Privacidade em Camadas

O Papel de Privacy Enhancing Technologies (PETs) em Sistemas de Inteligência Artificial

reglab
centro de estratégia & regulação

Sobre o Reglab

Somos um **centro de pesquisa privado especializado no setor de mídia e tecnologia**, que auxilia empresas, associações e formuladores de políticas a tomarem decisões estratégicas baseadas em dados e evidências.

Saiba mais em www.reglab.com.br

Sobre a Série Policy Briefs

A **Série Policy Briefs** engloba estudos que avaliam tendências, políticas públicas existentes ou propostas legislativas, utilizando dados qualitativos e quantitativos para informar e orientar decisões. O objetivo é trazer temas complexos de forma acessível, destacando os principais pontos de análise, impactos e possíveis recomendações.

Expediente

Diretor Executivo: Pedro Henrique Ramos

Coordenadora de Pesquisa: Marina Garrote

Autores(as): Pedro Henrique Ramos e Daniela Naomi Shimabukuro Nomura

Pesquisadores: Marina Garrote, Stephanie Souza, Giulia Brombine e Vinícius Pimenta

Diagramação Final: Eliza Natsuko Shiroma

Citação Sugerida: RAMOS, P.H.; NOMURA, D. N. S.Privacidade em Camadas: O Papel de Privacy Enhancing Technologies (PETs) em Sistemas de Inteligência Artificial. Policy Briefs Reglab n.3. São Paulo: Reglab, 2025.

Sumário Executivo

A adoção de Inteligência Artificial Generativa e de sistemas baseados em dados pessoais trouxe novas camadas de complexidade ao debate sobre privacidade. Neste estudo inédito no Brasil, o Reglab investigou como as *Privacy Enhancing Technologies* **(PETs)** podem ser aplicadas para mitigar riscos no treinamento e operação desses sistemas, a partir da visão de especialistas em IA, proteção de dados e cibersegurança.

As entrevistas qualitativas revelaram **consenso sobre o papel central das PETs** na redução de riscos de reidentificação, abuso ou exposição indevida. Embora não eliminem totalmente as vulnerabilidades, essas tecnologias oferecem **camadas adicionais de proteção** que tornam o tratamento de dados mais confiável e capaz de transmitir maior segurança a organizações que operam com soluções de IA.

Entre os principais achados, destacam-se:

- **Mitigação de riscos:** PETs aprimoram a proteção de dados pessoais, mas, atualmente, não são capazes de extinguir os riscos existentes à privacidade;
- **Viabilização de projetos:** tecnologias como aprendizado federado e ambientes de execução confiáveis tornam possível iniciativas de alto valor social, como pesquisas em saúde, que antes seriam inviáveis devido a riscos à privacidade;
- **Limitações persistentes:** especialistas ressaltam que não há "risco zero" em cibersegurança e que muitas soluções ainda carecem de testes amplos, o que limita a confiança do mercado e atrasa investimentos;
- **Fragmentação no Brasil:** o uso prático de PETs no país ainda é pontual e pouco sistematizado, dificultando ganhos de escala e integração com estratégias organizacionais mais amplas.

A partir desses elementos, o estudo propõe o **Modelo da Privacidade em Camadas**, framework inédito que organiza a proteção de dados em cinco níveis interligados:

i. **Ambiente institucional** (regulação e pressões externas);
ii. **Literacia** (educação e capacitação em privacidade);
iii. **Governança** (processos internos de alinhamento estratégico);
iv. **Segurança da informação** (infraestrutura de cibersegurança robusta);
v. **PETs** (ferramentas técnicas de mitigação de riscos).

Fonte: elaboração própria

Esse modelo permite visualizar diferentes graus de maturidade das organizações e orientar tanto **empresas** na autoavaliação e definição de prioridades de investimento, quanto **formuladores de políticas públicas** na construção de ambientes regulatórios que incentivem a adoção responsável de tecnologias de preservação de privacidade.

A principal contribuição deste estudo é mostrar que **PETs só realizam seu potencial quando integradas a uma estratégia mais ampla de governança, literacia e segurança.** Isoladamente, essas tecnologias não são suficientes para garantir proteção plena, mas em conjunto podem consolidar um ecossistema regulatório e empresarial mais robusto, alinhado ao uso responsável de dados.

SUMÁRIO

1. Introdução

A importância da inteligência artificial (IA) na sociedade não é mais objeto de debate: tornou-se um fato consolidado. Trata-se de um elemento transformador em várias dimensões: na economia, com potencial para impulsionar o PIB mundial em até 7 trilhões de dólares e elevar o crescimento da produtividade em 1,5% ao longo de 10 anos (Goldman Sachs, 2023); e na gestão organizacional, podendo reduzir em 10% o tempo dedicado à gestão de pessoas (Edin et al., 2025), apenas para citar alguns exemplos.

O desafio atual parece não estar somente em reconhecer os benefícios da IA, mas em **desenvolver modelos de governança que acompanhem sua acelerada disseminação e os processos disruptivos que ela desencadeia.** E um dos temas mais complexos é a relação dessa tecnologia com privacidade e proteção de dados pessoais.

À medida que diferentes aplicações se integram ao cotidiano e às decisões críticas da sociedade, **questões sobre como proteger dados pessoais e cumprir legislações sobre o tema** têm surgido com cada vez mais frequência em manchetes, decisões judiciais e discussões em fóruns jurídicos. Por outro lado, parece haver **um avanço tecnológico igualmente relevante no campo de soluções técnicas de segurança da informação.** Entre esses campos, não é incomum a percepção de que há um abismo de conhecimento entre ambos.

Esse estudo se propõe a ser **uma ponte entre o campo de proteção de dados e os avanços em tecnologias de aprimoramento de privacidade** - também conhecidas, em inglês, como PETs. Queremos explorar como essas tecnologias podem redefinir os contornos da governança de dados em sistemas de IA, especialmente em um cenário em que marcos regulatórios buscam acompanhar a velocidade do desenvolvimento tecnológico.

1.1. O que é IA e por que isso importa?

Como em qualquer trabalho de pesquisa, é preciso definir claramente nosso objeto de análise. Neste estudo, adotamos as seguintes definições:

- **Sistemas de IA:** sistemas de software que processam e analisam dados por meio de algoritmos e modelos matemáticos, utilizando técnicas estatísticas e computacionais para identificar padrões, fazer previsões e gerar resultados específicos. Para fins didáticos, podem ser classificados em dois tipos principais:
  - **Sistemas de IA Analítica:** sistemas projetados para resolver problemas específicos a partir de conjuntos de dados estruturados, operando dentro de parâmetros pré-definidos. Funcionam como “calculadoras sofisticadas” que executam tarefas determinísticas, como classificação, previsão ou reconhecimento de padrões existentes (Barr, 2023). Exemplos incluem sistemas de detecção de fraudes bancárias e recomendações personalizadas de assistentes virtuais.

- **Sistemas de IA Generativa:** sistemas que empregam técnicas estatísticas e de aprendizado de máquina para gerar novos conteúdos, como textos, imagens e vídeos (Barr, 2023). São baseados em grandes modelos de linguagem (LLMs) que transformam dados de treinamento em representações matemáticas (vetores) que capturam padrões e correlações estatísticas. Atualmente, as aplicações mais populares são modelos de chatbots (Stryker; Scapicchio, 2024).

Fonte: elaboração própria, a partir de Ramos (2023)

Dados pessoais são informações relacionadas a uma pessoa natural identificada ou identificáveis e que podem ser tratados de diversas maneiras em sistemas de IA, seja como dados de entrada inseridos diretamente por um usuário, ou como parte dos conjuntos analisados durante o treinamento. Vamos explorar essa relação no item 2.1 adiante.

1.2. A Proposta Metodológica desta Pesquisa

Entre 2016 e 2024, **diversos países desenvolveram ou atualizaram suas legislações de proteção de dados pessoais.** A maioria dessas leis, como a Lei Geral de Proteção de Dados (LGPD), de 2018, foi estabelecida antes da disseminação das aplicações de IA generativa e, consequentemente, nem sempre oferecem orientações claras para desafios emergentes.

