‹ radar tecnológico ›

inteligência artificial generativa

ANPD

Autoridade Nacional de Proteção de Dados

‹ radar tecnológico ›

n° 3

inteligência artificial generativa

Albert França Josuá Costa
Fabiana Faraco Cebrian
Lucas Costa dos Anjos
Marcelo Santiago Guedes
Thiago Guimarães Moraes

ANPD
Brasília, DF
2024

ANPD
Autoridade Nacional de Proteção de Dados

Diretor-Presidente
Waldemar Gonçalves Ortunho Junior

Diretores
Arthur Pereira Sabbat
Miriam Wimmer

Equipe de elaboração
Coordenação-Geral de Tecnologia e Pesquisa (CGTP)
Fabiana S. P. Faraco Cebrian
Albert França Josuá Costa
Lucas Costa dos Anjos
Marcelo Santiago Guedes
Thiago Guimarães Moraes

Projeto gráfico / editoração eletrônica
André Scofano Maia Porto

1ª edição
Publicação digital – PDF
Radar Tecnológico, Número 3, NOV 2024

ANPD
SCN, Qd. 6, Conj. A,
Ed. Venâncio 3000, Bl. A, 9º andar
Brasília, DF · Brasil · 70716-900
t. (61) 2025-8101
www.gov.br/anpd

‹ sobre a série ›

A série "Radar Tecnológico" é uma produção periódica da ANPD que objetiva realizar abordagens concisas de tecnologias emergentes que vão impactar ou já estejam impactando o cenário nacional e internacional da proteção de dados.

Sem a intenção de esgotar as temáticas ou firmar posicionamentos institucionais, o propósito da série é agregar informações relevantes ao debate da proteção de dados no País, com textos estruturados de forma didática e acessível ao público em geral.

Para cada tema, são abordados os conceitos principais, as potencialidades e as perspectivas de futuro, sempre com ênfase na proteção de dados no contexto brasileiro. ▪

nota explicativa

O estudo sobre Inteligência Artificial Generativa foi originalmente desenvolvido pela Coordenação-Geral de Tecnologia e Pesquisa como um insumo interno para subsidiar as unidades da Autoridade Nacional de Proteção de Dados (ANPD) no ano de 2023. Seu propósito inicial foi apoiar reflexões e discussões técnicas dentro do contexto institucional.

Destacamos que, desde a sua elaboração, o cenário tecnológico relacionado à Inteligência Artificial Generativa tem evoluído rapidamente. Por essa razão, esta edição do Radar Tecnológico pode conter informações que, embora relevantes no momento de sua produção, não refletem plenamente as atualizações mais recentes no campo como a seção sobre perspectivas de futuro.

Acreditamos que a divulgação desta edição mantém a sua relevância, pois oferece uma base sólida e um ponto de partida para reflexões e desafios relacionados com a Inteligência Artificial Generativa e a proteção de dados pessoais. ▪

‹ sumário ›

07
introdução

09
conceitos principais

16
inteligência artificial generativa e proteção de dados pessoais

27
inteligência artificial generativa no contexto brasileiro

29
perspectivas de futuro

32
considerações finais

34
referências

‹ introdução ›

A criação de máquinas inteligentes é uma busca já conhecida da humanidade, remetendo ao século XVIII, quando a máquina denominada de O Turco impressionava o mundo ao jogar xadrez, supostamente, provida de inteligência. O termo *Inteligência Artificial*[1] (IA) foi inicialmente cunhado somente na década de 1950, durante a conferência Dartmouth (DARTMOUTH, 2023). Desde então, a IA promoveu grandes avanços na ciência, bem como passou por mudanças internas notáveis.

No início, a IA foi baseada na abordagem simbólica, na qual um conjunto de regras é identificado por especialistas humanos e, posteriormente, traduzido em instruções lógicas para serem reproduzidas por máquinas. No entanto, essa abordagem apresenta limitações significativas, pois o mundo real possui uma alta complexidade que impossibilita a codificação manual do conhecimento em larga escala (RUSSEL E NORVIG, 2013).

Essas limitações levaram a comunidade científica a buscar mecanismos que permitissem as máquinas, de forma autônoma, aprenderem as regras que resolvem os problemas mais complexos com base em dados históricos dos próprios problemas, ao invés de programá-las explicitamente. Essa busca promoveu o surgimento do campo denominado de *Aprendizado de Máquina*[2] (AM).

Como explica Solove (2024), a maioria das vezes que a expressão IA é atualmente utilizada, tem por objetivo se referir a *sistemas de IA*[3] baseados em AM. É importante ter em mente que a IA é um campo muito mais largo do que o AM, mas para fins desse estudo, as expressões serão utilizadas de forma intercambiáveis.

No começo do século XXI, a área de AM passou por um crescimento exponencial, levando ao surgimento da subárea que ficou conhecida por *Aprendizado de Máquina Profundo*[4] (AMP). Os modelos de AMP são amplamente utilizados hoje em aplicações de IA. Mais recentemente, o surgimento e a popularização dos *modelos generativos*[5] têm ganhado notoriedade. Esses modelos fazem parte de uma subárea de AMP e

1 ***Inteligência Artificial*** *é a área do conhecimento humano que estuda o desenvolvimento de sistemas de inteligência artificial.*

2 ***Aprendizado de Máquina*** *é a capacidade de melhorar o desempenho [de uma máquina] na realização de alguma tarefa por meio da experiência (MITCHELL, 1997).*

3 ***Sistema de IA*** *é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir das entradas que recebe, como gerar saídas tais como predição, conteúdo, recomendações ou decisões que podem influenciar ambientes físico ou virtual. Diferentes sistemas de IA variam os seus níveis de autonomia e adaptabilidade após a sua implantação (OCDE, 2024).*

4 ***Aprendizado de Máquina Profundo*** *é uma variante do Aprendizado de Máquina que utiliza múltiplas camadas para resolver problemas extraindo o conhecimento a partir de dados brutos, e transformando-os em cada camada. Essas camadas incrementalmente obtêm características de alto nível dos dados brutos, permitindo a solução de problemas mais complexos com ›*

possuem enorme potencial para uso em diversos campos, tais como a formulação de novos produtos farmacológicos (TONG *et al*, 2021).

Este estudo técnico tem como objetivo fundamentar os conceitos relacionados aos modelos generativos de Inteligência Artificial, identificar potenciais riscos à privacidade e proteção de dados e relacioná-los, preliminarmente, à Lei Geral de Proteção de Dados Pessoais (LGPD).

› *uma alta acurácia e uma menor dependência de ajustes manuais (GARTNER, 2024).*

5 ***Modelos generativos*** *descrevem como um conjunto de dados é gerado em termos de um modelo probabilístico. Esse modelo é capaz de gerar novos dados por meio de amostragem (FOSTER, 2019).*

‹ conceitos principais ›

Dentre as diversas definições de AM, Mitchell (1997) o define como sendo a capacidade de melhorar o desempenho [de uma máquina] na realização de alguma tarefa por meio da experiência. O surgimento do AM permitiu que problemas mais complexos pudessem ser resolvidos por máquinas de forma automática, tais como a detecção de fraudes em operações de crédito, a detecção e a classificação de tumores em exames de imagens, a caracterização de elementos de discurso de ódio, entre outras tarefas.

Os avanços na área de AM trouxeram à luz novos termos, tais como modelo de linguagem, modelo generativo e modelo fundacional, sendo muitas vezes utilizados de forma intercambiável[6]. Entretanto, por mais que sejam conceitos relacionados, esses termos possuem significados distintos que devem ser diferenciados corretamente.

Os modelos de *Redes Neurais Artificiais*[7] (HAYKIN, 1999) são os mais conhecidos dentro da área de AM, porém há diversos outros modelos, como K-Vizinhos Mais Próximos (FIX e HODGES, 1951); Árvore de Decisão (QUINLAN, 1979) e Máquinas de Vetores de Suporte (CRISTIANINI e SHAWE-TAYLOR, 2000).

Independentemente do modelo de AM, sua geração ocorre, majoritariamente, em dois passos: *treino*[8] e *teste*[9] (HAYKIN, 1999).

Geração dos Modelos de Aprendizado de Máquina

A principal premissa existente para a construção de modelos de aprendizado de máquina é a existência de dados históricos que suficientemente representem o problema que será resolvido com o uso de IA. De forma simplificada, a premissa é a seguinte: o que falta de conhecimento de como modelar problemas complexos, deve sobrar em dados.

Em uma abordagem clássica, esses dados são divididos nos conjuntos de treino e teste. O primeiro é utilizado para treinar o modelo. Por sua vez,

6 *A distinção entre o conceito de Modelo de linguagem e Modelo generativo será apresentada no item 2.1.*

7 ***Rede Neural Artificial*** *é um modelo de aprendizado de máquina bioinspirado na forma em que o cérebro processa dados (HAYKIN, 2019).*

8 ***Treino:*** *o modelo de aprendizado de máquina aprende o padrão dos dados extraído de um conjunto de dados.*

9 ***Teste:*** *o desempenho do modelo é mensurado em um conjunto de dados diferente do utilizados para o treino do modelo.*

o segundo é utilizado para mensurar o desempenho do modelo. Essa abordagem é denominada de Treino e Teste (Figura 1).

