IA Generativa em Ambientes Corporativos: Análise dos Riscos de Segurança e Governança

Jefferson Dias Cardoso

Universidade Federal de Uberlândia
Faculdade de Computação
Bacharelado em Sistemas de Informação

Monte Carmelo - MG
2025

Jefferson Dias Cardoso

IA Generativa em Ambientes Corporativos: Análise dos Riscos de Segurança e Governança

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Área de concentração: Sistemas de Informação

Orientador: Prof[a] Dr[a] Ana Cláudia Martinez

Monte Carmelo - MG
2025

À minha família e amigos
por todo amor, apoio e compartilhamento da vida.

Agradecimentos

Este trabalho é possível graças a várias pessoas que passaram em minha vida e deixaram marcas importantes, mas citar todos esses nomes é uma tarefa impossível. A gratidão sempre foi palavra muito presente na minha vida – não por eu a falar ou demonstrar sempre, mas por sentir, o tempo todo, muita gratidão por tudo que tenho.

Sendo assim, uso este espaço como um lembrete de que sou muito grato, primeiramente a Deus pela vida e todas as bênçãos que recebo diariamente, o que envolve a oportunidade de cursar e concluir a graduação.

Grato à minha família por me apoiar e orientar nos caminhos da vida. Juscelena, minha mãe, por sempre ser luz e não medir esforços para me ver feliz; a meu pai, Wilmar, pelas brincadeiras, por toda orientação e conselhos acerca do mundo; aos meus avós, Yone e Valdemar (*in memoriam*) por terem me criado, educado e dado melhor de si para que eu pudesse ser quem sou hoje; à minha tia, Wilmariza, por ser inspiração e, sempre, ponto de apoio; à Vanusa, Gabriel e Marco por todo apoio e parceria durante a vida.

Grato por ter amigos incríveis que sustentam os momentos ruins e completam os bons. À Cassila (Cila), por ser minha "irmã mais velha" e ajudar a tornar a vida mais leve; à Dayane (Day) por todo conselho, confidencialidade e conversas que moldaram (e moldam) significativamente quem eu sou; ao Nícolas, por toda vivência e compartilhamento de vida; à Lílian por estar sempre na torcida por minha felicidade e realização; ao Victor por se fazer sempre presente. Ainda, aos meus amigos de Monte Carmelo, onde pude fazer morada, que são verdadeiramente irmãos: Jayne (jay), por ser sempre companheira e disposta a ajudar; à Maria Olívia (Mel), por ser quem sempre, verdadeiramente, ouve; ao Gabriel (Gabs), por todas as dores e felicidades que vivemos juntos, e; a todos os agregados da República Sede por todos os momentos incríveis vividos lá.

Grato aos professores que tive ao longo de toda a vida acadêmica por serem parte importante do que sou. Em especial à Ana Cláudia, por ter aceitado a orientação deste trabalho e por ter acreditado em mim desde os períodos iniciais do curso; por fim, à banca, por ter aceitado participar deste momento tão importante na trajetória acadêmica.

A computação, ao se fundir com outras áreas, reinventa a forma como compreendemos e interagimos com o mundo.

Resumo

A expansão atual do uso de Inteligência Artificial (IA) tem revolucionado o ambiente empresarial, proporcionando benefícios que vão desde a automação de tarefas simples até a análise de grandes volumes de dados. Contudo, o compartilhamento de dados internos de uma organização com tais ferramentas levanta preocupações significativas, incluindo riscos à segurança da informação, à privacidade dos dados e ao controle sobre informações sensíveis.

Este trabalho tem como objetivo principal avaliar e comparar a aderência de ferramentas gratuitas de Inteligência Artificial Generativa (IA Gen) às boas práticas de segurança, privacidade e governança. Para isso foram examinadas políticas de privacidade e documentação de ferramentas populares de IA Gen em relação a diretrizes estabelecidas por entidades de promoção à segurança da informação.

Os resultados indicam que, embora haja compromisso declarado com a proteção de dados, persistem lacunas importantes, especialmente quanto ao uso de dados fornecidos pelos usuários, à transparência no compartilhamento de informações e à conformidade legal. Tais fragilidades podem representar riscos adicionais para o uso dessas ferramentas em ambientes corporativos.

Palavras-chave: Inteligência Artificial Generativa, Privacidade de Dados, Compartilhamento de Dados, Governança de Dados, Segurança da Informação.