Neste cenário, é **importante investigar os conhecimentos práticos sobre como o próprio mercado,** que funciona com dinamismo e velocidade diferentes dos legisladores, tem desenvolvido soluções tecnológicas para a proteção de direitos no campo de IA adicionando camadas de segurança para reduzir riscos e fomentando a confiança na sua adoção.

> **Esta pesquisa examina a relação entre proteção de dados e PETs em sistemas de IA no Brasil. Nosso objetivo é indicar caminhos que apoiem o desenvolvimento de boas práticas e diretrizes de *privacy by design*, facilitando a incorporação da privacidade desde as etapas iniciais do desenvolvimento e utilização de sistemas de IA por empresas brasileiras.**

O estudo se baseia em uma das principais premissas metodológicas do Reglab: **a abordagem de policy translation**, ainda pouco explorada na governança digital no Brasil. Trata-se de uma metodologia que enfatiza o **processo ativo de interpretar e adaptar descobertas complexas de pesquisa em formatos que sejam compreensíveis, relevantes e aplicáveis por formuladores de políticas públicas** (Ingold; Monaghan, 2016).

Quando utilizamos quadros coloridos, gráficos, exemplos e anedotas, fazemos isso de forma intencional. Reconhecemos o risco de possíveis imprecisões técnicas, mas entendemos que traduzir evidências complexas em conhecimento aplicado exige tornar o conteúdo mais claro e acessível. Esta é uma escolha metodológica necessária — e um posicionamento que assumimos com total transparência.

Para nossa metodologia de coleta de dados, optamos por uma abordagem diferente das revisões bibliográficas e pesquisas documentais convencionais: as **entrevistas qualitativas em profundidade.** Inspirados por métodos de estudos de recepção, buscamos entender como **profissionais que enfrentam diariamente desafios técnicos de cibersegurança** compreendem o tratamento de dados pessoais em sistemas de IA, quais ferramentas PETs estão disponíveis, como são efetivamente implementadas na prática e de que maneira contribuem para uma melhor proteção da privacidade dos indivíduos.

Ao longo de um mês, realizamos 11 entrevistas com experts, focando em profissionais de nível sênior e com experiência prática em temas de cibersegurança e compliance em proteção de dados pessoais. As entrevistas seguiram roteiros pré-definidos e protocolos de confidencialidade sendo suas transcrições e memoriais avaliados por meio do software Atlas.ti a partir da técnica de análise temática.

Entrevistado(a)	Descrição
A	Homem, setor empresarial, engenheiro de machine learning
B	Homem, setor acadêmico, professor de cibersegurança
C	Homem, setor acadêmico, professor de cibersegurança
D	Homem, setor empresarial, consultor em engenharia de sistemas
E	Homem, setor empresarial, consultor em cibersegurança
F	Homem, setor empresarial, pesquisador em cibersegurança
G	Mulher, setor empresarial, consultora em gestão de privacidade e segurança
H	Homem, setor empresarial, consultor em cibersegurança
I	Mulher, setor empresarial, consultora em segurança da informação e privacidade
J	Mulher, setor empresarial, consultora em privacidade
K	Mulher, setor acadêmico, cientista de dados

A metodologia completa, com detalhes sobre os procedimentos adotados, encontra-se no final do estudo.

2. Resultados Principais

2.1. Como dados pessoais são usados nos modelos de IA?

CONCEITOS FUNDAMENTAIS

Na IA Analítica, dados pessoais são geralmente tratados de forma estruturada, em bases delimitadas e mantidas pelas organizações. Esse formato aumenta obrigações de proteção, mas facilita a aplicação dos princípios de minimização, finalidade e base legal, já que os contextos de uso são definidos previamente.

Na IA Generativa, grandes modelos de linguagem (LLMs) são treinados em volumes massivos de dados coletados na internet. Esses dados são fragmentados em unidades chamadas *tokens* e convertidos em representações matemáticas (*vetores*), que capturam relações estatísticas entre palavras e frases.

Em regra, os modelos não armazenam diretamente bases de dados pessoais. Contudo, como funcionam por padrões estatísticos, informações muito repetidas no treinamento podem ser "lembradas" e reproduzidas nos resultados, já que vetores funcionam como representações de conhecimento.

Modelos também podem gerar dados pessoais inexistentes ou não vistos no treinamento. Nesses casos, trata-se de inferências estatísticas, e não de memória. É o que ocorre, por exemplo, quando o modelo gera combinações de números com o formato de um CPF válido, mesmo sem ter armazenado aquele dado.

As entrevistas com os experts ajudaram a entender como funciona o uso de dados pessoais e como eles são anonimizados nos processos de tratamento. Contudo, embora pareça haver um conhecimento uniforme quando os entrevistados falam de sistemas de IA Analítica, observamos divergências e *gaps* de conhecimento ao aprofundarmos no funcionamento dos sistemas de IA Generativa e, particularmente, dos LLMs — que receberam mais destaque nas entrevistas.

Uso de dados – IA Analítica

Na chamada **IA Analítica**, os entrevistados ressaltaram que os dados pessoais são tratados de forma mais **estruturada, delimitada e vinculada diretamente às finalidades do projeto.**

Dados Estrurados: em algum momento do tratamento, as informações são organizadas em formatos padronizados que permitem classificação ou predição, como tabelas em bancos de dados ou conjuntos de variáveis numéricas.

Base de Dados Delimitada: quando se utilizam dados pessoais, esses registros estão associados a indivíduos de forma explícita (ex.: CPF, prontuário médico, número de conta) ou por meio de técnicas como a pseudonimização, que substitui identificadores diretos por códigos.

Finalidade definida: os dados são processados para resolver problemas específicos, como análise de risco de crédito, detecção de fraudes em transações financeiras ou apoio a diagnósticos médicos assistidos.

> *"Na prática, esses modelos são funções matemáticas que estão tentando encontrar padrões nos dados. Um exemplo clássico é você ter lá um dado estruturado no formato de tabela, que você utiliza como insumo e vai utilizar uma função específica para interpretar aqueles dados"* ***[Entrevistada K]***[1]

Em geral, os entrevistados destacaram que a IA Analítica oferece menor escala e maior precisão: o volume de dados pode ser menor, mas a **qualidade e relevância** da informação individual costuma ser mais determinante para a performance do modelo.

Contudo, alguns entrevistados ressaltaram que esses modelos possuem riscos maiores de reidentificação, sendo, comparativamente com modelos de LLM, até *mais perigosos para a privacidade* ***[Entrevistado B]***.

Isso acontece porque os dados permanecem em bases estruturadas e, ainda que pseudonimizados, podem ser facilmente reconectados a indivíduos quando cruzados com outras fontes. Esse funcionamento é diferente dos LLMs, que tendem a diluir padrões em representações estatísticas de larga escala, como veremos adiante.

Uso de Dados - IA Generativa e o papel dos LLMs

Seja de forma explícita ou implícita, diferentes entrevistados explicaram o uso de dados pessoais em modelos de IAG a partir de duas fases:

- **Treinamento:** momento em que grandes volumes de dados, potencialmente contendo informações pessoais, são coletados e transformados em *tokens* para gerar representações estatísticas. O risco está na eventual memorização de trechos recorrentes, que podem reaparecer nas respostas.

- **Inferência:** etapa em que o modelo gera saídas a partir de novos prompts do usuário. Aqui não há acesso direto ao banco de dados original, mas sim a recombinações estatísticas. Ainda assim, informações pessoais podem emergir tanto por memorização quanto por inferências plausíveis produzidas pelo padrão do modelo.

EXPLICAÇÃO DIDÁTICA

Treinamento: é como quando alguém estuda para uma prova lendo dezenas de livros e "aprende" padrões a partir do conhecimento adquirido. Se uma informação aparecer repetidamente nesse material (e.g. a tabela periódica), o modelo vai acabar "decorando" e reproduzindo essa informação posteriormente.

Inferência: é como usar esse conhecimento em uma prova sem consulta: o modelo não busca nos livros originais, mas combina estatísticas para criar a resposta. Assim, ele pode tanto repetir algo decorado quanto inventar algo novo que parece verdadeiro.