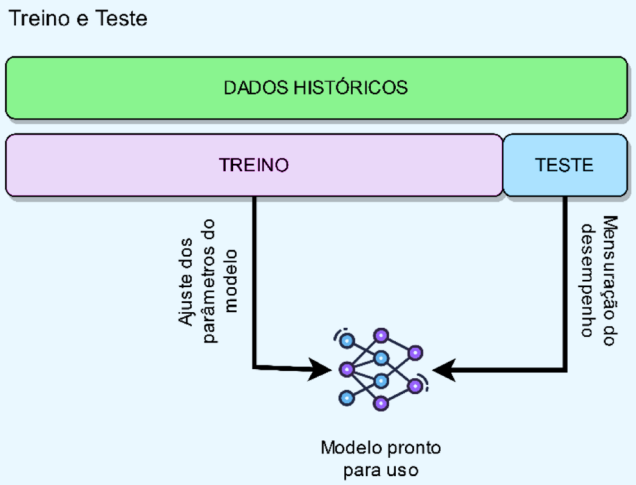

Figura 1 *Geração do modelo utilizando abordagem Treino e Teste.*

Fonte: ANPD

Antes de aprofundar nos modelos de IA Generativa, é importante estudar os modelos do tipo discriminativo, que são comumente utilizados com a finalidade de classificar grupos e/ou prever tendências ou comportamentos.

Modelo Discriminativo

No aspecto geral, os modelos de AM são do tipo discriminativo. Esse tipo tem como objetivo aprender, diretamente a partir dos dados, a relação entre as características dos dados e o que eles representam. Em síntese, os modelos discriminativos aprendem uma função que toma uma decisão sobre os dados.

Essa função de decisão é crucial para as tarefas realizadas pelos modelos discriminativos, como tarefa de classificação que consiste na categorização dos dados em classes previamente conhecidas, e a tarefa de regressão que em predizer o valor numérico de uma variável contínua de acordo com os valores dos dados.

O sucesso na tarefa de classificação depende de diversos fatores, desde a qualidade dos dados até a escolha criteriosa do algoritmo e seus parâme-

tros. Técnicas como pré-processamento e seleção de recursos podem ser essenciais para aprimorar o desempenho do modelo, garantindo que ele capture as informações relevantes e ignore ruídos irrelevantes.

Por sua vez, o domínio do problema de regressão abre um leque de oportunidades em diversos campos. Na área de finanças, por exemplo, permite prever tendências do mercado e avaliar riscos de investimentos. Já na área de saúde, contribui para o diagnóstico de doenças, personalização de tratamentos e otimização da gestão de recursos hospitalares. Nas indústrias, auxilia na otimização de processos, previsão de demanda e controle de qualidade.

Modelo Generativo

Em 2014, Ian Goodfellow publicou o trabalho percursor *Generative Adversarial Networks* (GOODFELLOW *et al*, 2014) que se tornou o ponto de partida para o desenvolvimento de um novo tipo de modelo de AM, denominado de generativo. Os modelos generativos, em vez de apenas aprender a função de associação dos modelos discriminativos, tentam capturar as características estatísticas subjacentes dos dados para gerar novos exemplos que se assemelham aos dados reais.

Um modelo generativo de AM tem como objetivo gerar *dados sintéticos*[10] com as mesmas características estatísticas dos dados reais. Idealmente, os dados sintéticos gerados devem ser tão próximos dos dados reais, que a distinção entres eles não deve ser possível.

Atualmente, há duas grandes abordagens para os modelos generativos: a primeira é baseada em *redes adversariais generativas*[11] (GAN) e a segunda em *transformadores generativos pré-treinados* (GPT)[12] [13] (CRS, 2023).

Para alcançar o objetivo, os modelos baseados em redes adversariais generativas são compostos por dois componentes principais, denominados de Gerador e Discriminador. O Gerador é responsável por tentar enganar o Discriminador ao gerar dados sintéticos com características estatísticas semelhantes aos dados reais. Por sua vez, o Discriminador tenta identificar quais dados são reais e quais não são. O modelo gene-

10 ***Dados sintéticos*** *são dados gerados artificialmente, em contraste com os dados reais que são oriundos da realidade (AEPD, 2023). O termo "dado sintético" é utilizado tanto no campo da Tecnologia da Informação (TI) quanto na área de privacidade e proteção de dados. Nesta última, o uso desses dados costuma estar associada a* privacy enhancing technolgies *(PETs). Deste modo, no capítulo 3, que discute a relação da IA generativa com a proteção de dados, será utilizada a expressão "conteúdo sintético" para se referir aos dados sintéticos gerados por modelos generativos. O conteúdo sintético não tem o objetivo de anonimizar dados pessoais.*

11 ***Redes Adversariais Generativas*** *são modelos de aprendizado de máquina que utilizam a abordagem adversarial para a geração de conteúdo sintético.*

12 *Do inglês* Generative Pre-Trained Transformer *(GPT). Ainda não há uma tradução formalmente estabelecida do termo para o português. Por isso, optou-se por fazer uma tradução própria.*

13 ***Transformadores generativos pré-treinados*** *são modelos de aprendizado de máquina que utilizam a abordagem codificador-decodificador para a geração de conteúdo sintético.*

rativo torna-se adequado quando o Discriminador não mais conseguir distinguir entre dados reais de dados sintéticos (Figura 2).

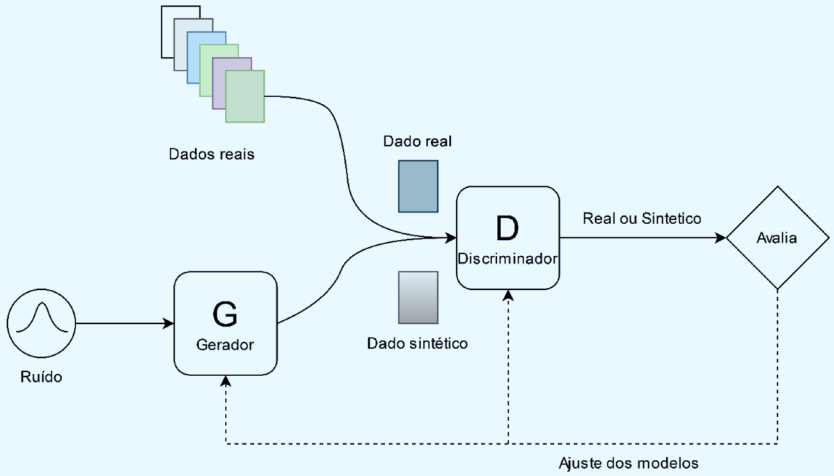

Figura 2 *Estrutura de um modelo generativo baseado em redes adversarias generativas.*

Fonte: ANPD

A abordagem baseada em transformadores generativos pré-treinados surgiu com o desenvolvimento de duas técnicas: (i) a arquitetura de Transformadores (FOSTER, 2019) e (ii) o mecanismo de Atenção (VASWANI *et al*, 2017).

Para alcançar o objetivo de geração de dados, a arquitetura de transformadores é composta por camadas de codificadores e de decodificadores. As camadas codificantes recebem um dado como entrada e a transformam em uma representação numérica chamada de *espaço de representação*[14], que representa o significado da entrada. Os decodificadores, por sua vez, geram a saída com base nas informações contidas no espaço de representações e em um contexto anterior (Figura 3).

14 ***Espaço de representação****: do inglês* embedding. *Ainda não há uma tradução formalmente estabelecida do termo para o português. Por isso, optou-se por fazer uma tradução própria.*

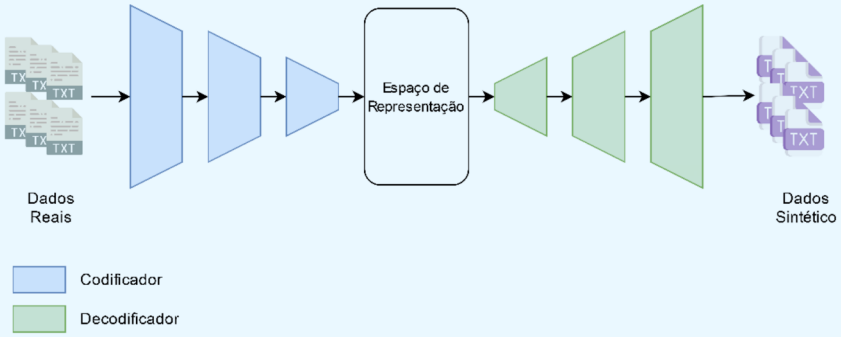

Figura 3 *Estrutura de um modelo generativo baseado em transformadores generativos pré-treinados.*

Fonte: ANPD

Por sua vez, o mecanismo de atenção permite que o modelo generativo se concentre em partes específicas do dado real, em vez de o tratar de maneira uniforme. Esse foco seletivo é útil quando as partes importantes para a execução de uma tarefa estão distribuídas por toda a entrada. O mecanismo atribui de forma dinâmica pesos distintos às partes do texto com base na sua importância para a execução da tarefa permitindo que as partes mais importantes sejam priorizadas em detrimento das partes de menor importância (VASWANI *et al*, 2017).