Lista de ilustrações

Lista de tabelas

Lista de siglas

API Interface de Programação de Aplicações - *Application Programming Interface*

CC Ciência da Computação

CNN Rede Neural Convolucional - *Convolutional Neural Network*

CLM Modelagem de Linguagem Causal - *Causal Language Modelling*

DOS Sistema Operacional em Disco - *Disk Operating System*

ENISA Agência Europeia para a Segurança das Redes e da Informação - *European Union Agency for Cybersecurity*

GAN Redes Generativas Adversárias - *Generative Adversarial Network*

GPT Transformador Generativo Pré-treinado - *Generative Pre-trained Transformer*

IA Gen Inteligência Artificial Generativa

IA Inteligência Artificial

LLM Modelos de Linguagem de Grande Escala - *Large Language Models*

LoRA Adaptação de Baixa Classificação - *Low-rank Adaptation*

LGPD Lei Geral de Proteção de Dados

MIT Instituto de Tecnologia de Massachusetts - *Massachusetts Institute of Technology*

MLM Modelagem de Linguagem Mascarada - *Masked Language Modelling*

NIST Insituto Nacional de Padrões e Tecnologia - *National Institute of Standards and Technology*

OECD Organização para Cooperação e Desenvolvimento Econômico - *Organisation de coopération et de développement économiques*

PLN Processamento de Linguagem Natural

RLHF Reforço com *Feedback* Humano - *Reinforcement Learning with Human Feedback*

RNN Rede Neural Recorrente - *Recurrent Neural Network*

RGPD Regulamento Geral sobre a Proteção de Dados (União Europeia)

SNARC Calculadora Neural-Análoga Estocástica de Reforço - *Stochastic Neural Analog Reinforcement Calculator*

SE Sistemas Especialistas

Sumário

Capítulo 1

Introdução

A inteligência artificial (IA) é uma área de estudo da Ciência da Computação que busca desenvolver sistemas capazes de realizar tarefas simulando a inteligência humana (MACHADO, 2023), como tradução de idiomas, tomada de decisões, resolução de problemas e conversas em linguagem natural. A IA evoluiu significativamente desde suas primeiras concepções, passando por diferentes fases de progresso e enfraquecimento.

Segundo McCarthy (2007), a história da IA como conhecida atualmente remonta à década de 1940, com o primeiro grande estudo que sugeria a implementação de redes neurais artificiais, fundamentado na compreensão da fisiologia cerebral, na lógica proposicional e na teoria da computação de Alan Turing. Já na década de 70, com pesquisas em redes neurais avançadas, pesquisas em Aprendizado de Máquina e Processamento de Linguagem Natural (PLN) foram alavancadas, o que deu espaço ao desenvolvimento de sistemas computacionais capazes de realizar atividades autônomas, *chatbot*, processamento de fala, entre outro. Para Cozman (2018), entretanto, a falta de poder de processamento fazia com que as IAs fossem bastante frágeis, apresentando falhas e respostas lentas, o que as colocou em um cenário conhecido como "O Inverno da IA", tornando-as alvo de críticas e desincentivo financeiro.

Posteriormente, o aumento do poder computacional deu espaço ao desenvolvimento de Sistemas Especialistas capazes de realizar tarefas complexas em curtos períodos de tempo, reacendendo o interesse do mercado corporativo em soluções de IA (MACHADO, 2023). Além disso, a popularização de dispositivos como microcomputadores e *smartphones*, juntamente com o crescimento exponencial da coleta de dados por meio de redes, abriu novos horizontes para o uso da IA e expandiu suas atribuições para além das academias de pesquisa. Esses avanços facilitaram a análise de grandes volumes de dados, impulsionaram o desenvolvimento de algoritmos mais sofisticados e expandiram ainda mais as aplicações da IA em diversos setores, como o de Inteligência Empresarial (do inglês *Bussiness Intelligence*), serviços automatizados de atendimento ao cliente ("*chatbots*"), saúde, finanças e transporte, além de ser a ferramenta crucial na viabilização da 4ª Revolução Industrial – caracterizada pelo desenvolvimento de indústrias inteligentes

(BAZZAN et al., 2023). A integração dessas tecnologias tem transformado a forma como as empresas operam, destacando a IA como um componente essencial para inovação e competitividade no mercado.

A partir dessa base de inovação e aplicação da IA, uma subárea tem ganhado foco nos últimos tempos: a chamada IA Gen, esta têm capacidade de gerar (daí a denominação "Generativa") imagens, textos, músicas e outros tipos de dados muitas vezes indistinguíveis daqueles criados por humanos. Machado (2023) destaca que apesar do sucesso atual as IA Gen têm origem atrelada às pesquisas da década de 70 em PLN, que buscavam o desenvolvimento de modelos computacionais capazes de ensinar máquinas a compreender e reproduzir a linguagem natural. Esse avanço, inicialmente voltado para o processamento de linguagem, abriu portas para a criação de modelos mais complexos, capazes de gerar e lidar com diversos tipos de dados e tem transformado o modo com que as empresas podem explorar a tecnologia.