1 Com o objetivo de preservar o anonimato e a confidencialidade dos participantes da pesquisa, foram realizadas modificações pontuais nas citações apresentadas neste estudo. Em determinadas circunstâncias, procedeu-se a adaptações linguísticas específicas para assegurar a intenção original dos entrevistados na transcrição textual. A preservação do registro discursivo foi mantida sempre que possível, respeitando os princípios metodológicos estabelecidos.

No treinamento, os entrevistados explicaram que os LLMs operam pela análise massiva de dados, convertendo textos em representações numéricas. Esses sistemas não funcionam como bancos de frases completas nem como repositórios de dados pessoais brutos: o que internalizam são **padrões estatísticos de linguagem**, isto é, as frequências e relações entre palavras e expressões (Stryker, 2025).

Na prática, isso significa que o **modelo prioriza correlações estatísticas em vez de registros individuais.** Assim como alguém que lê mil currículos não guarda todos os nomes, mas percebe padrões de carreira (por exemplo, que pessoas formadas em Administração costumam trabalhar em empresas privadas), um LLM absorve tendências de uso de palavras.

Contudo, há situações em que informações específicas podem ser reproduzidas, sobretudo quando são muito frequentes no material de treinamento (Kandpal; Wallace; Raffel, 2022):

> *"O modelo só sabe quem é o Harry Potter porque existem dezenas de milhares de páginas na internet mencionando o personagem. Isso está relacionado ao que chamamos de* ***relevância estatística****. Treinar ou não o modelo com os livros originais do Harry Potter é praticamente irrelevante, pois essas passagens já estão replicadas em milhares de outros sites que foram utilizados durante o treinamento"* ***[Entrevistado D]***

Há também pontos de atenção na etapa de **inferência**. Durante a inferência, portanto, dados pessoais podem ser tratados de forma transitória, especialmente quando usuários inserem dados pessoais em prompts ou compartilham documentos. Essas informações são processadas temporariamente pelo modelo para gerar resposta, **mas não se tornam parte permanente do modelo-base**, embora possam permanecer em memória de curto prazo ou serem usadas para personalização do perfil do próprio usuário.

> *"As pessoas têm essa sensação de que toda IA está aprendendo o tempo inteiro, né? E não é verdade isso. Na prática, ela não está aprendendo nada, principalmente nesses ambientes mais homologados, que é praticamente descartável toda interação que a gente tem do ponto de vista que foge ao usuário, fica ali dentro daquele usuário"* ***[Entrevistado H]***

Outro ponto recorrente nas entrevistas foi a capacidade da IA Generativa de produzir dados pessoais na saída, **mesmo que esses registros não tenham aparecido de forma literal no treinamento.** Os entrevistados descreveram esse processo como uma "invenção" do modelo, resultado de sua capacidade estatística de combinar padrões de forma verossímil. Exemplos incluem a criação de sequências válidas de CPFs (seguindo a regra de 11 dígitos com verificador), ou combinações comuns de nomes, como "Ana Maria".

Esse ponto é relevante porque mostra que, mesmo sem funcionar como bancos de dados pessoais, modelos generativos podem produzir informações identificáveis, o que traz desafios práticos para a interpretação e aplicação de leis de proteção de dados.

Anonimização de Dados na IA Generativa

A questão mais debatida entre os entrevistados foi **a anonimização de dados em modelos de IA Generativa.** O principal ponto emergente é a dificuldade em garantir a perda total do caráter identificável dos dados pessoais, especialmente nos modelos mais avançados.

Há uma convergência entre os especialistas: **os dados não são armazenados como registros brutos, mas transformados em vetores estatísticos.** Nesse processo, nomes, números e frases se convertem em *tokens* e pesos matemáticos que passam a representar padrões de linguagem. Essa característica altera, de forma significativa, o debate sobre proteção de dados, deslocando o foco da coleta e armazenamento literal para os **riscos de reidentificação e de uso estatístico das informações.**

Nesse ponto, muitos dos especialistas, pessoas com senioridade e experiência em suas áreas de atuação, **mostraram-se cautelosas nas respostas e, em alguns casos, até mesmo admitiram não saber como o processo funciona.** Suas frases e explicações não demonstraram a mesma segurança para explicar os outros pontos de funcionamento da tecnologia.

> *"Não é que eu transformo ele em números, não é uma "pseudonimização". Ele simplesmente se dissolve em tokens [...] ele não tem banco de dados, o que ele tem é de pedaço de palavra"* ***[Entrevistado D]***
>
> *"Eu li um artigo, estava lendo um artigo sobre isso, quando eu extraio dados de um modelo de IA para outro, eu posso reidentificar pessoas. E aí a pessoa pega isso e joga em outro modelo, e outro modelo, e aí eu consigo chegar em pessoas. [...] Sim, existe um risco de reidentificar, mesmo eu aplicando criptografia, anonimização, pseudonimização, existe. Se eu começo a cruzar ali, eu consigo".* ***[Entrevistada I]***
>
> *"Eu não acredito que perca o identificável não. Você volta e ele vai te identificar uma hora. É muito louco porque eu não sei dizer. Ele fala que não tem memória, mas tem. Eu não sei explicar".* ***[Entrevistada G]***

De todo modo, o debate revelou uma conclusão comum: **é preciso aumentar a literacia e o conhecimento público sobre anonimização na IA Generativa.** As divergências entre especialistas indicam que não se trata apenas de uma questão técnica, mas também de interpretação e de gestão de riscos.

2.2. O que são PETs

CONCEITOS FUNDAMENTAIS

- **Definição:** PETs é a sigla em inglês para Privacy Enhancing Technologies, que se traduz como Tecnologias de Aprimoramento da Privacidade. São soluções de software desenvolvidas para reduzir riscos à privacidade e melhorar a cibersegurança.

- **Termo Pouco Conhecido e Utilizado:** Embora seja popular na comunidade jurídica de proteção de dados, a maioria dos entrevistados concorda que o termo "PETs" é pouco conhecido ou não é amplamente utilizado no Brasil. Contudo, a padronização ajudaria a consolidar a área e a criar referências claras sobre quais mecanismos são eficazes e em quais cenários.

- **Importância de Aumentar a Conscientização:** Os entrevistados consistentemente apontam a falta de conhecimento e a baixa literacia técnica sobre PETs como um dos principais desafios para sua adoção no Brasil. Sem conhecimento, as empresas dificilmente investem em PETs, especialmente porque o custo financeiro e computacional dessas tecnologias é elevado.

PETs é um termo guarda-chuva que reúne diferentes tecnologias e técnicas voltadas a reduzir riscos à privacidade e à cibersegurança. Seu objetivo é possibilitar o uso responsável de dados, viabilizando projetos que, sem essas ferramentas, poderiam ser considerados inviáveis ou excessivamente arriscados do ponto de vista regulatório e de proteção de dados pessoais.

Apesar do termo ser reconhecido internacionalmente e entre especialistas jurídicos em privacidade, os entrevistados apontaram que ele ainda tem pouca penetração no Brasil. Muitos argumentaram que o termo deveria receber maior visibilidade no debate nacional, harmonizando-se com a importância que já possui em fóruns internacionais – o que revela uma disparidade entre a discussão global e a aplicação do conceito no contexto brasileiro.

POR QUE PETS APARECEM MAIS NO CAMPO JURÍDICO DO QUE ENTRE ESPECIALISTAS DE CIBERSEGURANÇA?

Uma possível explicação é que o termo Privacy Enhancing Technologies surgiu inicialmente a partir de relatórios de autoridades de proteção de dados pessoais do Canadá e Países Baixos nos anos 1990, e em seguida por um famoso relatório da Comissão Europeia, de 2007 (NicFab, 2023). A partir disso, a sigla PETs consolidou-se em guias e normas como uma categoria guarda-chuva para descrever práticas de anonimização, pseudonimização e privacidade por design, onde cumpre a função de traduzir diferentes técnicas em uma linguagem normativa única.

Já no campo técnico, parece-nos que profissionais preferem nomear métodos específicos, muitas vezes considerando rótulos guarda-chuva como vagos demais, já que não esclarecem quais medidas foram aplicadas, em que condições e com que garantias.