Dois exemplos são fornecidos sobre o conceito de mecanismo de atenção. O primeiro utiliza a tarefa de visão computacional, pois os modelos generativos (em especial os mecanismos de atenção) também são aptos a lidarem com diversas tarefas – a visão computacional é uma delas. Por sua vez, o segundo exemplo utiliza a tarefa de processamento de linguagem natural.

A Figura 4 exemplifica o funcionamento do mecanismo de atenção na tarefa de visão computacional. O modelo tem como objetivo identificar quais animais estão presentes na imagem original (parte superior da figura). As regiões mais claras nas duas últimas imagens representam as regiões para qual o modelo prestou mais atenção, considerando-as mais importantes na execução da tarefa. Já as regiões mais escuras foram consideradas menos importantes pelo mecanismo de atenção.

Figura 4 *Exemplo de mecanismo de atenção em visão computacional.*

Fonte: Keras (2024)

A Figura 5 ilustra o funcionamento do mecanismo de atenção na tarefa de processamento de linguagem natural. O modelo tem como objetivo encontrar quais são as palavras que merecem mais atenção, por serem consideradas mais importantes, ao ser alimentado com uma palavra de partida. Na matriz, as linhas representam as palavras de entrada e as colunas representam quais palavras devem receber mais atenção de acordo com cada palavra de entrada observada. Para exemplificar, quando o sistema observar a palavra "estava", a maior atenção do modelo será para a palavra "o" (0.59), seguida da palavra "cachorro" (0.08).

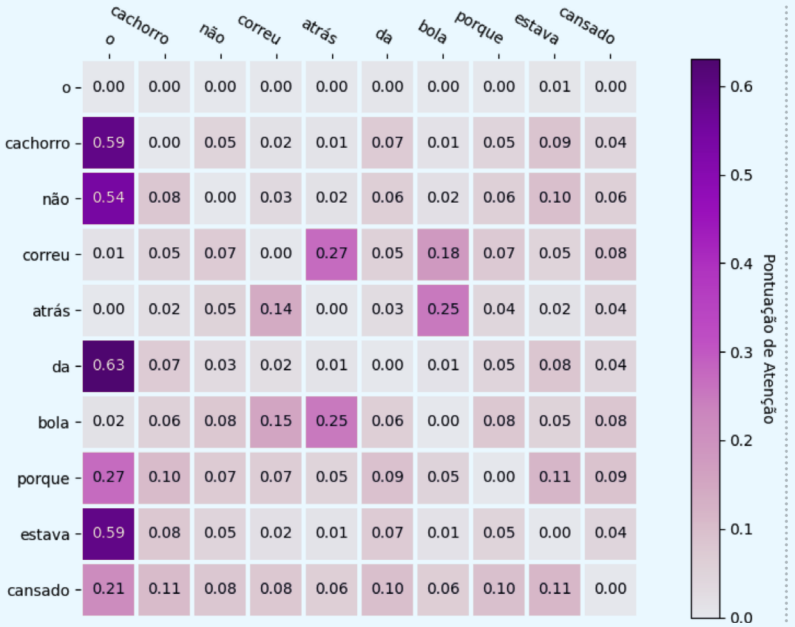

Figura 5 *Exemplo de mecanismo de atenção no processamento natural de linguagem.*

Fonte: ANPD

Um dos principais exemplos de modelos baseados em transformadores é conhecido por *Modelo de Linguagem em Larga Escala*[15] (MLLE). Os MLLEs são treinados em grandes volumes de dados textuais capturando as relações semânticas e sintáticas do contexto e do idioma dos textos da base de treinamento.

A geração de conteúdo textual é realizada com base na máxima probabilidade condicional das palavras do conjunto de treino, considerando o contexto criado com os trechos já gerados. Por esse motivo, os MLLEs são considerados modelos probabilísticos.

Esses modelos apresentam algumas vantagens. Em primeiro lugar, eles têm um desempenho significativamente melhor na compreensão e geração de texto em vários domínios. Além disso, possuem uma capacidade de generalização superior, pois incorporam um conhecimento mais abrangente da linguagem natural.

No entanto, esses modelos também apresentam alguns desafios, como o consumo elevado de recursos computacionais, grande quantidade de dados para treinamento e o risco de gerar respostas enganosas. Em relação a esse último, é preciso considerar que os MLLEs são baseados em probabilidades calculadas nos textos presentes na base de treino. De tal

15 ***Modelo de Linguagem em Larga Escala*** *é um modelo generativo capaz de gerar conteúdo sintético em grande quantidade e com alta precisão.*

forma, existe um risco considerável de geração de conteúdo verossímil, que observa regras de sintaxe e semântica, porém com conteúdo inverídico. Esse risco tem sido referido por especialistas com o termo "alucinações" (BEUTEL, GEERITS e KIELSTEIN, 2003).

Embora os transformadores pré-treinados generativos tenham alcançado resultados impressionantes em várias tarefas de linguagem natural, eles não possuem compreensão real do texto. Eles funcionam principalmente com base em padrões estatísticos aprendidos durante o treinamento e podem ocasionalmente gerar respostas que não fazem sentido ou são gramaticalmente incorretas.

É importante destacar que Modelos de Linguagem não são exclusivamente baseados em AM, havendo modelos baseadas em outras teorias, tal como n-gramas (BAEZA-YATES e RIBEIRO-NETO,2013).

Para terminar este capítulo, apresenta-se outro conceito importante para a IA Generativa: os modelos fundacionais.

Modelo Fundacional[16]

Modelo Fundacional[17] (MF), também referido como modelo de base, é uma rede neural extremamente profunda e complexa composta por bilhões de parâmetros. Os parâmetros são ajustados durante a fase de treinamento, que é realizado em uma quantidade massiva de dados, tipicamente por um longo período e de forma distribuída. Após a fase de treinamento, o MF pode representar complexas entidades, tais como linguagem humana, imagens, vídeos e outros (FREGLY, BARTH e EIGENBRODE, 2023).

São modelos que apresentam uma característica relevante: a capacidade de realizar tarefas para as quais não foi previamente treinado. Por exemplo, um modelo fundacional de análise de imagens pode ser utilizado para identificar anormalidades em imagens médicas, detectar possíveis falhas em infraestrutura para priorizar manutenção, dentre outros usos (ITI, 2023). Logo, a complexidade do MF permite que ele seja aplicado em diversas tarefas.

16 *Do inglês* Foundation Model. *Ainda não há uma tradução formalmente estabelecida do termo para o português, sendo muitas vezes traduzido como Modelo Fundacional, Modelo de Fundação ou Modelo de Base. Diante disso, optou-se por utilizar a tradução Modelo Fundacional.*

17 ***Modelo Fundacional** é um modelo com grandes parâmetros treinados em uma ampla gama de conjunto de dados de maneira auto supervisionada. O termo faz referência a sua importância e aplicabilidade em uma ampla gama de domínios (GARTNER,2024).*

É comum encontrar na literatura a associação de modelos fundacionais apenas a tarefas generativas. Isto ocorre pois os modelos fundacionais são amplamente utilizados no Processamento de Linguagem Natural para a tarefa de geração de texto. Dito isto, é importante entender que esses modelos também são utilizados em reconhecimento e classificação de imagens, conversão de fala em texto, análise de sentimentos, dentre outras tarefas discriminativas (ITI, 2023). Em síntese, não é correto assumir que os termos modelo generativo e modelo fundacional são intercambiáveis.

‹ inteligência artificial generativa e proteção de dados pessoais ›

A Lei Geral de Proteção de Dados Pessoais – LGPD (Lei n°. 13.709, de 14 de agosto de 2018) tem como objetivo proteger os direitos fundamentais de privacidade, liberdade e o livre desenvolvimento da personalidade (LGPD, art. 1º, *caput*), ao mesmo tempo em que tem por fundamento o desenvolvimento econômico e tecnológico e a inovação (LGPD, art. 2º, V). Portanto, a inovação tecnológica deve estar em harmonia com a proteção de dados pessoais. Para a adequada proteção de *dados pessoais*[18], torna-se necessário compreender como os dados são tratados em sistemas de IA Generativa. A recente disponibilização de tais sistemas para uso pela sociedade demonstrou sua rápida popularização e utilização para diferentes propósitos.

Sistemas de IA Generativa apresentam como características fundamentais: (i) necessidade de grandes volumes de dados para seu treinamento; (ii) capacidade de inferência que permite a geração de novos dados semelhantes aos dados de treinamento; e (iii) adoção de um conjunto diversificado de técnicas computacionais, como, por exemplo, as arquiteturas de transformadores para o *Processamento de Linguagem Natural* (PLN)[19] e algoritmos de AM, como as redes adversariais generativas para a geração de conteúdos visuais.

