synthetic Data

- **Definition:** Artificially generated data designed to simulate real information, used to expand samples, balance datasets, and reduce the use of raw personal data in training.
- **Didactic Example:** The system "invents" fictional customer records or creates fake photos based on real information, allowing the computer to "learn" without accessing real people's information.
- **Practical Example:** Creating artificial faces to train a facial recognition system without using real citizens' photos.
- **Relevance:** Expands training datasets, increases model robustness, and allows the replacement of personal information columns with synthetic versions while preserving the data structure.
- **Challenge:** May reinforce stereotypes or generate false information, producing less reliable models when used exclusively for training.

Synthetic data are artificially generated information created by algorithms that replicate the statistical properties of real datasets without copying actual records (Microsoft, 2025). This technique allows AI models to be trained and tested without exposing personal data, replacing it with fictional versions, either partially or entirely. In addition to reducing re-identification risks and mitigating the impact of leaks, synthetic data are especially valuable in sectors with strict regulatory compliance and data scarcity, such as healthcare (IBM, 2023).

In the interviews, Interviewee A highlighted the relevance of this technology for training image models, whether by creating variations of existing photos (data augmentation) or 3D models (digital twins) for simulations, increasing system robustness. On the other hand, Interviewee I warned about the drop in accuracy when models are trained exclusively with synthetic data:

> *"It's no use just generating synthetic images with AI, because it has already been proven that a model trained with synthetic data generated by other models suffers a massive drop in accuracy. It's our diversity, for example, that provides richness to a facial detection system that uses computer vision."* – ***[Interviewee I]***

Despite its potential, it is important to recognize that the generation of synthetic data generally depends on personal datasets to train the algorithms that produce it. As noted by Interviewee D, the direct use of real data, when safeguarded by robust mechanisms of governance, anonymization, and access control, can provide greater reliability and reduce the risk of distortions or "hallucinations" in the models. This perspective is consistent with recent academic analyses (Giuffrè; Shung, 2023).

Therefore, synthetic data should be regarded as complementary tools, not as complete substitutes in protection and innovation strategies.

Federated Learning

Definition: A technique where data remains on users' devices (phones, laptops), and only mathematical representations are sent to a central server to update the model.

Didactic Example: Imagine a team where each athlete trains in their own gym and only sends their training results. The coach uses these results to improve the team's strategy without ever seeing the complete training sessions.

Practical Example: A hospital can train a model across multiple medical centers without receiving the patient records from each. Each unit trains locally and shares only the statistical results.

Relevance: Reduces the need to centralize large volumes of personal data, increasing privacy and security.

Challenge: The technique is still limited outside sectors that heavily rely on data (such as healthcare and technology) and require high computational power on each end device.

Federated learning allows machine learning models to be trained collaboratively while keeping data at its original source. Instead of transferring data from peripheral devices to a central server, each participant uses their local information to train the model (Caballar; Stryker, 2025). After each training cycle, only the updated parameters—not the raw data—are sent to a central server, which consolidates them to improve the global model (Caballar; Stryker, 2025).

This technology enables collaboration between organizations that cannot share sensitive data due to privacy and security concerns. Interviewee D mentioned, as an example, a fraud detection project in the healthcare sector. In this case, companies needed to train a joint model but refused to share their databases and did not trust an intermediary to guarantee privacy. The solution was to process data locally at each institution and, in the end, combine only the training parameters. This arrangement overcame trust barriers and demonstrated how federated learning can create joint solutions even in contexts of high sensitivity and low willingness to share information.

However, due to the need for frequent communication between devices and the central server to update model parameters, interviewees highlighted that federated learning requires high processing power and connectivity (CIPL, 2025). For this reason, as confirmed by Interviewees A and B, its practical application is more viable for companies whose business is data-centric, such as technology firms, and that already have the infrastructure to maintain this continuous flow of information.

Homomorphic Encryption

Definition: A technique that allows computations to be performed directly on encrypted data without revealing the original content.

Didactic Example: Imagine performing calculations with numbers inside locked safes: you never see the numbers, but you can add and multiply them without opening the safe.

Practical Example: A bank could analyze customer balances to predict credit risk without ever accessing the actual account values.

Relevance: Considered one of the strongest protection solutions, as it keeps data encrypted throughout the entire processing workflow.