No ambiente corporativo, em especial o industrial, as ferramentas IA têm sido amplamente adotadas para otimizar processos, melhorar a eficiência e proporcionar automação de tarefas (BAZZAN et al., 2023). Durante a última década, as organizações começaram a implementar IA em diversas áreas, desde a automação de tarefas administrativas até a análise preditiva e apoio à tomada de decisão estratégica. O relatório AI Index Report[1] produzido pela Universidade de Stanford (2024), que apresenta os avanços e tendências de IA globalmente, destaca o aumento de 5% do uso de ferramentas de IA nas organizações, incluindo o uso de IA Gen e, se comparado ao ano de 2017 (ano do primeiro relatório em questão), esse aumento foi de 35%, desta forma, os dados apresentados no relatório evidenciam a tendência das organizações em incorporar em seus processos e operações o uso de tecnologias de IA.

Com o avanço do uso das ferramentas de IA, a segurança da informação empresarial torna-se uma questão crucial nesse contexto, as organizações precisam garantir que seus dados estejam protegidos contra ameaças e violações de privacidade. Segundo o mesmo relatório de Stanford, AI Index Report, o uso de IA Gen em 2023 se deu em sua maioria na criação de rascunhos de documentos de texto (9%), seguido do *marketing* personalizado (8%) e criação de imagens e/ou vídeos (8%).

Considerando que IA Gen aprende de maneira contínua através de entradas de usuários, uma determinada informação pode ser utilizada para responder outro usuário indevidamente, expondo propriedade intelectual ou segredos comerciais (KPMG, 2023), potencializando o risco de quebra de confidencialidade mesmo quando o uso é de boa fé. Além disso, há o risco de mau uso e de imprecisões nos conteúdos gerados pela IA Gen, o que pode afetar a qualidade dos resultados e tomadas de decisão. Pontos esses, indicam, portanto, que uma política de governança de dados bem estruturada é fundamental para

[1] AI Index Report, <https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf>

garantir que a utilização dessas ferramentas seja feita de maneira responsável.

Deste modo, este trabalho pretende contribuir para a compreensão dos riscos associados ao compartilhamento de dados empresariais com ferramentas de IA Gen. Ao explorar os principais desafios, espera-se que esta pesquisa forneça *insights* valiosos para organizações e colaboradores que buscam integrar IA em seus processos de forma segura e eficiente. O escopo deste concentra-se nos riscos associados ao compartilhamento de dados e como isso pode comprometer a privacidade, a segurança e a governança dos dados organizacionais.

1.1 Motivação

A ascensão da Inteligência Artificial Generativa representa um marco na evolução tecnológica e, tais ferramentas têm despertado o interesse de organizações em busca de inovação, produtividade e competitividade. Entretanto, o uso dessas tecnologias traz consigo uma camada de complexidade no que se refere à segurança, privacidade e governança de informações (Organisation for Economic Co-operation and Development (OECD), 2024).

Tais ferramentas operam com base em modelos treinados a partir de grandes volumes de dados, que continuam sendo alimentados por novas interações realizadas pelos usuários e, no contexto organizacional, isso levanta uma questão sensível: dados estratégicos podem estar sendo compartilhados de forma inadvertida por colaboradores, sem que haja plena consciência dos riscos envolvidos (KPMG, 2023). Ainda que empregadores e desenvolvedores de tecnologia adotem políticas de segurança, muitas vezes essas normativas não contemplam orientações claras sobre seu uso, o que resulta em lacunas operacionais e vulnerabilidades potenciais (Organisation for Economic Co-operation and Development (OECD), 2024).

Essa realidade evidencia a necessidade de analisar a clareza da documentação frente às formas com que os dados inseridos são tratados, armazenados e possivelmente utilizados, bem como avaliar se tais práticas estão alinhas a diretrizes técnicas e legais. A motivação central deste trabalho, portanto, é fomentar uma reflexão crítica sobre os impactos do uso indiscriminado de ferramentas de IA Gen no ambiente empresarial e oferecer subsídios para uma adoção mais segura, ética e responsável dessas tecnologias.

1.2 Problema

A atual expansão do uso de inteligências artificiais generativas traz consigo facilidades no cotidiano empresarial. No entanto, o compartilhamento de dados internos de uma organização pode gerar consequências importantes, como o aumento dos riscos à segurança da informação, à privacidade de dados e a capacidade de controle sobre as informações.

1.3 Hipótese

Este trabalho adota a hipótese de que as políticas de privacidade das ferramentas de IA Gen atuais são insuficientes para eliminar os riscos associados ao compartilhamento de dados com ferramentas de IA Gen.