18 ***Dados pessoais:*** *informação relacionada a pessoa natural identificada ou identificável.*

19 ***Processamento de Linguagem Natural:*** *é uma área interdisciplinar que envolve o campo da ciência da computação dedicado à interação entre computadores e a linguagem humana, por meio de programas capazes de processar, analisar, interpretar e gerar dados de linguagem natural. Isso inclui o resumo de textos, tradução automática, reconhecimento da voz, análise de sentimentos, geração de texto e voz, entre outros. A linguagem natural é qualquer língua humana, que pode ser expressa em texto, fala, linguagem de sinais etc.*

Tais características afetam diretamente o *tratamento de dados pessoais*[20] e os princípios que regem a LGPD. A necessidade de grandes volumes de dados pode resultar no tratamento tanto de dados pessoais quanto não pessoais. A coexistência desses dois tipos de dados eleva os riscos relacionados à proteção de dados pessoais, pois aumenta a probabilidade de que dados pessoais sejam tratados sem as devidas salvaguardas e que o princípio da necessidade não seja atendido. Além disso, a capacidade de geração de novos dados, ou conteúdo sintético, coloca em risco a proteção de dados pessoais, uma vez que o conteúdo sintético gerado pode ser indistinguível de dados pessoais e se relacionar a uma pessoa natural identificada ou identificável, bem como ser erroneamente associado a pessoas reais. Por fim, o conjunto de técnicas computacionais ensejam em metodologias complexas e opacas, de modo que o baixo nível de transparência não se deve necessariamente à natureza da técnica utilizada, mas à dificuldade em interpretar suas operações e compreender os processos realizados para a obtenção de resultados.

Assim, os *titulares*[21] e *agentes de tratamento*[22] se encontram diante de sistemas de IA que apresentam desafios significativos para a proteção de dados pessoais. Os riscos vão desde o potencial tratamento de dados pessoais e a geração de conteúdo sintético até a dificuldade de assegurar princípios como a transparência. Tais condições exigem atenção aos princípios e regras estabelecidas pela LGPD e impõem desafios relacionados ao risco de violações de direitos fundamentais.

A seguir, analisa-se a relação da IA generativa com diferentes operações de tratamento de dados pessoais. Além disso, ao final deste capítulo, serão trazidas breves reflexões sobre a IA generativa e alguns princípios da LGPD. Para tanto, adotou-se o seguinte conjunto de referências: CNIL (2023), Glenster; Gilbert (2023), ABNT NBR ISO/IEC 22989 (2023), AEPD (2023), OPC (2023), Solove (2024), Rana *et al*., (2022), Teffé (2017).

A relação entre o tratamento de dados pessoais e os sistemas de IA Generativa

A fim de evidenciar a relação entre o tratamento de dados pessoais e os sistemas de IA Generativa, este estudo delimitou 04 (quatro) conjuntos

20 ***Tratamento de dados pessoais:*** *toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração.*

21 ***Titulares:*** *pessoa natural a quem se referem os dados pessoais que são objeto de tratamento.*

22 ***Agentes de tratamento:*** *o controlador e o operador.*

de elementos que fazem parte do tratamento de dados pessoais. São eles: (i) coleta e armazenamento, (ii) processamento, (iii) compartilhamento e (iv) eliminação.

i) Coleta e Armazenamento de dados para treinamento

O desenvolvimento de modelos de IA Generativa envolvem várias etapas e um dos primeiros passos é a coleta de dados que posteriormente são armazenados e usados para treinar o modelo.

Uma das maneiras de coletar dados para a formação de grandes bases para treinamento e testes de sistemas de IA Generativa, é por meio da técnica de raspagem de dados da *web (data scraping ou web scraping)*. Essa técnica utiliza programas para navegar pela web que realizam a coleta, extração e/ou cópia automatizada de dados criados e disponibilizados pelos usuários da *web* ou por terceiros. Os dados coletados podem incluir qualquer informação presente na web como nomes, sobrenomes, endereços, endereços de e-mail, vídeos, áudios, imagens, comentários, opiniões, preferências, entre outros dados e identificadores, disponíveis em diferentes sítios eletrônicos ou em bases de dados.

A raspagem de dados, a depender do método de automação utilizado, pode selecionar dados específicos, como por exemplo, coletar apenas opiniões e até mesmo definir um intervalo de tempo regular para realizar uma nova coleta. Em outros casos, a raspagem pode operar em uma escala significativa e percorrer bilhões de páginas da web. A dinamicidade e a velocidade de disponibilização de novos dados na web permite o desenvolvimento de diferentes sistemas para realizar a atividade de raspagem e agregação de dados *(data aggregators)*[23].

Como exemplo de grandes bases de dados disponíveis e utilizadas no treinamento de *modelos de IA*[24] pode ser destacada a *Common Crawl*. Essa é uma organização sem fins lucrativos que realiza o rastreamento e a coleta regular de dados na web, por meio de *crawlers*, ou seja, programas automatizados, com o objetivo de gerar repositórios. Trata-se de uma infraestrutura centralizada que reúne dados de diferentes fontes, agrega-

23 ***Agregação de dados:*** *é uma técnica que por meio de sistemas computacionais agrupa dados coletados de uma grande variedade de fontes em um repositório centralizado para acesso e tratamento.*

24 ***Modelos de IA:*** *representação matemática e lógica de um sistema, entidade, fenômeno, processo ou dados. O modelo permite a IA processar, modelar, adaptar-se, interpretar e gerar dados para interagir com o mundo, com capacidade de generalizar o aprendizado para diferentes contextos. Os modelos podem ser divididos em diferentes tipos, de acordo com a sua função ou aplicação, por exemplo, modelos de aprendizado de máquina, modelos baseados em regras, modelos de redes neurais artificiais, entre outros.*

dos e armazenados de forma gratuita e aberta, disponível para qualquer pessoa que deseja realizar pesquisas e desenvolver inovações que requerem grandes conjuntos de dados, inclusive no campo da IA[25].

A natureza abrangente da coleta massiva de dados possibilita que repositórios, como os da *Common Crawl*, ofereçam uma fonte de dados ampla e diversificada obtidos por meio de raspagem a agregação de dados. Para o treinamento de modelos de IA, os desenvolvedores podem mesclar a raspagem de dados realizada de forma direta pelo próprio desenvolvedor, com fontes de outra organização como a *Common Crawl*, bem como com dados de fontes diversas como livros, artigos e publicações científicas, dissertações e teses, transcrições de áudio e vídeo, tabelas, códigos, legislações, entre outros.

A operação de raspagem e agregação de dados em larga escala amplia os riscos em relação à possibilidade de incluir dados pessoais. O amplo escopo de dados que podem ser coletados para o desenvolvimento de modelos, também apresenta a mesma preocupação.

A ausência de etapas de pré-tratamento adequadas para a eliminação ou *anonimização*[26] de dados pessoais possibilita a existência de riscos significativos relacionados ao tratamento indevido de dados pessoais. Os riscos apresentam agravantes nos casos que incluem o tratamento de *dados pessoais sensíveis*[27] e dados de crianças e adolescentes.

É importante destacar que o conteúdo de sites públicos ou acessíveis publicamente estão sujeitos à LGPD, visto que empresas que disponibilizam tais informações possuem obrigações em relação à proteção de dados pessoais em suas plataformas. Da mesma forma, desenvolvedores e empresas que realizam a raspagem na *web* devem garantir a conformidade com a proteção de dados pessoais.

O tratamento de dados pessoais sem o conhecimento dos titulares envolvidos limita o controle destes sobre os seus dados pessoais. A limitação do controle pode ocorrer mesmo após a eliminação dos dados pelos titulares na *web* em virtude da possibilidade de raspagem anterior e seu armazenamento em repositórios.

25 *Maiores informações em: https://commoncrawl.org.*

26 ***Anonimização:*** *utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo.*

27 ***Dados pessoais sensíveis:*** *dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural.*

A operação de raspagem deve indicar a hipótese legal para o tratamento de dados pessoais, de modo que empresas que realizam essa atividade precisam considerar uma das hipóteses legais presentes nos arts. 7 º e 11 da LGPD. Adicionalmente, conforme o art. 7º, §§ 3º e 4º, existe a expressa menção à aplicação dos princípios de proteção de dados nos casos em que os dados pessoais são tornados públicos pelo próprio titular ou de acesso público. Logo, o uso de dados pessoais raspados deve atentar sobretudo aos princípios da boa-fé, finalidade, adequação e necessidade.

Dessa forma, são necessários mecanismos que assegurem a transparência adequada nas etapas de coleta e armazenamento de dados para a formação de grandes bases de dados.

ii) Processamento

O processamento de dados envolve etapas do *ciclo de vida de sistemas de IA*[28] Generativa. A atividade de processar dados é iniciada na fase anterior ao treinamento do modelo, ou seja, durante a formação da base de dados de treinamento e teste e percorre o ciclo de vida dos sistemas de IA Generativa.