Challenge: Still under scientific development. Supported operations are basic and insufficient for training advanced models or performing complex calculations. Additionally, the computational cost remains extremely high, making large-scale use impractical.

Homomorphic encryption is considered one of the most active and challenging areas in information security research. Its theoretical proposal dates to the 1970s, and the first proof of concept was proposed in 2009 (Gentry, 2009). Since then, scientific literature has progressed toward more efficient schemes, but the consensus is that the technology remains experimental, with practical applications limited to low-complexity scenarios or controlled prototypes.

Typically, data is encrypted during storage or transmission. However, for operations such as updates, searches, analyses, or computations, it is usually necessary to decrypt the data first, exposing it to potential unauthorized access. **Homomorphic encryption offers a more secure solution: it allows computational operations to be performed directly on encrypted data, without the need to reveal it during processing** (CIPL, 2025).

In other words, this technique enables systems to perform calculations or analyses on protected data while maintaining confidentiality at all stages. For this reason, homomorphic encryption has been highlighted as a promising tool for applications in sectors that handle sensitive information, such as healthcare, finance, and electoral processes (Ruiz, 2022).

Inference on already-trained models emerges as the most viable use to date, allowing organizations to perform queries or predictions on encrypted data without risk of exposure and while maintaining result accuracy. Nevertheless, in the interviews, specialists emphasized that homomorphic encryption has **not yet reached full technical feasibility** and requires further research to become applicable at scale. The main limitation is the high computational cost required for its operation. Interviewee B noted that *"with homomorphic encryption in particular, the cost can reach a thousand percent higher."*

2.3. PETs and Privacy

There was consensus among the interviewees that PETs play a central role in mitigating risks related to the use of personal data. Specialists highlighted that these technologies significantly reduce the chances of re-identification, misuse, or improper exposure. It was emphasized that PETs provide additional layers of security, making data processing more reliable and instilling greater confidence in organizations using AI solutions.

By indicating that these technologies substantially reduce risks associated with data use, specialists also suggested that their adoption can make data processing more proportional and balanced. **In AI projects, this means bringing practice closer to the ideal of collecting only what is necessary and applying safeguards that demonstrate a commitment to responsible use of information.**

> *"[PETs] are very, very important because they mitigate risks more and provide greater security for companies that handle AI solutions with personal data, especially since the company controlling the data must comply with these legal requirements."* ***[Interviewee J]***

Furthermore, PETs enable projects that would otherwise be unfeasible due to high privacy risks. Examples include **using federated learning** to train AI models with health data from multiple institutions and secure processing of personal information in TEEs, allowing collaboration without improperly exposing data.

It is important to note that specialists recognized PETs do not entirely eliminate risks: *"there is never 100% risk elimination; this is a fundamental principle of cybersecurity and information security,"* stated Interviewee B. Nevertheless, the perception is that **the potential of these technologies remains underutilized:** although they offer significant gains in data protection and security, many solutions still lack extensive and consistent testing, which reduces market confidence and hinders new investments.

The **Layered Privacy Model** results from qualitative empirical research and thematic analysis of expert interviews and organizes a progressive data protection framework in a replicable way. It does not rely on abstract normative assumptions but emerges from the systematization of reported practices, structuring layers that integrate the institutional environment, literacy, governance, information security, and PETs.

3. Analysis and Comments

The research revealed a consistent set of practices, perceptions, and tensions among the interviewed specialists. Although their perspectives vary on specific aspects, they converge on one point: **PETs play a relevant role in risk mitigation, but their practical use in Brazil remains fragmented and poorly systematized.**