Apesar do baixo reconhecimento do termo PETs, os entrevistados demonstraram familiaridade com tecnologias específicas como privacidade diferencial, aprendizado federado e criptografia homomórfica quando mencionadas concretamente. Diversos relatos indicaram que os profissionais tendem a se orientar mais pelo nome do fabricante ou pela marca comercial das soluções de privacidade do que pelo termo técnico da tecnologia.

> **Isso parece evidenciar como o debate sobre PETs no Brasil está mais próximo do mercado e de produtos específicos, e não fundamentado em uma linguagem padronizada ou conceitual.**

Os entrevistados apontaram repetidamente que **o conhecimento, a uniformidade e a padronização em torno do termo PETs são cruciais para sua adoção no Brasil.** Eles enfatizaram que a baixa literacia técnica e a falta de awareness sobre o tema comprometem tanto as decisões de investimento quanto a consolidação da área, e que a ausência de padronização também impede a criação de referências claras sobre a eficácia de cada mecanismo em diferentes cenários.

O alto custo financeiro e computacional dessas tecnologias, somado à falta de familiaridade, dificulta a justificativa para investimentos. Esta lacuna impacta tanto o financiamento público quanto o privado: como vários entrevistados destacaram, o Brasil carece de linhas específicas de apoio à pesquisa em PETs, o que compromete o desenvolvimento e a viabilidade dessas soluções.

> *"**Sendo bem sincero: eu conheço o termo, mas nunca vi sendo usado, seja na academia ou mesmo na indústria (...)** acho que o termo PETs talvez seja adequado porque a gente precisa começar a construir esse conhecimento, tem muito trabalho para ser feito na área, e a oficialização de uma terminologia seria uma coisa interessante"* ***[Entrevistado B]***
>
> *"**As pessoas vão falar mais ou de uma ferramenta, de uma solução, ou de uma técnica**, mas elas dizem que para mim ainda falta uma literacia nesse sentido"* ***[Entrevistada I]***

Ou seja, para avaliar riscos e potencialidades de forma mais concreta, é preciso ir além do rótulo genérico e observar as **tecnologias específicas** que compõem esse universo. São essas ferramentas, com seus limites e aplicações práticas, que permitem medir de fato a utilidade de PETs em projetos de IA.

As tecnologias a seguir estão organizadas de acordo com a frequência de menção nas entrevistas e o nível de conhecimento demonstrado pelos entrevistados sobre cada uma delas.

Privacidade Diferencial

- **Definição:** Técnica que introduz ruído estatístico nos dados ou resultados, de modo que a presença ou ausência de um indivíduo seja praticamente indistinguível, preservando utilidade agregada.
- **Exemplo Didático:** É como misturar um pouco de barulho em várias vozes de uma multidão: você entende o que o grupo fala em geral, mas não consegue identificar uma pessoa sozinha.
- **Exemplo Prático:** Fabricantes de celular aplicam privacidade diferencial nos sistemas operacionais para coletar estatísticas de uso sem expor usuários individualmente.
- **Relevância:** Oferece garantias matemáticas formais de privacidade, permitindo treinar ou analisar dados sem revelar registros brutos.
- **Desafio:** A calibragem do ruído pode reduzir a acurácia dos modelos e tornar o treinamento mais custoso em termos de tempo e recursos.

A **privacidade diferencial** é uma técnica que protege contra riscos de reidentificação em grandes conjuntos de dados (RTT, 2024). Ela funciona adicionando ruídos aleatórios controlados às informações, o que reduz significativamente a possibilidade de associar os dados a indivíduos específicos (CIPL, 2025). Dessa forma, ainda é possível extrair padrões e fazer inferências úteis a partir dos dados, mantendo a privacidade das pessoas envolvidas.

Nas entrevistas, a privacidade diferencial foi caracterizada como um método complexo, porém bem estabelecido de anonimização, capaz de fortalecer a proteção em contextos de uso intensivo de dados. Como sintetizou o Entrevistado H:

> *"na privacidade diferencial, a gente não perde a capacidade da IA de aprender com aqueles dados, mas dificultamos a reidentificação daquele indivíduo, porque eu estou manipulando o dado de origem sem que ele perca a sua essência"* ***[Entrevistado H]***

Nesse sentido, a privacidade diferencial tem ganhado destaque em setores que dependem fortemente de dados, como a saúde (Feretzakis et al., 2024). Em um projeto concreto compartilhado pela Entrevistada I, a introdução de ruídos controlados em bases sensíveis, como prontuários médicos, permitiu que pesquisadores e desenvolvedores treinassem modelos de IA capazes de identificar padrões clínicos relevantes sem expor informações individuais dos pacientes.

Apesar de seu potencial elevado de anonimização, os entrevistados também alertaram para o **trade-off** entre proteção da privacidade e utilidade: a técnica pode reduzir a precisão dos modelos de IA e diminuir o valor analítico dos dados, preocupação confirmada por pesquisas recentes (CIPL, 2025).

Especialistas indicaram que a privacidade diferencial **funciona melhor em IA Analítica com dados tabulares** (idade, registros médicos), onde ruídos controlados reduzem riscos de reidentificação sem comprometer significativamente a utilidade dos dados. **Na IA Generativa, a adição de ruído pode acabar sendo interpretada como um sinal estatístico**, fazendo com que o modelo aprenda esse "padrão artificial" em vez dos dados originais, reduzindo a eficácia do treinamento e comprometendo sua utilidade prática.

Ambiente de Execução Confiável

- **Definição:** Cria um ambiente isolado (enclave) dentro do hardware, onde os dados são processados de forma protegida contra acessos não autorizados, inclusive do próprio provedor da infraestrutura.
- **Exemplo Didático:** É uma espécie de "cofre digital" em que os dados podem ser usados para cálculos, mas, lá dentro, ninguém consegue espiar o que acontece.
- **Exemplo Prático:** Serviços de nuvem que permitem processar dados de saúde em TEEs, de modo que nem os engenheiros da empresa de nuvem conseguem acessar as informações brutas.
- **Relevância:** Diferencia-se de técnicas como anonimização porque não altera os dados, mas garante que eles só sejam processados em um ambiente controlado e de alta segurança. É fundamental para setores como saúde ou finanças.
- **Desafio:** A maioria das soluções de TEE é oferecida por grandes empresas estrangeiras de tecnologia, levantando preocupações sobre soberania de dados e dependência tecnológica.

Ambientes de Execução Confiável (TEE) são áreas isoladas dentro de um sistema computacional que permitem processar dados com alto grau de segurança (CIPL, 2025). Diferentemente de outras PETs que atuam diretamente sobre os dados, por meio de técnicas como anonimização ou criptografia, os TEEs protegem o **ambiente** no qual esses dados são processados. Como explicou o Entrevistado B: *"as outras PETs trabalham na camada dos dados [...]. Já o ambiente de execução segura não opera sobre os próprios dados, mas no ambiente em que eles vão ser processados".*

Essa arquitetura é especialmente útil em aplicações que envolvem dados confidenciais. Ao processar informações sensíveis dentro de um TEE, garante-se que elas permaneçam protegidas mesmo em contextos em que o restante do sistema possa estar vulnerável.

Apesar de suas vantagens, os entrevistados também expressaram preocupações com relação à **soberania de dados.** Como a maioria dos TEEs disponíveis hoje pertence a empresas estrangeiras, destacou-se a importância de desenvolver **data centers nacionais protegidos,** capazes de oferecer ambientes de execução confiáveis sem que o próprio provedor da infraestrutura tenha acesso às informações processadas. O Entrevistado D explica que mesmo que uma instância de nuvem esteja localizada no Brasil, se a infraestrutura pertencer a uma empresa dos Estados Unidos, o *Cloud Act* pode permitir o acesso remoto a dados de titulares brasileiros a outras jurisdições (Teofilo, Rocillo, 2018).