A utilização de dados de repositórios originados de processos de raspagem da *web*, bem como a mesclagem com dados raspados de forma direta pelo próprio desenvolvedor com outras fontes podem implicar no uso e reuso de dados para o treinamento e refinamento de diferentes modelos em IA. Na existência de dados pessoais, isso pode conduzir a um tratamento contínuo, irrestrito e ilimitado de dados pessoais em diferentes sistemas de IA Generativa.

A capacidade de sistemas de IA Generativa de gerar novo conteúdo sintético vai além do processamento básico de dados e abrange o aprendizado e a modelagem para gerar novas representações com base nos dados de treinamento.

Durante o treinamento, os parâmetros do modelo recebem seus respectivos valores que representam os *pesos*[29] em relação à capacidade do modelo de gerar, por exemplo, linguagem natural precisa. Os pesos refle-

28 ***Ciclo de vida de sistemas de IA:*** *é um modelo que descreve a evolução e etapas de um sistema de IA, desde o início de seu desenvolvimento até a sua desativação. As etapas não são sequenciais e podem ocorrer muitas vezes de forma iterativa, tais como, gestão de riscos, correções, refinamento do modelo, implementação de melhorias e atualizações do sistema. A decisão de desativar um sistema de IA pode ocorrer em qualquer momento durante a fase de operação e monitoramento.*

29 ***Pesos:*** *variável interna de um modelo que afeta a forma de cálculo das saídas ou resultados. No caso de IA Generativa, os pesos determinam a importância relativa de diferentes entradas para o cálculo das saídas. Para isso, os pesos são ajustados iterativamente durante o processo de treinamento do modelo para capturar padrões e características dos dados de treinamento.*

tem padrões e relações aprendidas, ou seja, como o modelo responde e aprende a partir dos dados de treinamento.

Os modelos não processam e armazenam informações específicas, pois sua característica matemática invisibiliza os dados, ou seja, a existência de dados pessoais nas bases de dados pode ser ocultada por meio de processos matemáticos durante o treinamento. Desta forma, os dados pessoais podem não ser diretamente identificados no modelo. Cumpre ressaltar, que recentes estudos colocam em discussão a possibilidade de pessoas naturais serem identificáveis em razão de vulnerabilidades a ataques do tipo de inversão de modelo *(model inversion)* e de *membership inference* (VEALE, BINNS, EDWARDS, 2018)

Porém, os sistemas de IA Generativa permitem interações dos usuários em linguagem natural com o modelo treinado para a geração de respostas. Desta forma, a depender da forma de interação, instruções e o contexto informado pelo usuário por meio do *prompt*[30], dados pessoais podem ser gerados como resposta em MLLE.

O *conteúdo*[31], embora sintético, ou seja, gerado pelo modelo, apresenta narrativas que podem resultar na produção de conteúdo falso ou inverídico sobre uma pessoa real. Essa possibilidade apresenta riscos em relação à proteção de dados e ao livre desenvolvimento da personalidade, especialmente no que tange ao direito de imagem do titular.

Neste ponto, não se trata apenas da fisionomia e retrato da pessoa, ou seja, sua imagem-retrato como uma expressão externa da pessoa humana, mas também de sua imagem-atributo que está relacionada com um conjunto de características que podem ser associadas a uma pessoa em sua representação no meio social, como a honra, reputação, dignidade ou prestígio.

Por exemplo, em uma interação com o *prompt* para análise de currículos, o sistema pode gerar conteúdo sintético que apresente fatos infundados e inverídicos sobre a vida pessoal do candidato, interpretações errôneas de opiniões, informações inverídicas e qualquer outro conteúdo que pode disseminar informações falsas ou prejudicar a reputação de alguém. Outro exemplo é o caso ocorrido com a atriz Taylor Swift que foi alvo de disseminação viral de imagens sexuais falsas geradas por deepfakes[32].

30 ***Prompt:*** *pode ser traduzido como entrada ou comando. Na computação o* prompt *é uma mensagem ou uma linha de comando em uma interface de usuário. No contexto da IA Generativa, o* prompt *é uma entrada de texto usada para dar instruções ao modelo de IA sobre o que fazer ou qual pergunta esse deve responder.*

31 ***Conteúdo sintético:*** *conjunto de informações composta por dados sintéticos. Nesta seção, optou-se por utilizar esta expressão, para frisar que o conteúdo gerado poderá incluir dados pessoais. É importante, ainda, ter em mente que o termo "dado sintético" é utilizado tanto no campo da computação quanto na área de privacidade e proteção de dados. Na computação, os dados sintéticos podem ser utilizados no desenvolvimento e testes de sistemas de IA quando os dados reais não estão disponíveis nas quantidades necessárias, não existem ou não podem ser tratados, e para o balanceamento de bases de dados. Na área de privacidade e proteção de dados, o uso de dados sintéticos costuma estar associado às* Privacy Enhancing Technologies *(PETs). Neste sentido, a AEPD destaca que o dado sintético será uma técnica de PET, se usada para gerar conjuntos de dados não pessoais com a mesma utilidade que os* ›

Daniel Solove (2024) se refere a esses conteúdos inverídicos gerados pela IA generativa como "materiais malevolentes", devido à sua capacidade de amplificar casos de fraudes e outros golpes. Ele alerta, contudo, que nem sempre esses materiais são gerados de forma intencional pelo operador da IA Generativa, devido ao fenômeno da *alucinação*[33]. A intenção pode ser um elemento importante em regimes de responsabilidade que se baseiam numa concepção de culpa subjetiva, contudo ela tem menor relevância em regimes de responsabilidade objetiva ou baseados na concepção normativa de culpa.

Assim, há um desafio em definir quem deve se responsabilizar pela geração de alucinações referentes a pessoas naturais que possuam efeito danoso. Outro desafio está na conformidade com a LGPD, visto que sistemas de IA Generativa podem gerar dados pessoais sem que tenham sido especificamente treinados para essa finalidade.

iii) Compartilhamento

Em virtude da abrangência do conceito de compartilhamento no tratamento de dados pessoais, que pode ser visto por diferentes perspectivas, o estudo dividiu o compartilhamento em três etapas: (1) compartilhamento de dados pelo usuário do sistema de IA Generativa que pode ser ou não o titular de dados; (2) compartilhamento dos resultados obtidos por meio da interação com o *prompt* em sistemas de IA Generativa com dados pessoais por terceiros; e (3) compartilhamento do modelo pré-treinado com dados pessoais.

1. *Compartilhamento de dados pelo usuário do sistema de IA Generativa que pode ser ou não o titular de dados*

Em primeiro lugar, cabe aqui refletir sobre o compartilhamento de novos dados pessoais por usuários que interagem com esses sistemas (sejam usuários na posição de titular de dados ou na de *agentes de tratamento*[34]) a partir de *prompts* de comando.

› *pessoais. Note, porém, que "conteúdos sintéticos" não são gerados com a finalidade de anonimização.*

32 ***Deepfakes:*** *o termo* deepfake *é derivado do termo* deep learning *e* fake*, ou seja, aprendizado profundo e falso, em português. O termo descreve o conteúdo realístico manipulado como fotos e vídeos gerados pelo aprendizado profundo.*

33 ***Alucinação:*** *o termo alucinação em Modelos de Linguagem em Larga Escala (MLLEs) são caracterizados por conteúdo gerado que não é representativo, verídico ou não faz sentido em relação à fonte fornecida, por exemplo, devido a erros na codificação e decodificação entre texto e representações. No entanto, deve-se notar que a alucinação artificial não é um fenômeno novo (BEUTEL, GEERITS e KIELSTEIN, 2003).*

34 ***Agentes de tratamento:*** *o controlador e o operador. Controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais. Operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador.*

O *prompt* dos atuais sistemas de IA Generativa possibilita compartilhar uma vasta gama de dados para a geração de respostas. Com a evolução dos sistemas, a funcionalidade do *prompt* foi aprimorada e passou a comportar a inclusão de anexos em diferentes formatos. Logo, os usuários podem fornecer instruções e adicionar documentos que implicam diretamente no compartilhamento e tratamento de dados.

As instruções fornecidas pelos usuários podem incluir uma diversidade de informações, como trechos de documentos, mensagens de texto, e-mails, comentários e opiniões disponíveis em diferentes plataformas, detalhes de experiências pessoais, pesquisas acadêmicas, histórico de compras, interações com clientes, registros médicos, dúvidas e relatos sobre procedimentos médicos, entre outros, que podem apresentar dados pessoais e dados pessoais sensíveis.