1.4 Objetivos

Este trabalho tem como objetivo principal avaliar e comparar a aderência das ferramentas gratuitas de IA Gen às boas práticas e normas relacionadas à segurança da informação, privacidade de dados e governança. Quanto aos objetivos específicos, tem-se:

- ❑ Examinar como ferramentas de IA Gen populares tratam os dados inseridos pelos usuários, com foco em coleta, uso, compartilhamento e medidas de segurança;
- ❑ Comparar essas práticas com diretrizes internacionais, como Agência Europeia para a Segurança das Redes e da Informação (ENISA), Insituto Nacional de Padrões e Tecnologia (NIST) e Organização para Cooperação e Desenvolvimento Econômico (OECD), e observar conformidade com legislações de proteção de dados, como a Lei Geral de Proteção de Dados (LGPD);
- ❑ Identificar lacunas, riscos e vulnerabilidades relacionadas ao uso dessas ferramentas no ambiente empresarial;

1.5 Organização da Monografia

Este trabalho estrutura-se inicialmente na apresentação da fundamentação teórica sobre IA, Inteligência Artificial Generativa (IA Gen) e governança de dados, estabelecendo bases conceituais necessárias para o desenvolvimento da pesquisa. A partir desse referencial teórico, o estudo parte para o capítulo dedicado aos experimentos, onde são realizados levantamentos documentais, seguido de análises que sustentam as conclusões do trabalho.

Capítulo 2

Fundamentação Teórica

2.1 A Inteligência Artificial

A definição de Inteligência Artificial não é um consenso literário, variando conforme o contexto e o objetivo do estudo. Para Russell e Norvig (2013), a definição pode ser dividida em quatro principais estratégias de estudo e abordagem metodológica (sem qualquer ordem): máquinas capazes de simular o pensamento humano de forma abrangente, incluindo processos como tomada de decisões, resolução de problemas e aprendizado; máquinas capazes de realizar tarefas que requerem inteligência humana; sistemas que simulam a percepção, raciocínio e ações humanas, reproduzindo capacidades cognitivas por meio de modelos computacionais; 4) o estudo do projeto de agentes inteligentes, onde a IA está ligada ao desempenho inteligente de artefatos. Assim, a IA pode ser definida de diferentes formas, dependendo do objetivo de quem a estuda, não se limitando a uma definição universal.

Partindo deste pressuposto, diversos autores definem conceitualmente a Inteligência Artificial. A exemplo, McCarthy em seu trabalho intitulo *"What is Artificial Intelligence?"* (2007), define que a IA se consiste no desenvolvimento de máquinas e programas inteligentes, inerentes à ciência e engenharia e, embora este esteja relacionada à modelagem da inteligência humana, ela não deve se limitar a padrões observáveis na biologia.

A definição de IA também pode ser abordada de uma forma mais prática e aplicada, como sugerido por Luger (2008), que considera a IA como o campo da Ciência da Computação (CC) focado em criar sistemas capazes de realizar tarefas que, quando realizadas por seres humanos, requerem inteligência – o que inclui tarefas como raciocínio lógico, aprendizado, reconhecimento de padrões e tomada de decisões autônomas, isto é, o desenvolvimento da automação do comportamento inteligente humano.

Diante do vasto acervo literário que contempla fundamentalmente os conceitos de Inteligência Artificial, este trabalho utiliza das aplicações práticas atribuídas à tecnologia independente do método e abordagem, para definir a IA como área da ciência que objetiva desenvolver sistemas que executam tarefas que exigem inteligência humana, como

comunicação, raciocínio e aprendizado, e deste modo, incorporando características como processamento de linguagem natural, que possibilita interações fluentes; representação de conhecimento para armazenar e recuperar informações; raciocínio automatizado para tomar decisões e formular respostas; e aprendizado de máquina, permitindo adaptação e identificação de padrões em novos contextos que contribuem para aplicações práticas em diversas áreas, como sistemas financeiros, de mobilidade urbana, traduções automáticas, saúde, automatização de sistemas produtivos, dentre outros (BAZZAN et al., 2023).

2.1.1 Origem e Evolução da Inteligência Artificial

O caminho percorrido pela IA até o cenário atual foi marcado por períodos de grande entusiasmo e profundas decepções. Esses "picos e vales", como conhecidos na literatura, moldaram a trajetória da tecnologia ao longo das décadas. Compreender essa evolução histórica é fundamental para evitar a repetição de erros que já prejudicaram o avanço da área e para aproveitar as lições de seus momentos de sucesso. Essa compreensão histórica permite que a IA continue a evoluir de maneira mais robusta e responsável, garantindo que seu impacto positivo na sociedade seja maximizado e riscos sejam minimizados.