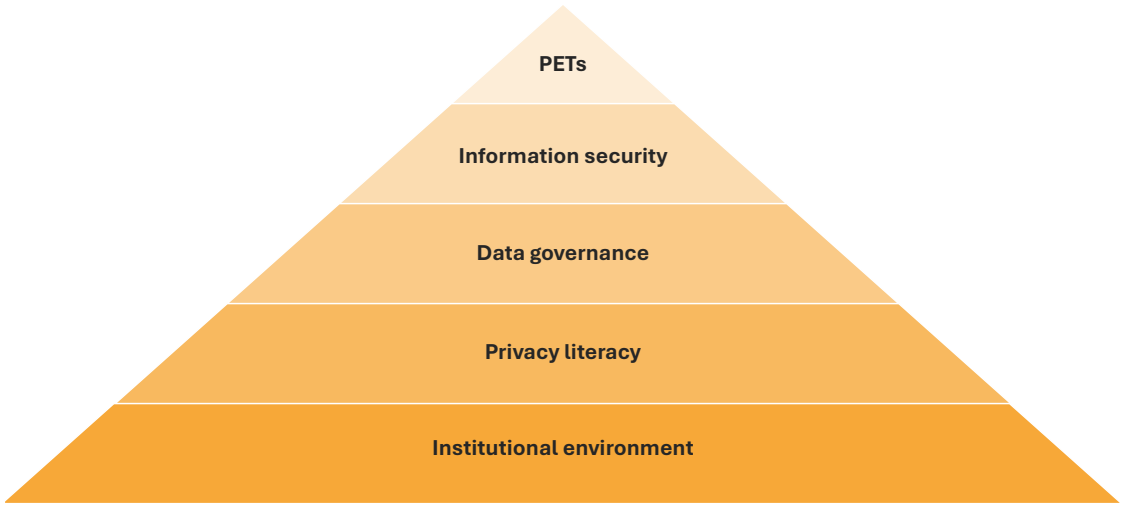
This perception reinforces the need to organize the different elements that support data protection in AI projects—from governance and information security to the adoption of PETs themselves—in order to build a clearer framework for how these technologies can be effectively applied in the Brazilian context.

This convergence made it possible to systematize the findings into the **Layered Privacy Model**, which organizes and gives structure to the patterns identified throughout the interviews. It is an exploratory model, not derived from normative assumptions established a priori, but emerging directly from the thematic analysis of the collected empirical material.

> **Its value lies in the ability to organize elements already present in professional practice, offering a structured representation that can guide both companies and policymakers.**

In academic literature, **models are understood as simplified and functional representations of reality** (Bhattacherjee, 2025; Martins, 2005). Unlike theories, models do not aim to provide complete explanations of phenomena but capture specific relationships between variables to guide analyses and predictions.

Policymakers also frequently use analytical models to structure evaluation processes, and it is from this perspective that the proposal developed here should be understood: **a practical tool to organize a field still characterized by conceptual and methodological dispersion.**



Source: author's own elaboration

The Layered Privacy Model is a conceptual framework designed to guide the adoption of PETs in AI projects. It organizes data protection across five interrelated levels—institutional environment, literacy, governance, information security, and PETs themselves. The logic is progressive: each layer establishes the conditions for privacy-enhancing technologies to be applied consistently, reducing risks and reinforcing responsible practices.

The **Layered Privacy Model** emerges directly from the qualitative empirical research conducted in this study: its layers are derived from the interviews and are interdependent:

- **Institutional Environment:** encompasses the regulatory context and external pressures that shape decisions regarding the use of personal data.;

> *"When we analyze projects in organizations operating* **in highly regulated sectors, with greater concern for privacy***, it is more common to observe the use of this type of protective technology."* ***[Interviewee E]***

- **Literacy:** covers the level of technical and organizational knowledge, frequently identified by interviewees as essential for the full adoption of advanced protection techniques.

> *"The* ***biggest challenge is related to ethics, governance, and education*** *regarding the use of these technologies: to what extent are they really useful, and to what extent is human discernment still necessary?* ***To address this, it is not enough to have tools; people need to be trained, educated, and empowered." [Interviewee I]***

- **Governance:** refers to internal processes of coordination and strategic alignment, necessary to avoid fragmented or disjointed technological implementations.

> *"***Governance** *involves ensuring that if there is data such as CPF, it will be anonymized in some hash, and can only be decrypted after a request goes through the legal or compliance team, some type of governance."* ***[Interviewee A]***

- **Information Security:** constitutes a transversal foundation, as the protection of personal data is hardly sustainable without robust cybersecurity practices.

> *"There is concern with the* ***infrastructure of these environments, linked to traditional information security,*** *which remains relevant to prevent attacks by malicious actors."* ***[Interviewee G]***

- **PETs** form the technological core itself, providing concrete instruments to mitigate privacy risks. Without the prior layers, the effectiveness of PETs is unlikely to be fully realized.