Adicionalmente, os usuários podem ter a opção de anexar diferentes tipos de documentos em sua integralidade de modo a ampliar a interação e a capacidade de tratamento de dados. Sendo assim, documentos empresariais confidenciais, receitas e laudos médicos, documentos públicos como escrituras e procurações, atas de reuniões, tabelas, figuras, entre outros, podem ser anexados a fim de receber como resultado uma análise, interpretação ou qualquer outro questionamento que o usuário considere pertinente. Assim, as informações disponibilizadas ao *prompt* são utilizadas para o aprendizado do contexto.

Um aspecto relevante sobre os MLLE refere-se à geração de respostas personalizadas. As informações fornecidas no *prompt* são utilizadas para modelar as respostas dentro do contexto, de modo que o contexto da resposta anterior pode ser utilizado para responder uma pergunta posterior dentro do mesmo assunto.

O agente de tratamento e o titular de dados, em muitos casos, podem não ter o conhecimento sobre os riscos envolvidos neste compartilhamento de informações ou confiar no sistema em virtude dos benefícios proporcionados pelos resultados ou assistência recebida. Além disso, se uma pessoa natural, usuário do sistema de IA, compartilha dados pessoais de outros titulares com o sistema de IA Generativa, ele poderá, a depender do contexto, ser considerado um agente de tratamento.

Dentro desse aspecto, os sistemas devem ser desenvolvidos de modo a proteger a privacidade dos usuários em interações que podem envolver o compartilhamento de dados pessoais no *prompt*. Outro aspecto relevante é a ausência de informações claras e facilmente acessíveis sobre o tratamento dos dados pessoais disponibilizados no *prompt*.

A transferência de responsabilidade sobre a proteção de dados pessoais para o usuário a fim de garantir o uso adequado de sistemas de IA Generativa que adotam MLLE não parece ser suficiente para lidar com as consequências atualmente existentes do compartilhamento de dados pessoais via *prompt* de comando.

2. *Compartilhamento dos resultados obtidos por meio da interação com o prompt em sistemas de IA Generativa com dados pessoais por terceiros*

Sistemas de IA Generativa podem permitir que dados pessoais sejam compartilhados com terceiros, quando esses dados compõem o conteúdo sintético gerado por estes sistemas.

Nesse caso, as observações mencionadas na seção anterior (Processamento) devem ser observadas com atenção, principalmente considerando o risco de os dados compartilhados serem reutilizados para finalidades secundárias que dificilmente o desenvolvedor do sistema de IA Generativa conseguirá controlar.

Estabelecer uma cadeia de responsabilidade entre os diferentes agentes envolvidos nesse compartilhamento de dados pessoais se torna um ponto relevante para garantir conformidade à LGPD, embora desafiador.

3. *Compartilhamento do modelo pré-treinado com dados pessoais*

Como os modelos pré-treinados podem ser considerados um reflexo da base de dados utilizada no treinamento, a popularização de criação de *APIs*[35] que adotam modelos fundacionais como os MLLE pré-treinados, traz um novo desafio. O compartilhamento de modelos tende a envolver também os dados que estão matematicamente presentes neles.

35 ***APIs:*** *sigla para* Application Programming Interface*, em português, Interface de Programação de Aplicações; uma forma de permitir, por intermédio de programas de software, a interação entre diferentes aplicativos e extrair dados.*

Este tipo de compartilhamento permite o desenvolvimento de aplicações independentes que realizam um ajuste fino ou *refinamento do modelo fundacional*[36], por meio do treinamento com um conjunto de dados específicos para o domínio pretendido.

Ao relacionar o refinamento do modelo fundacional, com a possibilidade de uso dos resultados obtidos por meio da interação com o *prompt* para a geração de bases de treinamento para o refinamento, a existência de dados pessoais permite um ciclo contínuo de tratamento, como será descrito no tópico seguinte (Eliminação).

O compartilhamento de modelos fundacionais que foram treinados com dados pessoais, bem como o uso desses dados para seu refinamento, pode envolver riscos relacionados à proteção de dados a depender da finalidade desejada.

iv) Eliminação

A etapa de eliminação é aquela na qual o dado pessoal ou o conjunto de dados armazenados em banco de dados são eliminados ao término do seu tratamento.

A definição do término do tratamento de dados pessoais em sistemas de IA Generativa precisa considerar três novos elementos relevantes: a geração de conteúdo sintético, a interação com o *prompt* que permite o compartilhamento de novos dados e o refinamento contínuo do modelo.

Essa integração de elementos em um único sistema de IA Generativa pode apresentar dados pessoais. Esse novo contexto tecnológico pode resultar na possibilidade de tratamento contínuo de dados pessoais e exige novas abordagens.

Desse modo, há um desafio em delimitar o fim do período de tratamento, bem como se a finalidade ou necessidade foram alcançadas, além de dificuldades relacionadas com efetivação da revogação do consentimento do titular em sistemas de IA Generativa (caso essa hipótese legal seja utilizada).

36 ***Refinamento do modelo fundacional:*** *Em inglês* fine tuning *é uma técnica que pode ocorrer no Aprendizado de Máquina para o ajuste de um modelo que já foi treinado por um grande conjunto de dados para seu uso em um domínio específico. O ajuste ocorre por meio de um novo treinamento com um conjunto de dados mais restrito.*

Assim, é importante observar o princípio da responsabilização e prestação de contas (art. 6º, X) por todos os atores da cadeia produtiva de sistemas de IA Generativa, enquanto esses dados pessoais não tenham sido terminantemente eliminados.

Em suma, todo o ciclo de vida do tratamento de dados pessoais e o uso de elementos da Inteligência Artificial Generativa devem ser compatíveis com direitos e liberdades dos indivíduos, de modo que os direitos e princípios que orientam a LGPD sejam observados.

A IA Generativa e os princípios da LGPD

Em relação aos princípios é possível realizar alguns apontamentos. O *princípio da transparência* requer informações claras, precisas e facilmente acessíveis aos titulares de dados. É comum que os titulares de dados não sejam informados sobre a raspagem de seus dados na *web*, a inclusão de seus dados pessoais nas bases de treinamento dos modelos, bem como da possibilidade de que interações com o *prompt* envolvam o compartilhamento de seus dados pessoais ou de terceiros. Portanto, é comum observar a ausência de disponibilização de documentação técnica e não técnica detalhada sobre o tratamento de dados pessoais em diferentes sistemas de IA Generativa.

A existência de documentação técnica detalhada seria um ponto inicial para a verificação de conformidade em relação à proteção de dados e às fontes de dados utilizadas, agregadas e filtradas, de modo a evidenciar as técnicas ou práticas adotadas que podem conduzir a não utilização de dados pessoais. Da mesma forma, a produção de documentação adequada poderia auxiliar no monitoramento dos sistemas de IA Generativa em seu ciclo de vida, de modo a identificar melhorias e permitir o exercício de direitos, no caso de existência de dados pessoais. A documentação poderia reduzir os riscos relacionados a aplicações que envolvem o tratamento de dados pessoais e garantir a transparência.

Por sua vez, o *princípio da necessidade* apresenta um desafio adicional relacionado ao uso de grandes bases de dados em modernos sistemas

de IA Generativa no atendimento ao critério de limitação ao tratamento mínimo necessário para o alcance da finalidade. O princípio não indica uma proibição em relação ao treinamento de sistemas de IA Generativa com grandes volumes de dados, mas envolve reflexões e cuidados antes do treinamento, para evitar a existência de dados pessoais não úteis nas bases de treinamento, bem como inseridos posteriormente por meio do *prompt* ou de anexos.

De maneira similar, embora os demais princípios da LGPD também não proíbam o crescimento e a inovação no campo da IA Generativa, trazem aspectos que precisam ser observados para o desenvolvimento e uso responsável dessas tecnologias. Afinal, é necessário garantir o desenvolvimento responsável e o pleno progresso da Inteligência Artificial Generativa em diferentes áreas em conjunto com o respeito à privacidade e proteção de dados em todo o seu ciclo de vida.

‹ inteligência artificial generativa no contexto brasileiro ›

No contexto brasileiro, algumas iniciativas que envolvem a adoção de sistemas de IA Generativa podem ser observadas. Este estudo apresenta, para fins ilustrativos 03 (três) casos de adoção: (i) Instituição pública; (ii) Área da saúde; (iii) Setor bancário. Contudo, é importante frisar que existem diversos outros contextos de uso para além dos aqui destacados, os quais não foram trazidos por limitação de escopo deste material.

i) Instituição pública

O Tribunal de Contas da União (TCU), em 2023, adotou um modelo personalizado de assistente de redação com base em Processamento de Linguagem Natural e Inteligência Artificial, denominado ChatTCU (TCU, 2023).

O ChatTCU[37] foi desenvolvido após estudos de um Grupo de Trabalho (GT) para avaliar os riscos e as oportunidades de uso de ferramentas de IA. O assistente tem como objetivo inicial otimizar o tempo e auxiliar as equipes do Tribunal na produção de textos, adaptações para linguagem simples, traduções e análises de ações de controle externo.

O Tribunal afirma que o ChatTCU receberá atualizações à medida que novas funcionalidades sejam desenvolvidas e no futuro o sistema terá acesso à base de dados do TCU na íntegra para fornecer informações relevantes e precisas sobre processos específicos no Tribunal.