A trajetória da Inteligência Artificial antecede em séculos sua formalização como um campo acadêmico, tendo o desejo de representação de inteligência humana em objetos retratado até mesmo em mitologias antigas. Nos séculos XVII e XVIII, o avanço do conhecimento filosófico e científico aliado à técnica prática consolidou as bases tecnológicas que visavam criar máquinas para simular ações humanas, abrindo inquietações à possibilidade de imitação das máquinas ao pensamento humano. Entretanto, no que diz respeito à Inteligência Artificial moderna, a qual se debruça este trabalho, seu início é marcado pelo período pós-Segunda Guerra Mundial (1939-1945), com o marco-zero estabelecido em 1956 durante um seminário no colégio Dartmouth (intitulado *Dartmouth Summer Research Project*), em New Hampshire, Estados Unidos. Essa conferência, organizada por John McCarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester é amplamente reconhecida como o ponto de partida formal da IA como disciplina acadêmica e campo de pesquisa. A proposta do evento era descrever aspectos da inteligência e aprendizado de forma que fosse possível implementá-los em máquina. Neste momento foi então cunhado o termo "Inteligência Artificial" (BARBOSA; BEZERRA, 2020). As ideias apresentadas em Dartmouth impulsionaram uma série de avanços experimentais, levando ao desenvolvimento de algoritmos iniciais e sistemas que, embora rudimentares, serviram de alicerce para os modelos mais complexos que viriam nas décadas seguintes, (RUSSELL; NORVIG, 2013).

Apesar de o marco inicial da IA ser amplamente reconhecido como o seminário de Dartmouth, suas raízes remontam a um período anterior. Para Barbosa e Bezerra (2020), a história da IA não tem uma caminho linear e portanto não é progressiva, diante disso, a literatura aponta que um dos primeiros trabalhos fundamentais para o desenvolvimento

da Inteligência Artificial foi realizado por McCulloch e Pitts em 1943. Baseando-se na fisiologia do cérebro humano, na lógica proposicional e na Teoria da Computação de Alan Turing, eles propuseram um modelo de neurônio artificial. Nesse modelo, os "neurônios" recebem sinais numéricos como entrada e os combinam por meio de funções de soma ponderada – dessa forma, os pesos associados às conexões podem intensificar ou atenuar a interação entre os neurônios da rede (BAZZAN et al., 2023).

Posterior a isso, em 1950 Turing realizou um importante trabalho à área ao propor o Teste de Turing, uma forma de avaliar a capacidade de uma máquina em exibir um comportamento inteligente indistinguível do de um ser humano. Esse teste foi descrito em seu artigo de intitulado "*Computing Machinery and Intelligence*" (em português "Computadores e Inteligência") - cuja a hipótese central era a de que máquinas poderiam pensar como humanos -, considerando como o texto fundador da IA (BARBOSA; BEZERRA, 2020). O Teste de Turing consiste em uma experiência na qual um avaliador interage, por meio de uma interface de texto, com uma máquina e um ser humano, sem saber qual é qual. Se a máquina conseguir enganar o avaliador, fazendo-o acreditar que ela é humana em uma porcentagem significativa das interações, então ela pode ser considerada capaz de "pensar" ou ter inteligência similar à humana. Esse trabalho foi fundamental porque trouxe uma perspectiva filosófica e prática sobre o que significa uma máquina ser inteligente e ajudou a estimular o desenvolvimento de sistemas de IA nas décadas seguintes. A partir dos conceitos de Turing e das redes neurais propostas por McCulloch e Pitts, Marvin Minsky cria então, em 1951, o primeiro computador de rede neural, operando em Sistema Operacional em Disco (DOS) (do inglês "*Disk Operating System*"), o Calculadora Neural-Análoga Estocástica de Reforço (SNARC) – do inglês *Stochastic Neural Analog Reinforcement Calculator*, a qual foi testada a capacidade da máquina em aprender sair sozinha de um labirinto virtual (BARBOSA; BEZERRA, 2020), alavancando o aprendizado de máquina, área que se tornariam pilares da IA moderna.

Nas décadas posteriores surge então um grande interesse da comunidade acadêmica e comercial em ferramentas de IA, entretanto, dada a capacidade de *hardware* à época, os produtos desenvolvidos não passavam de objetos capazes de realizar operações matemáticas. Neste contexto, Russell e Norvig (2013) destacam o que chamam de "Lista X" de Turing, que reflete uma visão crítica sobre as limitações das máquinas, em especial às tarefas mais complexas e subjetivas da experiência humana, tal lista descreve uma série de itens propostos que uma máquina nunca seria capaz de fazer, como ser amável, diligente, aprender a partir da experiência, usar palavras corretamente, fazer algo realmente novo.