> *"The most advanced PETs mitigate risks more effectively than the traditional technologies* ***used in industry." [Interviewee B]***

Although exploratory, the model allows the structuring of a practical framework for companies that develop and use AI solutions. Its progressive layered structure makes it easier to visualize that organizations may be at different levels of maturity, with progress achieved through the coordinated strengthening of these dimensions.

In practice, the model can be used for:

Organizational self-assessment: mapping which layer the company has the strongest foundation in (e.g., robust information security policies) and where there are weaknesses (e.g., lack of training in PETs).

Defining investment priorities: aligning resources to critical areas, such as creating privacy literacy programs before adopting advanced technologies.

Regulatory and institutional planning: evaluating whether the normative environment and contracts allow for the secure use of PETs (e.g., data-sharing clauses).

Progressive implementation of PETs: starting with more accessible PETs (e.g., anonymization and pseudonymization) and advancing to sophisticated techniques (e.g., homomorphic encryption, federated learning).

Integration with corporate governance: incorporating privacy maturity metrics into compliance or risk audit reports.

4. Conclusion

The advancement of AI and the intensive use of data present new challenges for protecting the privacy of data subjects. In this context, **PETs emerge as essential tools to mitigate risks**, offering technical solutions that reduce the exposure of personal data across various stages of the information lifecycle.

However, the findings of this study reveal that **these technologies are not sufficient when applied in isolation.** The effective implementation of PETs must be integrated into a **broader data governance perspective,** which includes continuous training, organizational awareness, and privacy literacy. Without these complementary elements, even the most sophisticated solutions remain vulnerable.

It is important to emphasize that this study did not conduct a technical or operational analysis of the individual effectiveness of PETs. Our objective was to **highlight these technologies** by mapping expert perspectives and demonstrating their role as instruments to mitigate privacy risks in the Brazilian context. By introducing the topic into the public debate, we also aim to encourage greater **engagement from companies, regulators, and society**, fostering investment and contributing to the maturation of these solutions in the country.

Finally, protecting privacy in AI systems requires more than innovative technologies—it demands a **robust institutional environment** grounded in effective governance, clear regulations, good practices, and investment in training. Only this combination will allow PETs to realize their potential in strengthening trust in the responsible use of data and consolidating a regulatory ecosystem prepared for the challenges of AI.

5. Direction for future studies

This study addressed the topic of PETs and analyzed their potential application in data protection within AI systems. Nevertheless, several questions remain open. The following highlight gaps that can guide future research and contribute to the ongoing regulatory debate, expanding knowledge about these technologies and their impacts in the Brazilian context:

Standardization of language and effects on adoption: The lack of uniform terminology for PETs in Brazil was a recurring theme in the interviews. Future studies could investigate how regulatory bodies, sector associations, and standardization organizations can disseminate clear nomenclatures and encourage the adoption of these technologies.

Practical application of the model: Gaps remain regarding the technical functioning and operation of PETs in real-world scenarios. Future research, developed with the support of STEM specialists, could map use cases in Brazil, testing costs, technical barriers, and regulatory impacts arising from the application of these technologies.

Political economy of adoption: The high cost of PETs and the concentration of expertise in large international companies create asymmetries. New studies could explore incentives, financing mechanisms, and the role of the state in democratizing their adoption.

Effects on the debate about legitimate interest: PETs may influence the legal interpretation of legitimate interest as a legal basis for data processing. Research could examine whether the use of these technologies strengthens arguments of proportionality and necessity in Brazilian regulatory contexts.

References

CABALLAR, Rina Diane; STRYKER, Cole. **O que é aprendizado federado?** 2025. Available at: https://www.ibm.com/br-pt/think/topics/federated-learning. Accessed on: 28 jul. 2025.

CENTRE FOR INFORMATION POLICY LEADERSHIP (CIPL). **Privacy-Enhancing and Privacy-Preserving Technologies in AI: Enabling Data Use and Operationalizing Privacy by Design and Default.** 2025. Available at: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_pets_and_ppts_in_ai_mar25.pdf. Accessed on: 10 set. 2025.

EDIN, Per et al. **Quantifying the GenAI opportunity: Lessons learned from benchmarking 17 million+ companies worldwide.** 2025. Available at: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/quantifying-genai-opportunity.pdf. Accessed on: 04 set. 2025.