Uma das características do sistema é que as conversas são protegidas por contrato que garantem a confidencialidade das informações fornecidas pelos usuários. Além disso, a existência de uma IA de uso interno oferece maior segurança quando comparada à uma IA de uso público (como o ChatGPT).

37 *Maiores informações em: https://portal.tcu.gov.br/imprensa/noticias/tcu-adota-modelo-personalizado-de-assistente-de-redacao-baseado-em-inteligencia-artificial.htm.*

ii) Área da saúde

A Fundação Bill & Melinda Gates, em 2023, selecionou a *healthtech* de Curitiba Munai para desenvolver um projeto de Inteligência Artificial a partir de Modelos de Linguagem de Larga Escala, como o ChatGPT (PMC, 2023).

O projeto busca solucionar um problema sanitário que - ocorre no mundo relacionado ao combate da resistência antimicrobiana (RAM) que é a capacidade de micróbios ou bactérias não sucumbirem aos efeitos de medicações como antibióticos, o uso desnecessário de antibióticos e a falta de adesão ao Programa de Gerenciamento de Antimicrobianos (PGA).

Uma assistente virtual será desenvolvida no projeto, responsável por realizar a automação de protocolos hospitalares, suporte à decisão clínica, aprimorar a acessibilidade, interpretação e aplicação de protocolos. Na prática a solução pretende auxiliar médicos e prescritores a fazer o uso racional de antibióticos, visto que na atualidade cerca de metade das prescrições de antimicrobianos são desnecessárias. Ao final, o projeto poderá ser aplicado em hospitais brasileiros e instituições internacionais de saúde.

A empresa foi uma das 51 selecionadas pela fundação que visa a promover soluções com impacto social para questões globais a partir de Modelos de Linguagem de Larga Escala. A Munai já atua na área de gestão em saúde com o uso de dados e IA.

iii) Setor bancário

O Banco do Brasil desenvolveu um assistente que utiliza IA Generativa com base no Modelo de Linguagem de Larga Escala do ChatGPT, com respostas relacionadas ao contexto negocial da instituição, a fim auxiliar os colaboradores responsáveis pelo atendimento ao cliente (BB, 2023).

A técnica utilizada combina interações com o Modelo de Linguagem por meio de engenharia de *prompt* e a recuperação de informações com indexação semântica que utiliza espaço de representação para capturar e organizar o significado dos dados. O sistema encontra-se em fase de implementação e os temas tratados envolvem questões relacionadas ao crédito imobiliário, veículo, energia renovável, ações de sustentabilidade e cartão.

O colaborador poderá solicitar informações ao assistente inteligente quando o cliente entrar em contato com o banco a respeito de algum dos temas. As respostas geradas serão utilizadas tanto para auxiliar o colaborador a fornecer informações quanto para auxiliar na tomada de decisão do cliente.

O modelo será atualizado à medida que for utilizado e treinado para oferecer respostas cada vez mais assertivas, refinadas e personalizadas conforme o número de interações dos temas mais questionados pelos clientes.

‹ perspectivas de futuro ›

Os modelos generativos têm demonstrado um potencial significativo em geração de conteúdos cada vez mais realistas. Tal potencial abre espaço para diversas oportunidades de inovação e desenvolvimento tecnológico,

mas certamente trarão importantes questões relacionadas à proteção de dados e à privacidade. Essas questões se apresentarão como desafios a serem enfrentados pela sociedade.

Um recente avanço na área que merece destaque está associado ao paradigma de *Aprendizado por Reforço*[38]: melhorias no algoritmo conhecido como *Q-Learning*[39] têm o potencial de remodelar o cenário de IA generativa. As implicações dessas melhorias são vastas, abrindo caminho para aplicações com potencial de superar as capacidades dos modelos generativos. Assim, há uma expectativa com o Q-Learning de que a atual limitação dos modelos generativos em resolver problemas matemáticos onde só existe uma resposta certa - ao contrário da geração de conteúdo e tradução onde há uma ampla variedade de respostas certas – seja superada. Isto seria feito por meio do alinhamento de aspectos determinísticos dos algoritmos tradicionais de IA com o potencial criativo e generativos dos MLLE (MCINTOSH *et al*, 2023).

No cenário de produtos tecnológicos, a empresa Google recentemente lançou o produto Gemini (GOOGLE, 2024). O Gemini é um *modelo de linguagem multimodal*[40] com a capacidade de processar dados de diversos tipos, tais como texto, imagens ou código. Em termos técnicos, o Gemini é um produto construído utilizando a arquitetura de transformadores multimodal com bilhões de parâmetros, e permite que o modelo compreenda o significado das informações de forma profunda.

A carência de modelos abertos e detalhados no campo da IA limita oportunidades para estudos mais abrangentes para a compreensão dos impactos e exploração de seu potencial inovador. Desta forma, gradativamente surgem modelos para superar essa lacuna, como o projeto OLMo *(Open Language Model)*, desenvolvido pelo Allen Institute for AI (AI2). O OLMo fornece acesso total ao MLLE, dados de treinamento, código, pesos do modelo e documentação (AI2, 2024).

O modelo passou por um treinamento extensivo em 3 trilhões de *tokens* que garante robustez de desempenho na geração de texto e compreensão de leitura. A possibilidade de acesso ao código aberto do projeto e sua documentação são recursos importantes que podem promover colaboração, transparência e novos desenvolvimentos no campo da IA Generativa.

38 ***Aprendizado por Reforço*** *é um modelo de aprendizado de máquina onde o modelo recebe treinamento somente em termos de reforço positivo (recompensa) e reforço negativo (penalidade). Durante a resolução de problemas, o modelo realiza ações para que a recompensa geral seja maximizada e, ao mesmo tempo, minimize as punições (GARTNER, 2024).*

39 ***Q-Learning*** *é um algoritmo do paradigma de aprendizado por reforço que objetiva encontrar a melhor ação a ser tomada em determinada circunstância.*

40 ***Modelo de linguagem multimodal*** *é capaz de processar e gerar dados de diferentes modalidades, como texto, imagens, sons e código.*

Os desafios da confiabilidade e precisão de MLLE estão passando por avanços por meio do uso de técnicas como o CRAG *(Contextual Retrieval--Augmented Generation)* que procuram enriquecer o processo de geração de conteúdo sintético e reduzir a produção de alucinações. Atualmente, esses desafios ocorrem porque a precisão dos textos gerados não pode ser assegurada apenas pelo conhecimento paramétrico adquirido pelo modelo (Yan *et al.*, 2023). O CRAG introduz um avaliador de recuperação que realiza a avaliação contextual de documentos externos que foca em ações de recuperação de informações chave. Esse processo almeja garantir que informações confiáveis sejam incorporadas ao conteúdo sintético gerado (Yan *et al.*, 2024).

Os avanços na área são constantes e os passos atuais demonstram o desejo de trilhar o caminho da *Inteligência Artificial de Propósito Geral* (IAPG)[41]. A IAPG caracteriza-se por ser um tipo de IA com habilidades gerais de resolução de problemas. A despeito de que a IAPG esteja no horizonte do avanço científico, ela representa um grande passo em direção à criação de sistemas mais capazes de se adaptarem a uma variedade de cenários. De tal forma, há ainda um longo caminho a ser trilhado para alcançar a IAPG. É importante destacar que o conceito de IAPG não deve ser confundido com os conceitos de *Inteligência Artificial Forte*[42], esse último está relacionado as questões de senciência e consciência.

41 ***Inteligência Artificial de Propósito Geral*** *(General Purpose AI) é a forma de uma inteligência artificial que possui a habilidade de entender, aprender e aplicar o conhecimento em uma ampla gama de tarefas e domínios. A Inteligência Artificial de Propósito Geral pode ser aplicada a um conjunto muito mais amplo de casos de uso e incorpora flexibilidade cognitiva, adaptabilidade e habilidades gerais de resolução de problemas. (GARTNER, 2024). O Parlamento Europeu considera o ChatGPT como um exemplo de IA de propósito geral (EUROPEAN PARLIAMENT, 2023)*

42 ***Inteligência Artificial Forte*** *é o ramo da inteligência artificial que considera a capacidade hipotética de uma máquina realmente pensar, ao invés de simular o pensamento (RUSSEL, NORVIG, 2013).*

‹ considerações finais ›

A proteção de dados no contexto da IA Generativa deve utilizar uma perspectiva ética, jurídica e sociotécnica. A inovação tecnológica no campo da IA Generativa poderá trazer novos riscos ou amplificar alguns já conhecidos.

A ampla gama de aplicações da IA Generativa poderá causar diversos impactos na sociedade, como o distanciamento do que é real, produzido, pensado e certificado por um ser humano. Isso poderá afetar a percepção de mundo e implicar riscos ainda desconhecidos para a sociedade. A confiança depositada nas respostas geradas por tais modelos requer maior exploração. Passa a ser necessário verificar os aspectos positivos e negativos envolvidos em sua utilização.