Dados Sintéticos

- **Definição:** Dados gerados artificialmente para simular informações reais, usados para ampliar amostras, balancear conjuntos de dados e reduzir o uso de dados pessoais brutos em treinamentos.
- **Exemplo Didático:** O sistema "inventa" fichas fictícias de clientes ou cria fotos falsas, a partir de informações verdadeiras, para que o computador "aprenda" sem precisar acessar informações de pessoas de verdade.
- **Exemplo Prático:** Criar rostos artificiais para treinar um sistema de reconhecimento facial sem usar fotos reais de cidadãos.
- **Relevância:** Ampliam a base de treinamento, aumentam a robustez de modelos e permitem substituir colunas de informações pessoais por versões sintéticas, preservando a estrutura dos dados.
- **Desafio:** Podem reforçar estereótipos ou gerar informações falsas, produzindo modelos menos confiáveis quando usados exclusivamente para treinamento.

Dados sintéticos são informações artificiais criadas por algoritmos que reproduzem as propriedades estatísticas de bases reais, sem copiar registros verdadeiros (Microsoft, 2025). Esta técnica permite treinar e testar modelos de IA sem expor dados pessoais, substituindo-os por versões fictícias, total ou parcialmente. Além de reduzir riscos de reidentificação e mitigar impactos de vazamentos, os dados sintéticos são especialmente valiosos em setores com rígida conformidade regulatória e escassez de dados, como a área da saúde (IBM, 2023).

Nas entrevistas, o Entrevistado A destacou a relevância dessa tecnologia para o treinamento de **modelos de imagens**, seja criando variações de fotos existentes (*data augmentation*) ou modelos 3D (*digital twins*) para simulações, aumentando a robustez dos sistemas. Por outro lado, a Entrevistada I alertou para a queda de precisão em modelos treinados **exclusivamente** com dados sintéticos.

> *"E não adianta só eu gerar imagens sintéticas com IA, porque já foi provado que um modelo treinado com dados sintéticos gerados por outros modelos, a precisão dele cai, absurdamente. É a nossa diversidade, por exemplo, que vai dar riqueza para um sistema de detecção facial, que usa visão computacional" –* ***[Entrevistada I]***

Apesar de seu potencial, é importante reconhecer que a geração de dados sintéticos geralmente depende de bases de dados pessoais para treinar os algoritmos que os produzem. Como observou o Entrevistado D, o uso direto de dados reais, quando protegidos por mecanismos robustos de governança, anonimização e controle de acesso, pode oferecer maior confiabilidade e reduzir o risco de distorções ou "alucinações" nos modelos. Essa visão alinha-se com análises acadêmicas recentes (Giuffrè; Shung, 2023).

Portanto, os dados sintéticos devem ser vistos como instrumentos complementares, não como substitutos completos nas estratégias de proteção e inovação.

Aprendizado Federado

Definição: Técnica em que os dados permanecem nos dispositivos dos usuários (celulares, notebooks), e apenas representações matemáticas são enviadas a um servidor central para atualizar o modelo.

Exemplo Didático: Imagine um time em que cada atleta pratica em sua academia e manda apenas os resultados dos treinos. O técnico usa esses resultados para melhorar a estratégia da equipe, sem nunca ver os treinos completos de cada um.

Exemplo Prático: Um hospital pode treinar um modelo em vários centros médicos diferentes sem precisar receber prontuários de cada um deles. Cada unidade treina localmente e compartilha somente os resultados estatísticos.

Relevância: Reduz a necessidade de centralizar grandes volumes de dados pessoais, aumentando a privacidade e a segurança.

Desafio: A técnica ainda tem uso limitado fora de setores que dependem fortemente de dados (como saúde e tecnologia) e demanda alto poder computacional nos dispositivos de cada ponta.

O aprendizado federado é uma técnica que permite treinar modelos de machine learning colaborativamente, mantendo os dados em suas fontes originais. Em vez de transferir dados de dispositivos periféricos para um servidor central, cada participante utiliza suas próprias informações locais para treinar o modelo (Caballar; Stryker, 2025). Após cada ciclo de treinamento, apenas os parâmetros atualizados, e não os dados brutos, são enviados a um servidor central, que os consolida para aprimorar o modelo global (Caballar; Stryker, 2025).

A tecnologia viabiliza a colaboração entre organizações que não poderiam compartilhar dados sensíveis por questões de privacidade e segurança. O Entrevistado D mencionou como exemplo um projeto de detecção de fraudes no setor de saúde. Neste caso, as empresas precisavam treinar um modelo conjunto, mas se recusavam a compartilhar suas bases de dados e não confiavam em um intermediário para garantir a privacidade. A solução foi processar os dados localmente em cada instituição e, ao final, combinar apenas os parâmetros do treinamento. Este arranjo superou barreiras de confiança e demonstrou como o aprendizado federado pode criar soluções conjuntas mesmo em contextos de alta sensibilidade e pouca disposição para compartilhar informações.

No entanto, devido à necessidade de comunicação frequente entre dispositivos e o servidor central para atualizar os parâmetros do modelo, os entrevistados ressaltaram que o aprendizado federado exige alta capacidade de processamento e conectividade

(CIPL, 2025). Por este motivo, conforme confirmado pelos Entrevistados A e B, sua aplicação prática é mais viável para empresas que têm dados como núcleo de seu negócio, como empresas de tecnologia, e que já possuem a infraestrutura necessária para manter esse fluxo contínuo de informações.

Criptografia Homomórfica

Definição: Técnica que permite realizar cálculos diretamente sobre dados criptografados, sem precisar revelar o conteúdo original.

Exemplo Didático: Imagine fazer contas com números dentro de cofres trancados: você nunca vê os números, mas consegue somar e multiplicar sem abrir o cofre.

Exemplo Prático: Um banco poderia analisar saldos de clientes para prever riscos de crédito sem nunca acessar os valores reais das contas.

Relevância: É considerada uma das soluções mais fortes de proteção, pois mantém os dados criptografados do início ao fim do processamento.

Desafio: É uma solução ainda em desenvolvimento científico. As operações suportadas ainda são básicas, insuficientes para treinar modelos avançados ou realizar cálculos complexos. Além disso, o custo computacional ainda é altíssimo, tornando o uso em larga escala impraticável.

A criptografia homomórfica é considerada uma das áreas mais ativas e desafiadoras em pesquisa de segurança da informação. Sua proposta teórica data dos anos 1970, e a primeira prova de conceito foi proposta em 2009 (Gentry, 2009). Desde então, a literatura científica tem avançado em esquemas mais eficientes, mas o consenso é que a tecnologia ainda se encontra em fase experimental, com aplicações práticas limitadas a cenários de baixa complexidade ou protótipos controlados.

Normalmente, dados são criptografados durante armazenamento ou transmissão. Porém, para operações como atualizações, buscas, análises ou cálculos, geralmente é necessário descriptografá-los primeiro, o que os expõe a possíveis acessos não autorizados. **A criptografia homomórfica oferece uma solução mais segura: permite realizar operações computacionais diretamente nos dados criptografados, sem necessidade de revelá-los durante o processamento** (CIPL, 2025).

Em outras palavras, essa técnica possibilita que sistemas realizem cálculos ou análises com dados protegidos, mantendo sua confidencialidade em todas as etapas. Por isso, a criptografia homomórfica tem sido apontada como uma ferramenta promissora para aplicações em setores que lidam com dados sensíveis, como saúde, finanças e processos eleitorais (Ruiz, 2022).

A **inferência em modelos já treinados** desponta como o uso mais viável até o momento, permitindo que organizações realizem consultas ou previsões a partir de

dados criptografados, sem risco de exposição e mantendo a precisão dos resultados. Entretanto, nas entrevistas, especialistas ressaltaram que a criptografia homomórfica ainda não atingiu **viabilidade técnica plena** e demanda pesquisas adicionais para se tornar aplicável em larga escala. O principal entrave é o elevado custo computacional necessário para sua operação. O Entrevistado B chegou a citar que *"em criptografia homomórfica, em especial, o custo pode chegar a ser mil por cento maior"*.

2.3. PETs e Riscos de Privacidade

Houve consenso entre os entrevistados de que as PETs têm um papel central na mitigação de riscos relacionados ao uso de dados pessoais. Os especialistas apontaram que essas tecnologias reduzem significativamente as chances de reidentificação, abuso ou exposição indevida. Destacou-se que PETs oferecem camadas adicionais de segurança, capazes de tornar o tratamento de dados mais confiável e de transmitir maior segurança a organizações que utilizam soluções de IA.