FERETZAKIS, Georgios; PAPASPYRIDIS, Konstantinos; GKOULALAS-DIVANIS, Aris; VERYKIOS, Vassilios S.. Privacy-Preserving Techniques in Generative AI and Large Language Models: a narrative review. **Information,** [S.L.], v. 15, n. 11, p. 697, 4 nov. 2024. MDPI AG. http://dx.doi.org/10.3390/info15110697. Available at: https://www.mdpi.com/2078-2489/15/11/697. Accessed on: 19 set. 2025.

GENTRY, Craig. Fully homomorphic encryption using ideal lattices. **Proceedings Of The Forty-First Annual Acm Symposium On Theory Of Computing**, [S.L.], p. 169-178, 31 maio 2009. ACM. http://dx.doi.org/10.1145/1536414.1536440. Available at: https://dl.acm.org/doi/10.1145/1536414.1536440. Accessed on: 17 set. 2025.

GIUFFRÈ, Mauro; SHUNG, Dennis L.. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. **Npj Digital Medicine,** [S.L.], v. 6, n. 1, p. 1-8, 9 out. 2023. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/s41746-023-00927-3. Available at: https://www.nature.com/articles/s41746-023-00927-3#citeas. Accessed on: 11 set. 2025.

GOLDMAN SACHS. **Generative AI could raise global GDP by 7%**. 2023. Available at: https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent. Accessed on: 04 set. 2025.

IBM. **What is synthetic data?** Available at: https://www.ibm.com/think/topics/synthetic-data. Accessed on: 23 jul. 2025.

INGOLD, Jo; MONAGHAN, Mark. Evidence translation: an exploration of policy makers use of evidence. **Policy & Politics**, [S.L.], v. 44, n. 2, p. 171-190, abr. 2016. Bristol University Press. http://dx.doi.org/10.1332/147084414x13988707323088. Available at: https://bristoluniversitypressdigital.com/view/journals/pp/44/2/article-p171.xml. Accessed on: 16 set. 2025.

KANDPAL, Nikhil; WALLACE, Eric; RAFFEL, Colin. Deduplicating Training Data Mitigates Privacy Risks in Language Models. **Arxiv**, [S.l.], p. 1-11, dez. 2022. Available at: https://arxiv.org/abs/2202.06539. Accessed on: 19 set. 2025.

MARR, Bernard. **The Difference Between Generative AI And Traditional AI: An Easy Explanation For Anyone**. 2023. Available at: https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-tradit... Accessed on: 19 set. 2025.

MICROSOFT LEARN. **Geração de dados sintéticos no portal do Azure AI Foundry**. Available at: https://learn.microsoft.com/pt-br/azure/ai-foundry/concepts/concept-synthetic-data. Accessed on: 23 jul. 2025.

NICFAB. **Privacy Enhancing Technologies (PETs): an evergreen category - part 1**. 2023. Available at: https://notes.nicfab.eu/en/posts/pet01/?utm_source=chatgpt.com. Accessed on: 19 set. 2025.

RADAR DE TENDÊNCIAS TECNOLÓGICAS. **Tecnologias de Aprimoramento de Privacidade.** 2024. Available at: https://radar.apps.bb.com.br/tendencia/Radar-2024/Temas/TechGuardian/Tecnologias-de-Aprimoramento-de-Privacidade/21001. Accessed on: 23 jul. 2025.

RUIZ, Evandro Eduardo Seron. **Criptografia homomórfica: essa técnica resolve o problema da segurança dos dados pessoais?.** 2022. Available at: https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/375269/criptografia-homomorfica. Accessed on: 22 jul. 2025.

STRYKER, Cole. **What are LLMs?** 2025. Available at: https://www.ibm.com/think/topics/large-language-models. Accessed on: 19 set. 2025.

STRYKER, Cole; SCAPICCHIO, Mark. **O que é IA generativa?** 2024. Available at: https://www.ibm.com/br-pt/think/topics/generative-ai. Accessed on: 19 set. 2025.

TEOFILO, Davi; ROCILLO, Paloma. **CLOUD Act: um caso de Direitos Humanos e Jurisdição.** Available at: https://irisbh.com.br/cloud-act-um-caso-de-direitos-humanos-e-jurisdicao/. Accessed on: 11 set. 2025.