Durante as décadas seguintes, entre 1950 e 1960, houveram avanços significativos no que tange a IA na comunidade acadêmica, científica e comercial (BARBOSA; BEZERRA, 2020). Em 1957, foi desenvolvido o primeiro *hardware* que implementava o algoritmo Perceptron por Frank Rosenblatt (1928-1971), definindo o Teorema da Convergência do Perceptron, que marcou um dos primeiros modelos de aprendizado supervisionado e rede

neural de uma camada capazes de realizar classificações binárias simples, utilizando os conceitos propostos por McCulloch e Pitts em 1943. Esse modelo foi pioneiro em demonstrar que uma máquina poderia aprender a classificar padrões a partir de exemplos, gerando um grande entusiasmo e estimulando pesquisas adicionais sobre redes neurais. Entretanto, apesar do impacto inicial, as limitações teóricas e práticas do Perceptron, como foi demonstrado por Minsky e Papert (1969), levaram a um período de estagnação nas pesquisas da área de redes neurais (RAUBER, 2005), e ganhando novo impulso em meados 1980 com o desenvolvimento de algoritmos de aprendizado por retropropagação, conforme abordado posteriormente, para redes neurais de múltiplas camadas (RUSSELL; NORVIG, 2013) e servindo como base para estudos em Aprendizado do Máquina ou *Machine Learning* (MACHADO, 2023).

Outro marco importante na história da IA foi dado por McCarthy ao desenvolver, em 1958, a linguagem LISP, uma das primeiras linguagens de programação voltadas para a Inteligência Artificial. LISP, que vem do termo em inglês *LISt Processing*, foi projetada para lidar com símbolos e manipulação de listas, o que a tornava particularmente adequada para expressar e processar conhecimento em IA, especialmente nas áreas de raciocínio simbólico e resolução de problemas (NILSSON, 1998). A flexibilidade da linguagem permitiu que fosse rapidamente adotada em várias pesquisas de IA, tornando-se uma das ferramentas essenciais para o desenvolvimento de algoritmos em áreas como aprendizado de máquina e redes neurais, se consagrando, segundo Russell e Norvig (2013), a principal linguagem para desenvolvimento de IA pelos próximos anos. O desenvolvimento da LISP proporcionou à comunidade acadêmica e comercial uma plataforma poderosa para a criação de sistemas inteligentes que podiam aprender e resolver problemas de maneira autônoma e foi pilar para o desenvolvimento de linguagens como Python, JavaScript e Swift.

Como dito anteriormente, a história da IA não é linear e faz-se necessário a volta na linha do tempo entre meados de 1950 e 1960, onde começaram a ser realizados os primeiros estudos em Processamento de Linguagem Natural (PLN), um ramo multidisciplinar que combina conceitos de linguística e computação para permitir que as máquinas compreendam, interpretem e realizem tarefas com base na linguagem humana (BAZZAN et al., 2023). O objetivo do PLN era criar sistemas capazes de interagir com os usuários de forma mais intuitiva, facilitando a comunicação entre seres humanos e computadores. Esses avanços iniciais começaram a dar forma a um campo que se expandiria enormemente nas décadas seguintes e serviriam de base à IA Gen, que será abordada em outro momento.

Um dos marcos importantes a partir do desenvolvimento da PLN foi a criação do *chatbot* ELIZA, em 1966, pelo cientista da computação Joseph Weizenbaum no Instituto de Tecnologia de Massachusetts (MIT). ELIZA foi um dos primeiros programas a simular uma conversa humana, identificando palavras-chave na entrada do usuário e gerando

respostas pré-programadas com base nessas palavras. O mais famoso dos *scripts* de ELIZA foi o "DOCTOR", que imitava um psicoterapeuta Rogeriano[1], criando a ilusão de uma conversa inteligente (SHARMA; GOYAL; MALIK, 2017). Esse experimento pioneiro ajudou a popularizar a ideia de que as máquinas poderiam, de alguma forma, compreender e processar a linguagem humana, servindo como um precursor das tecnologias de *chatbots* modernas e dos assistentes virtuais baseados em PLN.

Ainda neste período, entre as décadas de 1960 e 1970, o foco dos estudos em IA foi fortemente direcionado para sistemas baseados em conhecimento, surgindo os chamados Sistemas Especialistas (SE), programas projetados para manipular conhecimento e informações de forma inteligente, visando resolver problemas que exigem uma grande quantidade de conhecimento especializado. Esses sistemas possuem uma base de conhecimento, onde são armazenados dados relevantes de uma área específica, permitindo simular a tomada de decisões de um especialista humano e assim oferecendo soluções para situações complexas, automatizando decisões com base em conhecimento especializado e aprimorando a eficiência e precisão no processo decisório (MENDES, 1997). Nos anos seguintes, os Sistemas Especialistas, que manipulam grandes volumes de conhecimento explícito, passaram a integrar-se com os avanços das redes neurais e aprendizado de máquina. Essa integração permitiu que o conhecimento fosse adquirido de forma automática, ou com pouca interferência humana, ampliando as capacidades dos sistemas. Exemplos de técnicas automáticas incluem mineração de dados e aprendizado de máquina, como redes neurais e árvores de decisão (COSTA; SILVA, 2005).