Os dados são elementos essenciais para o desenvolvimento de diferentes sistemas de IA, contudo no campo da IA Generativa os sistemas passam a produzir dados e permitir novas formas de compartilhamento de informações. Os dados passam a fazer parte de todo o ciclo de vida do sistema e seu fluxo contínuo é renovado a cada resposta gerada ou arquivo compartilhado que podem conter dados pessoais.

Na IA Generativa, a preocupação com a presença de dados pessoais no modelo treinado passa a percorrer também o refinamento do modelo e a possibilidade de novos desenvolvimentos com modelos treinados que podem envolver o tratamento de dados pessoais.

A variedade de possibilidades de aplicações futuras, que já se encontram em desenvolvimento, envolvendo áreas como a saúde, instituições públicas e setor bancário, demonstram o seu impacto na sociedade e relação direta com proteção de dados pessoais. Tais pontos merecem maiores debates, pois o desenvolvimento econômico, tecnológico e a inovação, bem como o respeito à privacidade e aos fundamentos da LGPD, são essenciais para o crescimento ético e responsável da Inteligência Artificial e IA Generativa em diferentes domínios e tarefas.

Neste ponto, por se tratar de um estudo preliminar, identificaram-se apenas alguns riscos e fragilidades dos sistemas relacionadas ao tratamento de dados pessoais, não podendo ser considerados como os únicos existentes. Logo, a continuidade de estudos neste campo se faz necessária a fim de garantir a adequada proteção aos titulares de dados ao longo de todo o ciclo evolutivo das tecnologias e promover o desenvolvimento de sistemas éticos e responsáveis.

‹ referências ›

ABNT. Associação Brasileira de Normas Técnicas. ABNT NBR ISO/IEC 22989: 2023. Tecnologia da informação — Inteligência artificial — **Conceitos de inteligência artificial e terminologia**. 1 ed. Rio de Janeiro, 2023.

AEPD. Agencia Española Protección Datos. **Synthetic data and data protection**. 2023. Disponível em https://www.aepd.es/en/prensa-y-comunicacion/blog/synthetic-data-and-data-protection. Acesso em 11 de janeiro de 2024.

AI2. Allen Institute for AI. **Open Language Model: OLMo**. Disponível em: https://allenai.org/olmo. Acesso em: 05 de fev. 2024.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. 2. Ed. Bookman, 2013.

BB. Banco do Brasil. **BB usa tecnologia generativa para apoiar atendimento**. 2023. Disponível em: https://www.bb.com.br/pbb/pagina-inicial/imprensa/n/67461/bb-usa-tecnologia-generativa-para-apoiar-atendimento#/. Acesso em: 01 de fev. 2024.

BEUTEL, G.; GEERTIS, E.; KIELSTEIN JT. A**rtificial Hallucination: GPT on LSD?**. Critical care, 2023.

BRASIL, **Lei Geral de Proteção de Dados**. 2018. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm. Acessado em 11 de julho de 2023.

CNIL. Commission Nationale de l'Informatique et des Libertés. **AI how-to sheets**. Disponível em: https://www.cnil.fr/en/ai-how-sheets. Acesso em: 02. fev. 2024.

CRISTIANINI, N.; SHAWE-TAYLOR, **An Introduction to support Vector Machines and Other Kernel-based Learning Methods**. Cambridge University Press. 2000.

CRS. **Generative Artificial Intelligence and Data Privacy: A primer**. Congressional Research Service, 2023.

DARTMOUTH. **Artificial Intelligence Coined at Dartmouth**. Disponível em <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>. Acesso em 07 de julho de 2023.

EUROPEAN PARLIAMENT. **General-Purpose artificial intelligence**. 2023. Disponível em: https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf. Acesso em 06 de fevereiro de 2024.

FIX, E.; HODGES, J. **Discriminatory Analysis Nonparametric Discrimination: Consistency Properties**. USAF School of Aviation Medicine, 1951.

FOSTER, D. **Generative Deep Learning: Teaching Machines to Paint, Write, Compose and Play**, O`Reilly, 2019.

FREGLY, C.; BARTH, A.; EIGENBRODE, S. **Generative AI on AWS: Building Context-Aware Multimodal Reasoning Applications**. O`Reilly, 2023.

GARTNER. **Gartner Glossary**. Disponível em: https://www.gartner.com/en/information-technology/glossary/foundation-models. Acesso em 02 de fevereiro de 2024.

GLENSTER, Ann Kristin; GILBERT, Sam. **Policy Brief: Generative AI Report**. University of Cambridge. 42 p. 2023.

GOODFELLOW, I., J.; ABADIE, J., P.; MIRZA, M.; XU, B; FARLEY, D., W; OZAIR, S.; COURVILLE, A.; BENGIO, Y. **Generative Adversarial Networks**. *In*: Communications of the ACM, V. 63, N.11, 2014.

GOOGLE. **Gemini**. Disponível em https://deepmind.google/technologies/gemini/#introduction. Acesso em: 2 de janeiro de 2024.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2. ed. Upper Saddle Rive, Prentice-Hall, 1999.

ITI. INFORMATION TECHNOLOGY INDUSTRY COUNCIL. **Understanding AI Foundation Models & The AI Value Chain: ITI's Comprehensive Policy Guide**. 11 p. Ago., 2023.

KERAS. **Grad-CAM class activation visualization**. Disponível em: https://keras.io/examples/vision/grad_cam/#lets-testdrive-it. Acesso em 9 de janeiro de 2024.

MCINTOSH, T., R., SUSNAJAK, T., LIU, T., WATTERS, P., HALGAMUGE, M.K. **From Google to OpenAI Q* (Q-Star): A Survey of Reshaping the Generative Artificial Intelligence (AI) Research Landscape**. arXiv:2312.10868, 2023.

MITCHEL, T. **Machine Learning**. McGraw-Hill Science, 1997.

MONTGOMERY, B. **Taylor Swift AI Images Prompt US Bill to Tackle Nonconsensual, Sexual Deepfakes**. The Guardian (Jan. 30, 2024). Disponível em: https://www.theguardian.com/technology/2024/jan/30/taylor-swift-ai-deepfake-nonconsensual-sexual-images-bill?ref=biztoc.com. Acesso em: 2 de fev. de 2024.

OCDE. **OECD Ai Principles overview**. Disponível em: https://oecd.ai/en/ai-principles. Acesso em 17 de abr. de 2024.

OPC. Office of the Privacy Commissioner of Canada. **Joint statement on data scraping and protection of privacy**. 2023. Disponível em: https://www.priv.gc.ca/en/opc-news/speeches/2023/js-dc_20230824/. Acesso em: 02. fev. 2024.

PMC. Prefeitura Municipal de Curitiba. **Startup de Curitiba é selecionada pela Fundação de Bill Gates para desenvolver projeto de IA**. 2023. Disponível em: https://www.curitiba.pr.gov.br/noticias/startup-de-curitiba-e-selecionada-pela-fundacao-de-bill-gates-para-desenvolver-projeto-de-ia/69799. Acesso em: 31 de jan. de 2024.

QUILAN, R. **Discovering rules by induction from large collections of examples**. *In:* Expert Systems in the Microelectronic, 1979.

RANA, Md Shohel; *et al*. **Deepfake Detection: Systematic Literature Review**. IEEE Access. v. 10, p. 25494 - 25513, fev. 2022.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. São Paulo, GEN LTC, 2013.

SOLOVE, DANIEL J. **Artificial Intelligence and Privacy**. 2024. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4713111. Acesso em 2 de fev. de 2024.

TCU. Tribunal de Contas da União. **TCU adota modelo personalizado de assistente de redação baseado em inteligência artificial**. 2023. Disponível em: https://portal.tcu.gov.br/imprensa/noticias/tcu-adota-modelo-personalizado-de-assistente-de-redacao-baseado-em-inteligencia-artificial.htm. Acesso em: 16 de jan. de 2024.

TEFFÉ, C., S. de. **Considerações sobre a proteção do direito à imagem na internet**. Revista de Informação Legislativa: RIL, v. 54, n. 213, p. 173-198, jan./mar. 2017.

TONG, X., LIU, X., TAN, X., LI, X., JIANG, J., XIONG, Z., XU, T., JIANG, H., QIAO, N., & ZHENG, M. **Generative Models for De Novo Drug Design**. *Journal of medicinal chemistry, 64* (19), 2021.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A., N.; KAISER, L., POLOSUKHIN, I. **Attention Is All You Need**. *In:* 31st Conference on Neural Information Processing System (NIPS), Long Beach, 2017.

VEALE, M.; BINNS, R.; EDWARDS, L. Algorithms that remember: Model inversion attacks and data protection law. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2133, 2018.

YAN, Shi-Qi; *et al*. **Corrective Retrieval Augmented Generation**. arXiv. p. 01 – 155. 2023.

www.gov.br/anpd

ANPD

Autoridade Nacional de
Proteção de Dados