VAN AUDENHOVE, Leo; DONDERS, Karen. Talking to People III: Expert Interviews and Elite Interviews. In: BULCK, Hilde van Den; PUPPIS, Manuel; DONDERS, Karen; VAN AUDENHOVE, Leo (ed.). **The Palgrave Handbook of Methods for Media Policy Research.** [S.l.]: Palgrave Macmillan Cham, 2019. p. 179-197.

Methodology Annex

Title	**Layered Privacy:** Privacy-Enhancing Technologies (PETs) and Their Role in AI
Research question	How do experts in AI, privacy, and cybersecurity assess the use of Privacy Enhancing Technologies (PETs) in artificial intelligence systems and their effectiveness in protecting personal data?
Methodology summary	This research adopts a qualitative and exploratory approach, combining primary data collection through semi-structured qualitative expert interviews with secondary data collection (documents, literature, and practical cases). Data analysis followed the reflexive thematic analysis technique with the aid of Atlas.ti software. Visual tools within the software were used to identify central patterns and themes, which were later validated against the empirical corpus.
Data Collection	The study employed the expert interviews methodology (Audenhove & Donders, 2019), conducting semi-structured qualitative exploratory interviews. This method was chosen due to the technical nature of the topic and the lack of conceptual standardization, making the expertise of professionals in the field in Brazil essential. The sample was composed according to diversity and representativeness criteria, including minimum participation of women; representatives from academia or research centers; professionals from Brazilian companies; and experts from large companies and technology consultants. Participants were selected through active LinkedIn search as the main strategy, complemented by convenience sampling and snowballing to expand the network of AI, privacy, and cybersecurity specialists. Of the 37 people contacted, 12 agreed to participate, 7 reported unavailability, and 18 did not respond. Interviews were conducted online (via Teams) between July 29 and August 29, 2025, lasting between 45 and 60 minutes. All were conducted by at least one Reglab researcher, following a questionnaire attached to the study. One interview was conducted as a pilot to test the script and validate initial hypotheses. The remaining 11 interviews were considered sufficient for theoretical saturation, as in qualitative semi-structured in-depth approaches, thematic recurrence and analytical density generally consolidate with few participants (Guest et al., 2006). Interviews were recorded with participants' consent, fully transcribed, and accompanied by interviewer memos. All material was stored and coded in Atlas.ti, with participant names and institutions anonymized.
Data analysis	Data analysis followed the reflexive thematic analysis technique (Braun & Clarke, 2006), suitable for exploratory qualitative studies in high-complexity contexts. This approach prioritizes contextualized interpretation rather than exhaustive coding, allowing the use of open analytical strategies. All interviews were fully transcribed and processed in Atlas.ti, which performed an initial intentional automated coding round. This procedure generated 719 first-level codes and 23 second-level codes, which were later manually reviewed by the research team In the next stage, various visual tools within the software (concept clouds, maps, correlation graphs, chatbots) were used to identify patterns and relationships between codes. This process led to the emergence of central themes, which were tested and validated against the empirical corpus, ensuring adherence to the original data. Analysis was conducted between August 29 and September 5, 2025.

Bias Reduction Procedures

Consolidated theoretical-methodological references: data collection and analysis techniques followed practices recognized in academic literature. The methodology was discussed internally before and after preliminary interviews, allowing incorporation of feedback into the final research design.

Complementary verification tool: given that initial coding was performed with software, a second tool (NotebookLM) was used to check the consistency of Atlas.ti codings and identify blind spots. This was performed by researchers who directly conducted interviews to capture nuances potentially overlooked in automated coding.

Method triangulation: empirical findings were contrasted with secondary source document analysis to compare, validate, and reinforce the consistency of interpretations derived from interviews. These references were explicitly cited throughout the text when used.

Independent dual analysis: two researchers cross-reviewed all codes and themes, reducing individual biases. Final theme definitions were made collectively, ensuring multiple perspectives and bias control in data interpretation.

Documentation and methodological transparency: all analytical steps were documented, including successive file versions and coding decisions. This practice ensures traceability according to Reglab guidelines for transparency and replicability.

Other Methodological Limitations

Initial automated coding: although Atlas.ti is a well-established software for qualitative analysis and coding was validated with a second tool, automated coding may have introduced noise in the initial phase, which could constitute blind spots in later analysis.

Dependence on external tools: part of the analysis relied on proprietary software, potentially limiting replicability in other contexts.