Após isso, por volta de 1974, a IA entra em uma fase de desaceleração significativa no entusiasmo e nas investigações que duraria até 1980. Isso aconteceu devido a uma série de fatores, como as limitações práticas das tecnologias da época, a falta de poder computacional e a incapacidade de muitas das promessas iniciais da IA de entregar resultados concretos (COZMAN, 2018). Durante esse período, as expectativas em relação à IA foram excessivamente otimistas, mas os resultados não corresponderam às previsões (RUSSELL; NORVIG, 2013), o que levou a uma diminuição dos investimentos, tanto do setor público quanto privado, em pesquisas de IA.

> A maioria dos pesquisadores em IA e em áreas relacionadas confessa um sentimento pronunciado de decepção com o que foi alcançado nos últimos vinte e cinco anos. Os trabalhadores entraram na área por volta de 1950 e, até mesmo em 1960, com grandes expectativas que estão muito distantes de terem sido realizadas em 1972. Nenhuma parte do campo teve as descobertas feitas até agora resultando no

[1] A Psicoterapia Rogeriana, criada por Carl Rogers, explicada brevemente, foca em ouvir atentamente o cliente e, a partir disso, fazer perguntas que o incentivem a refletir sobre seus sentimentos e encontrar suas próprias soluções, com empatia e aceitação.

> grande impacto que foi prometido na época. (LIGHTHILL, 1973)

Esse "inverno" foi interrompido na década seguinte, principalmente em 1986, quando houve um renascimento da IA com o aprimoramento das redes neurais e o melhorias do algoritmo de retropropagação, que possibilitou o treinamento mais eficiente de redes neurais de múltiplas camadas.

Em meados de 1980, houve uma redescoberta e aprimoramento do algoritmo de retropropagação (RUSSELL; NORVIG, 2013), originalmente proposto em meados 1969. Esse algoritmo é uma técnica usada para treinar redes neurais artificiais, ajustando os pesos das conexões entre os neurônios com base nos erros – ele propaga esse erro de volta através das camadas da rede para melhorar a precisão das previsões, ajustando os pesos de modo a minimizar o erro durante o treinamento. Aplicado a problemas de aprendizado em várias áreas, como CC e Psicologia, o algoritmo ganhou destaque após a publicação da coletânea *Parallel Distributed Processing* - em português Processamento Distribuído Paralelo, (MCCLELLAND; RUMELHART; HINTON, 1988), que trazia uma série de resultados dos casos de uso do algoritmo. O desenvolvimento das redes neurais conexionistas — aquelas que conectam neurônios em várias camadas para processar informações — foi visto como uma alternativa aos modelos simbólicos tradicionais da IA – que se baseiam em símbolos e lógica–, desafiando a ideia de que a manipulação de símbolos seria fundamental para explicar a cognição humana, como dito por Russell e Norvig (2013). Além disso, a retropropagação permitiu o treinamento mais eficaz de redes neurais profundas, o que abriu caminho para a criação de modelos mais complexos que conseguem capturar padrões intricados na linguagem natural, alavancando a PLN. É possível destacar ainda que no que diz respeito ao acesso a dados e conectividade, à medida que a internet se expandiu durante a década de 1990, um aumento massivo na quantidade de dados disponíveis e a possibilidade de compartilhá-los em tempo real proporcionaram um terreno fértil para o desenvolvimento (ou pelo menos ideias) de algoritmos de IA, especialmente no contexto de aprendizado de máquina e redes neurais. Esse avanço nas redes neurais, aliado ao aumento do poder computacional, foi essencial para revitalizar o interesse em IA, dando início ao que seria o "Verão da IA" nas anos seguintes (BARBOSA; BEZERRA, 2020).

Ainda à de 1980, a IA experimentou uma expansão significativa no ambiente empresarial, com empresas começando a explorar suas aplicações para melhorar a eficiência operacional e impulsionar a inovação. Simultaneamente, a expansão da internet proporcionou o aumento de dados disponíveis e possibilitou a conexão entre sistemas, a partir disso, ferramentas de indexação de páginas e sistemas de navegação foram desenvolvidos (BARBOSA; BEZERRA, 2020). Deste modo, a IA passou a ser vista não apenas como uma área de pesquisa, mas como uma ferramenta prática e estratégica para a resolução de problemas empresariais.

Entretanto, como pontuado por Cozman e Neri (2021), a história da IA é marcada por grandes promessas e expectativas - 1987 data então o Segundo Inverno da IA, evento este marcado por diversos favores, como pontua Favaron (2024), sendo eles: o colapso do mercado de computadores especializados em rodar LISP, que eram caros e de difícil manutenção; a obsolescência dos Sistemas Especialistas, que não conseguiam resolver problemas complexos de maneira eficiente e começaram a se mostrar limitados; o fracasso do Projeto de Quinta Geração do Japão, que visava criar máquinas com capacidades cognitivas humanas, mas não gerou avanços concretos; e, reduções significativa nos investimentos no setor, somada à crescente desconfiança de investidores e do público em relação às promessas não cumpridas da área.