Ao indicar que essas tecnologias reduzem de maneira relevante os riscos associados ao uso de dados, os especialistas também sugerem que sua adoção pode tornar o tratamento mais proporcional e equilibrado. Em projetos de IA, isso significa **aproximar a prática do ideal de coletar apenas o necessário e empregar salvaguardas que demonstram compromisso com o uso responsável das informações.**

> *"[PETs] são muito, muito importantes porque mitigam mais os riscos e trazem mais segurança para aquela empresa que tem ali a solução de IA com dados pessoais, até porque, sendo a empresa controladora do dado, ela precisa observar esses requisitos legais".* ***[Entrevistada J]***

Além disso, as PETs possibilitam a realização de projetos que seriam inviáveis devido aos altos riscos à privacidade. Exemplos incluem o **uso do aprendizado federado** para treinar modelos de IA com dados de saúde de diferentes instituições e o **processamento seguro de informações pessoais em TEEs**, possibilitando colaboração sem expor indevidamente os dados.

É importante destacar que os especialistas reconheceram que as PETs não eliminam totalmente os riscos - *"nunca existe a eliminação 100% dos riscos, isso é uma primitiva da cibersegurança e da segurança da informação"*, afirmou o Entrevistado B. Ainda assim, a percepção é de que **o potencial dessas tecnologias segue subaproveitado:** embora ofereçam ganhos relevantes em proteção e segurança de dados, muitas soluções ainda carecem de testes amplos e consistentes, o que reduz a confiança do mercado e dificulta novos investimentos.

O **Modelo da Privacidade em Camadas** resulta de pesquisa empírica qualitativa e análise temática, a partir de entrevistas com especialistas, e organiza de forma replicável um framework progressivo de proteção de dados. Ele não parte de pressupostos normativos abstratos, mas emerge da sistematização de práticas relatadas, estruturando camadas que integram ambiente institucional, literacia, governança, segurança da informação e PETs.

3. Análise e Comentários

A pesquisa revelou um conjunto consistente de práticas, percepções e tensões entre os especialistas entrevistados. Embora suas perspectivas variem em aspectos específicos, convergiram em um ponto comum: **PETs desempenham papel relevante na mitigação de riscos, mas seu uso prático no Brasil ainda é fragmentado e pouco sistematizado.**

Essa percepção reforça a necessidade de organizar os diferentes elementos que sustentam a proteção de dados em projetos de IA desde a governança e a segurança da informação até a própria adoção das PETs, de modo a construir um quadro mais claro sobre como essas tecnologias podem ser efetivamente aplicadas no contexto brasileiro.

Esta convergência permitiu sistematizar os achados no **Modelo da Privacidade em Camadas**, que organiza e confere estrutura aos padrões identificados ao longo das entrevistas. Trata-se de um modelo de caráter **exploratório**, que não deriva de pressupostos normativos estabelecidos a priori, mas emerge diretamente da análise temática do material empírico coletado.

> **Seu valor reside na capacidade de organizar elementos já presentes na prática profissional, oferecendo uma representação estruturada que pode orientar tanto empresas quanto formuladores de políticas públicas.**

Na literatura acadêmica, **modelos são compreendidos como representações simplificadas e funcionais da realidade** (Bhattacherjee, 2025; Martins, 2005). Diferentemente das teorias, modelos não buscam explicações completas dos fenômenos, mas capturam relações específicas entre variáveis para guiar análises e previsões.

Formuladores de políticas públicas também utilizam frequentemente modelos de análise para estruturar processos de avaliação, e é nesta perspectiva que se insere a proposta aqui desenvolvida: **uma ferramenta prática para organizar um campo ainda caracterizado pela dispersão conceitual e metodológica.**

Fonte: elaboração própria

O Modelo da Privacidade em Camadas é um framework conceitual voltado a orientar a adoção de PETs em projetos de IA. Ele organiza a proteção de dados em cinco níveis interligados - ambiente institucional, literacia, governança, segurança da informação e as próprias PETs. A lógica é progressiva: cada camada cria as condições para que as tecnologias de privacidade sejam aplicadas de forma consistente, reduzindo riscos e fortalecendo práticas responsáveis.

O **Modelo da Privacidade em Camadas** resulta da própria pesquisa empírica qualitativa desse estudo: suas camadas emergem das entrevistas, interdependentes entre si:

- **Ambiente institucional:** compreende o contexto regulatório e as pressões externas que moldam decisões sobre uso de dados pessoais;

> *"Quando analisamos projetos em organizações que atuam **em setores altamente regulados, com maior preocupação em relação à privacidade**, é mais comum observar o emprego desse tipo de tecnologia de proteção." **[Entrevistado E]***

- **Literacia:** abarca o nível de conhecimento técnico e organizacional, frequentemente identificado pelos entrevistados como essencial para adoção plena de técnicas avançadas de proteção;

> *"O **maior desafio está relacionado à ética, à governança e à educação** sobre o uso dessas tecnologias: até que ponto elas são realmente úteis e até que ponto ainda é necessário o discernimento humano? **Para lidar com isso, não basta dispor de ferramentas; é preciso treinar, educar e capacitar as pessoas.**" **[Entrevistada I]***

- **Governança:** refere-se aos processos internos de coordenação e alinhamento estratégico, necessário para evitar implementações tecnológicas fragmentadas ou desarticuladas;

> *"A **governança** envolve garantir que, se houver dados como CPF, estes estarão anonimizados em algum hash, e só poderão ser desencriptados após um pedido que passe por um time de legal, time de compliance, algum tipo de governança" **[Entrevistado A]***

- **Segurança da informação:** constitui base transversal, uma vez que a proteção de dados pessoais dificilmente se sustenta sem práticas robustas de cibersegurança;

> *"Existe uma preocupação com a infraestrutura desses ambientes, vinculada à **segurança da informação tradicional**, que permanece relevante para prevenir ataques de agentes mal-intencionados." **[Entrevistada G]***

- **PETs** compõem o núcleo tecnológico propriamente dito, oferecendo instrumentos concretos de mitigação de riscos à privacidade. Sem as bases anteriores, a eficácia das PETs dificilmente será aproveitada.

> *"As* **PETs mais elaboradas mitigam mais o risco do que as tecnologias mais tradicionais** *utilizadas na indústria"* ***[Entrevistado B]***

Embora exploratório, o modelo permite a estruturação de um **framework prático** para empresas que desenvolvem e utilizam soluções de IA. Sua estrutura em camadas progressivas facilita visualizar que organizações podem estar em diferentes níveis de maturidade, com avanços obtidos através do fortalecimento articulado dessas dimensões.

Na prática, o modelo pode ser utilizado para:

Autoavaliação organizacional: mapear em qual camada a empresa tem maior solidez (ex.: políticas de segurança da informação robustas) e onde há fragilidades (ex.: ausência de capacitação em PETs).

Definição de prioridades de investimento: alinhar recursos a áreas críticas, como criar programas de literacia em privacidade antes de adotar tecnologias avançadas.

Planejamento regulatório e institucional: avaliar se o ambiente normativo e os contratos permitem uso seguro de PETs (ex.: cláusulas de compartilhamento de dados).

Implementação progressiva de PETs: começar a implementação de PETs por soluções mais acessíveis (ex.: anonimização e pseudonimização) e avançar para técnicas sofisticadas (ex.: criptografia homomórfica, aprendizado federado).

Integração com governança corporativa: incorporar métricas de maturidade em privacidade em relatórios de compliance ou de auditoria de risco.

4. Conclusão

O avanço da IA e o uso intensivo de dados apresentam novos desafios para a proteção da privacidade dos titulares. Nesse cenário, as **PETs emergem como ferramentas essenciais para mitigar riscos**, oferecendo soluções técnicas que reduzem a exposição de dados pessoais em diversas etapas do ciclo de vida das informações.

Os achados deste estudo revelam, no entanto, que **essas tecnologias não são suficientes quando aplicadas isoladamente**. A implementação efetiva de PETs deve integrar uma **perspectiva mais ampla de governança de dados**, que inclui capacitação contínua, conscientização organizacional e letramento em privacidade. Sem esses elementos complementares, mesmo as soluções mais sofisticadas permanecem vulneráveis.