Qualitative scope: findings reflect perceptions from a limited set of interviews with analytical depth but are not intended for statistical generalization.

Convenience sampling: selection may reflect availability and professional network biases, despite diversity criteria.

Technological evolution: results reflect the state of the art of AI tools and practices at the time of research. Rapid changes in the field may affect some conclusions.

Uses of softwares

SOFTWARE	USE IN THE RESEARCH
Suíte MS Office	Text editing, spreadsheets and charts, conducting interviews (Teams)
Suíte Adobe C	Layout and finalization of charts and illustrations
Atlas.ti	Organization, coding, and analysis of qualitative data
Cockatoo	Audio transcription of interviews into text
ChatGPT 5o	Brainstorming, information systematization, grammatical review (spelling, grammar, synonym search), language adjustment, compliance with Reglab Writing Manual
Notion AI	Text editing and review (spelling, grammar, synonym search, language adjustment, translations), research organization, and schedule structuring
Lex.page	Text review (brevity, clichés, readability, passive voice, unsupported statements, repetitions)
More UFSC	Generation of bibliographic references in ABNT format

Ethical Guidelines	**Research funding:** this publication is part of a series of studies sponsored by Google, Meta, and b/luz, with Reglab maintaining full editorial control. Unlike commissioned research, Reglab independently defined the scope, objectives, and methodology of this study. The authors maintained complete professional independence and assume full responsibility for the content and conclusions presented. **Personal data treatment:** the research involved the processing of personal data only during collection and analysis, in a limited and proportionate manner aligned with the study's objectives, in accordance with Law No. 13,709/2018 (LGPD). **Legal basis:** all participants formally authorized their participation by signing a consent form, acknowledging the research objectives and the use of their data. **Purpose and adequacy:** data were used exclusively for the purposes of this research, as per the consent obtained, and not for other purposes. **Minimization and anonymization:** personally identifiable information not relevant to the study's objectives was anonymized in transcripts and removed from the active dataset. **Confidentiality:** in presenting results, data were kept confidential, and citations were adjusted as needed to protect source confidentiality. Only a limited number of researchers directly involved in the project had access to personal data and original documents. **Recordkeeping and information security:** files were stored with password-protected access in accordance with Reglab's internal information security policies. **Retention and disposal:** data will be stored for up to 12 months solely for methodological audit purposes and possible replication, after which they will be deleted. **Responsible use of public data:** although some analyzed data are public, their use was conducted responsibly and ethically, exclusively for independent research purposes. **Methodological transparency:** the research methodology was described in detail to ensure transparency and replicability, contributing to scientific integrity and enabling independent validation of results. **Non-discrimination and respect for diversity:** the research was conducted with respect for diversity and avoiding any form of discrimination.

ANNEX II - SEMISTRUCTURED INTERVIEWS SCRIPT

#	SCRIPT
1	To start, could you tell us a bit about your experience with AI projects? More specifically, your experience with privacy or data protection issues in the context of AI?
2	In your view, how is personal data actually used in AI models? From a technical standpoint, does it lose its identifiability during the process?
3	In your experience, what are the main techniques used today to reduce privacy risks and improve cybersecurity for personal data in AI projects?
4	I will mention some technologies. Are you familiar with or have you worked with any of them? Could you comment, if you wish, on their practical relevance? • Differential privacy • Trusted Execution Environment (TEE) • Synthetic data • Federated learning • Homomorphic encryption
5	In your view, does the use of PETs in AI systems actually eliminate privacy risks, or do relevant concerns remain, such as re-identification or data leakage?
6	In your daily work, is the term “PETs” commonly used? Or do people more often refer directly to specific technologies?
7	Have you seen cases where the use of PETs enabled projects that would otherwise have been unfeasible or high risk due to privacy concerns?
8	Have you encountered projects involving sensitive data—such as health, racial origin, or children’s data—where the use of PETs was considered (or adopted) to enable data collection or reduce risks? How was this decision handled?
9	What do you consider the main challenges to increasing the awareness and adoption of PETs in AI? Have you personally faced any of these obstacles in any project?
10	Would you like to highlight any points that were not addressed, or leave a recommendation for future research in this area? Can you recommend someone else to participate in the interviews?


reglab
center for strategy
& regulation