É importante ressaltar que este estudo não conduziu análise técnica ou operacional da eficácia individual das PETs. Nosso objetivo foi **destacar essas tecnologias** ao mapear perspectivas de especialistas e evidenciar seu papel como instrumentos de mitigação de riscos à privacidade no contexto brasileiro. Ao introduzir o tema no debate público, buscamos também estimular maior **engajamento de empresas, reguladores e sociedade,** fomentando investimentos e contribuindo para o amadurecimento dessas soluções no país.

Por fim, a proteção da privacidade em sistemas de IA requer mais que tecnologias inovadoras – demanda um **ambiente institucional robusto** fundamentado em governança eficaz, normas claras, boas práticas e investimento em capacitação. Somente esta combinação permitirá que as PETs realizem seu potencial de fortalecer a confiança no uso responsável de dados e consolidar um ecossistema regulatório preparado para os desafios da IA.

5. Sugestões para estudos futuros

Este estudo apresentou o tema das PETs e analisou seu potencial de aplicação na proteção de dados em sistemas de IA. Ainda assim, várias questões permanecem em aberto. A seguir, destacamos lacunas que podem orientar pesquisas futuras e contribuir para a continuidade do debate regulatório, ampliando o conhecimento sobre essas tecnologias e seus impactos no contexto brasileiro.

Padronização de linguagem e efeitos na adoção: A ausência de terminologia uniforme para PETs no Brasil foi recorrente nas entrevistas. Estudos futuros podem investigar como entidades reguladoras, associações setoriais e organismos de padronização podem difundir nomenclaturas claras e incentivar a adoção dessas tecnologias.

Aplicação prática do modelo: Há lacunas sobre o funcionamento e a operação técnica de PETs em cenários reais. Pesquisas futuras, desenvolvidas com o apoio de especialistas em STEM, podem mapear casos de uso no Brasil, testando custos, barreiras técnicas e impactos regulatórios decorrentes da aplicação dessas tecnologias.

Economia política da adoção: O custo elevado das PETs e a concentração de expertise em grandes empresas internacionais criam assimetrias. Novos estudos podem explorar incentivos, mecanismos de financiamento e o papel do Estado para democratizar sua adoção.

Efeitos para o debate sobre legítimo interesse: As PETs podem influenciar a interpretação jurídica do legítimo interesse como base legal para tratamento de dados. Pesquisas podem examinar se o uso dessas tecnologias fortalece argumentos de proporcionalidade e necessidade em contextos regulatórios brasileiros.

Referências

CABALLAR, Rina Diane; STRYKER, Cole. **O que é aprendizado federado?** 2025. Disponível em: https://www.ibm.com/br-pt/think/topics/federated-learning. Acesso em: 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. Disponível em: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_pets_and_ppts_in_ai_mar25.pdf. Acesso em: 10 set. 2025.

EDIN, Per et al. **Quantifying the GenAI opportunity: Lessons learned from benchmarking 17 million+ companies worldwide.** 2025. Disponível em: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/quantifying-genai-opportunity.pdf. Acesso em: 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. Disponível em: https://www.mdpi.com/2078-2489/15/11/697. Acesso em: 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. Disponível em: https://dl.acm.org/doi/10.1145/1536414.1536440. Acesso em: 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. Disponível em: https://www.nature.com/articles/s41746-023-00927-3#citeas. Acesso em: 11 set. 2025.

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

IBM. **What is synthetic data?** Disponível em: https://www.ibm.com/think/topics/synthetic-data. Acesso em: 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. Disponível em: https://bristoluniversitypressdigital.com/view/journals/pp/44/2/article-p171.xml. Acesso em: 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. Disponível em: https://arxiv.org/abs/2202.06539. Acesso em: 19 set. 2025.

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

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

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

RADAR DE TENDêNCIAS TECNOLÓGICAS. **Tecnologias de Aprimoramento de Privacidade**. 2024. Disponível em: https://radar.apps.bb.com.br/tendencia/Radar-2024/Temas/TechGuardian/Tecnologias-de-Aprimoramento-de-Privacidade/21001. Acesso em: 23 jul. 2025.

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

STRYKER, Cole. **What are LLMs?** 2025. Disponível em: https://www.ibm.com/think/topics/large-language-models. Acesso em: 19 set. 2025.

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

TEOFILO, Davi; ROCILLO, Paloma. **CLOUD Act: um caso de Direitos Humanos e Jurisdição**. Disponível em: https://irisbh.com.br/cloud-act-um-caso-de-direitos-humanos-e-jurisdicao/. Acesso em: 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.

Anexo de Metodologia Reglab

Título	**Privacidade em Camadas:** O Papel de Privacy Enhancing Technologies (PETs) em Sistemas de Inteligência Artificial
Pergunta de pesquisa	Como especialistas em IA, privacidade e cibersegurança avaliam o uso de Privacy Enhancing Technologies (PETs) em sistemas de inteligência artificial e sua efetividade na proteção de dados pessoais?
Resumo de metodologia	Esta pesquisa adota uma abordagem qualitativa e exploratória, combinando coleta de dados primários por meio de entrevistas qualitativas semiestruturadas com especialistas (*expert interviews*), complementada por coleta de dados secundários (documentos, literatura e casos práticos). A análise dos dados seguiu a técnica de análise temática reflexiva com auxílio do software Atlas.ti. Ferramentas visuais do software foram utilizadas para identificar padrões e temas centrais, posteriormente validados contra o corpus empírico.
Coleta de dados	A pesquisa utilizou a metodologia de expert interviews (Audenhove e Donders, 2019), realizando entrevistas qualitativas semiestruturadas de caráter exploratório. Este método foi escolhido devido ao caráter técnico do tema e à falta de padronização conceitual sobre o assunto, o que torna essencial o conhecimento de especialistas atuantes na área no Brasil. A amostra foi composta seguindo critérios de diversidade e representatividade, incluindo: participação mínima de mulheres; representantes da academia ou centros de pesquisa; profissionais de empresas brasileiras; e especialistas de grandes empresas e consultores em tecnologia. A seleção dos participantes combinava busca ativa no LinkedIn como estratégia principal, complementada por amostragem por conveniência e técnica de snowballing para expandir a rede de especialistas em IA, privacidade e cibersegurança. Das 37 pessoas contatadas, 12 aceitaram participar da pesquisa, enquanto 7 informaram indisponibilidade e 18 não responderam ao convite. As entrevistas ocorreram entre 29 de julho e 29 de agosto de 2025, em formato online (via Teams), durando entre 45 e 60 minutos. Todas foram conduzidas por ao menos uma pesquisadora do RegLab, seguindo um roteiro de perguntas anexo ao estudo. Uma entrevista foi realizada como piloto para testar o roteiro e validar hipóteses iniciais. As 11 entrevistas restantes foram consideradas suficientes para saturação teórica, pois em abordagens qualitativas com entrevistas semiestruturadas e em profundidade, a recorrência temática e a densidade analítica geralmente se consolidam com poucos participantes (Guest et al, 2006). As entrevistas foram gravadas com autorização dos participantes, transcritas integralmente e acompanhadas por memorandos dos entrevistadores. O material foi armazenado e codificado no software Atlas.ti, com nomes e instituições dos entrevistados devidamente anonimizados.
Análise de dados	A análise dos dados seguiu a técnica de análise temática reflexiva (Braun; Clarke, 2006), adequada a estudos qualitativos exploratórios em contextos de alta complexidade. Essa abordagem prioriza a interpretação contextualizada em vez de uma codificação exaustiva, permitindo o uso de diferentes estratégias analíticas abertas. Todas as entrevistas foram transcritas integralmente e processadas no software Atlas.ti, que realizou uma primeira rodada de codificação automatizada intencional. Esse procedimento gerou 719 códigos de primeiro nível e 23 códigos de segundo nível, que foram posteriormente revisados manualmente pela equipe de pesquisa. Na etapa seguinte, diferentes ferramentas visuais do software (nuvens de conceitos, mapas, gráficos de correlação, chatbots) foram utilizadas para identificar padrões e relações entre os códigos. Esse processo resultou na emergência de temas centrais, os quais foram testados e validados contra o corpus empírico, assegurando sua aderência aos dados originais. A análise foi conduzida entre os dias 29 de agosto e 5 de setembro de 2025.

Procedimentos de redução de vieses	**Referências teórico-metodológicas consolidadas:** as técnicas de coleta e análise de dados adotadas neste estudo seguiram práticas reconhecidas na literatura acadêmica. A abordagem metodológica foi discutida internamente antes e após a realização das entrevistas preliminares, permitindo a incorporação de críticas e sugestões ao desenho final da pesquisa, antes do início do processo de análise. **Ferramenta complementar de checagem:** tendo em vista que a codificação inicial foi realizada por meio de software, empregamos um segundo software (NotebookLM) para verificar a consistência das codificações produzidas no Atlas.ti e identificar pontos cegos. O uso esse software foi feito por pesquisadoras que participaram diretamente das entrevistas, com o objetivo de capturar nuances que podem ter sido ignorados na codificação automatizada. **Triangulação de métodos:** os achados empíricos foram contrastados com análise documental de fontes secundárias, com o objetivo de comparar, validar e reforçar a consistência das interpretações construídas a partir das entrevistas. Essas referências, quando utilizadas, foram expressamente citadas ao longo do texto **Dupla análise independente:** dois pesquisadores revisaram o conjunto de códigos e temas de forma cruzada, reduzindo vieses individuais. A definição final dos temas foi realizada em discussão coletiva entre os dois autores, assegurando múltiplas perspectivas e controle de vieses individuais na interpretação dos dados. **Registro e transparência metodológica:** todas as etapas do processo analítico foram documentadas, incluindo versões sucessivas dos arquivos e decisões de codificação. Essa prática permite a rastreabilidade do percurso metodológico, conforme as diretrizes do Reglab para transparência e replicabilidade
Outras Limitações Metodológicas	**Codificação automatizada inicial:** embora o Atlas.ti seja um dos softwares mais consolidados para análise qualitativa e sua codificação tenha sido validada com o uso de um segundo software, o uso de codificação automatizada pode ter gerado ruídos na etapa inicial, que podem constituir pontos-cego na análise posterior. **Dependência de ferramentas externas:** parte do processo analítico dependeu do desempenho de softwares proprietários, o que pode limitar a replicabilidade em contextos diferentes. **Escopo qualitativo:** os achados refletem percepções de um conjunto delimitado de entrevistas, com profundidade analítica, mas sem pretensão de generalização estatística. **Amostragem por conveniência:** a seleção pode ter refletido vieses de disponibilidade e círculos profissionais, apesar dos critérios de diversidade. **Evolução tecnológica:** os resultados refletem o estado da arte das ferramentas e práticas de IA no momento da pesquisa. Mudanças rápidas nesse campo podem alterar parte das conclusões.
Uso de software	

SOFTWARE	USO NA PESQUISA
Suíte MS Office	Edição de texto, planilhas e gráficos, entrevistas (Teams)
Suíte Adobe C	Diagramação e finalização de gráficos e ilustrações
Atlas.ti	Organização, codificação e análise dos dados qualitativos
Cockatoo	Transcrição de áudio das entrevistas em texto
ChatGPT 5o	Brainstorm, sistematização de informações, revisão gramatical (ortografia, gramática busca de sinônimos), adequação da linguagem, adequação ao Manual de Redação Reglab
Notion AI	Edição e revisão de texto (ortografia e gramática, busca de sinônimos, adequação de linguagem, traduções), organização da pesquisa e estruturação de cronograma
Lex.page	Revisão de texto (brevidade, clichês, legibilidade, voz passiva, afirmações sem evidências, repetições)
More UFSC	Geração de referências bibliográficas no modelo ABNT

Diretrizes éticas	**Financiamento da pesquisa:** esta publicação integra uma série de pesquisas patrocinadas pelas empresas Google, Meta e b/luz, com o Reglab mantendo controle editorial integral. Diferentemente de pesquisas comissionadas, o Reglab definiu o escopo, os objetivos e a metodologia deste estudo com completa autonomia. Os autores preservaram total independência profissional e assumem integral responsabilidade pelo conteúdo e pelas conclusões apresentadas. **Tratamento de dados pessoais:** a pesquisa envolveu o tratamento de dados pessoais apenas nas etapas de coleta e análise, de forma limitada e proporcional aos objetivos do estudo, em conformidade com a Lei nº 13.709/2018 (LGPD). **Base legal:** todos os participantes autorizaram formalmente sua participação mediante assinatura de termo de consentimento, com ciência sobre os objetivos da pesquisa e sobre o uso dos dados. **Finalidade e adequação:** os dados foram utilizados exclusivamente para os fins desta pesquisa, de acordo com o consentimento obtido, não sendo empregados para outras finalidades. **Minimização e anonimização:** informações pessoalmente identificáveis que não eram relevantes para os objetivos do estudo foram anonimizadas nas transcrições e excluídas da base ativa. **Sigilo e confidencialidade:** na apresentação dos resultados, os dados foram mantidos sob sigilo e as citações foram ajustadas, quando necessário, para preservar a confidencialidade das fontes. Apenas um número restrito de pesquisadores diretamente envolvidos no projeto teve acesso aos dados pessoais e documentos originais. **Registro e segurança da informação:** os arquivos foram armazenados mediante controle de acesso por senha e em conformidade com as políticas internas de segurança da informação do Reglab. **Retenção e descarte:** os dados serão armazenados por até 12 meses, exclusivamente para fins de auditoria metodológica e eventual replicação, sendo posteriormente eliminados. **Uso responsável de dados públicos:** embora alguns dados analisados sejam públicos, seu uso foi realizado de maneira responsável e ética, com o objetivo exclusivo de pesquisa independente. **Transparência metodológica:** a metodologia de pesquisa foi descrita de forma detalhada para assegurar transparência e replicabilidade, contribuindo para a integridade científica e possibilitando a validação independente dos resultados. **Não Discriminação e Respeito à Diversidade:** a pesquisa foi conduzida de modo a respeitar a diversidade e a evitar qualquer forma de discriminação.

ANEXO II - ROTEIRO SEMIESTRUTURADO DAS ENTREVISTAS

#	PERGUNTA
1	Para começarmos, você poderia contar um pouco sobre sua experiência com projetos de IA? E, mais especificamente, sua experiência com questões de privacidade ou proteção de dados no contexto de IA?
2	Na sua visão, como dados pessoais são realmente usados nos modelos de IA? Do ponto de vista técnico, eles perdem seu caráter identificável ao longo do processo?
3	Quais são, na sua experiência, as principais técnicas usadas hoje para reduzir riscos à privacidade e melhorar a cibersegurança de dados pessoais em projetos que envolvem IA?
4	Vou citar algumas tecnologias. Você conhece ou já trabalhou com alguma delas? Poderia comentar, se quiser, sobre a relevância que elas têm na prática? • Privacidade diferencial • Ambiente de execução confiável (TEE) • Dados sintéticos • Aprendizado federado • Criptografia homomórfica
5	Na sua visão, o uso de PETs nos sistemas de IA realmente elimina riscos à privacidade, ou ainda restam preocupações relevantes, como reidentificação ou vazamento?
6	No seu dia a dia, o termo “PETs” costuma ser usado? Ou é mais comum se referirem diretamente às tecnologias específicas?
7	Você já viu casos em que o uso de PETs permitiu viabilizar projetos que, inicialmente, seriam inviáveis ou de alto risco por questões de privacidade?
8	Você já se deparou com projetos que envolviam dados sensíveis — como saúde, origem racial ou dados de crianças — em que o uso de PETs foi considerado (ou adotado) para viabilizar a coleta ou reduzir riscos? Como essa decisão foi tratada?
9	Quais você considera os principais desafios para aumentar a disseminação e adoção de PETs em IA? Você já enfrentou algum desses obstáculos em algum projeto?
10	Gostaria de destacar mais algum ponto que não foi abordado ou deixar uma recomendação para futuras pesquisas nessa área? Indica alguém para participar também das entrevistas?