O Segundo Inverno da IA tem seu fim dado em 2000 impulsionado por uma combinação de avanços tecnológicos e novos modelos teóricos. O aumento significativo no poder computacional, especialmente com o uso de unidades de processamento gráfico (GPUs), permitiu a realização de cálculos complexos necessários para o desenvolvimento de redes neurais profundas (*Deep Learning*). Além disso, o crescimento exponencial dos dados disponíveis, aliado à infraestrutura escalável de computação em nuvem, facilitou o treinamento de modelos mais robustos e precisos. Pesquisas inovadoras em aprendizado de máquina renovaram o campo de pesquisa, enquanto os investimentos de grandes empresas, e a popularização de aplicativos práticos, como assistentes virtuais e reconhecimento de voz, ajudaram a reverter a desconfiança do público e impulsionar o interesse pela área (FAVARON, 2024). Ainda neste período, em 1997 o computador "*Deep Blue*" ("Azul Profundo" em tradução literal) da IBM venceu uma partida de xadrez contra a campeão mundial Garry Kasparov (RUSSELL; NORVIG, 2013), chamando a atenção de todo o mundo sobre a capacidade da IA.

Entre anos de 2000 e 2013, a IA foi amplamente aplicada em diversas áreas, com impactos significativos em setores como saúde, logística, entretenimento e automação. Além de ser utilizada em veículos autônomos, reconhecimento de voz e planejamento de missões espaciais, a IA passou a ser essencial para a análise de grandes volumes de dados (Big Data), otimizando processos em empresas e melhorando a tomada de decisões. A introdução de assistentes virtuais como Siri, Google Now e Cortana trouxe a IA para o cotidiano das pessoas, permitindo a realização de tarefas diárias através de comandos de voz. No campo da robótica, dispositivos como aspiradores automáticos e robôs de desarmamento começaram a ganhar popularidade, enquanto algoritmos de aprendizado de máquina impulsionaram sistemas de recomendação, personalizando a experiência do usuário em plataformas como Netflix e Amazon (BARBOSA; BEZERRA, 2020), (RUSSELL; NORVIG, 2013). Além disso, a IA foi aplicada em diagnósticos médicos, com ferramentas que ajudaram a analisar imagens e prever doenças, e em sistemas de reconhecimento de imagens e vídeos, contribuindo para avanços em segurança e vigilância. Essas inovações marcaram o início de uma revolução tecnológica, que continuaria a expandir e evoluir nas

décadas seguintes.

No cenário atual IA é amplamente usada em sistemas de recomendação, como os encontrados em plataformas de *streaming* e *e-commerce*, otimizando a experiência do usuário. Outro avanço importante é sua aplicação na criação de conteúdo por meio de modelos generativos com geração de texto e imagem, como ChatGPT e Leonardo AI. A IA também se destaca no campo da automação industrial, logística e planejamento estratégico, otimizando processos e economizando tempo e recursos. No entanto, com seu crescimento, surgem também questões éticas e de segurança, como a privacidade dos dados e a necessidade de regulamentação para evitar usos indevidos da tecnologia (KPMG, 2023).

Smith et al. (2006) descreve que a IA passa por três fases em sua relação com a sociedade: 1) Curiosidade, 2) Desilusão e 3) Confiança. Essas fases refletem o ciclo exposto neste capítulo, que começa com um entusiasmo impulsionado por feitos dados à época como extraordinários, seguido pela perda de popularidade e confiança devido a expectativas não atendidas e relatórios duramente críticos, até a aceitação gradual dos benefícios práticos da tecnologia. Mesmo durante os períodos de "inverno" da IA, a pesquisa na área continuou, evidenciando às aspirações mais antigas da humanidade de criar máquinas capazes de simular a inteligência e as ações humanas. Esse objetivo sobreviveu por séculos, atravessando momentos de grandes avanços e hiatos, até a chegada das atuais IA Gen.

2.1.2 Da Inteligência Artificial Tradicional à IA Generativa

O começo do que é conhecido como "Inteligência Artificial Generativa" remonta, naturalmente, a tópicos mencionados na linha do tempo da IA, como estudos iniciais em Processamento de Linguagem Natural, o desenvolvimento de redes neurais artificiais e aplicações de *machine learning* e *big data*. No entanto, a verdadeira inovação na área de Inteligência Artificial Generativa ocorreu mais tarde, com o desenvolvimento das Redes Generativas Adversárias (GAN) por Ian Goodfellow em 2014 (MACHADO, 2023). As GANs introduziram um novo paradigma para gerar conteúdo, onde duas redes neurais competem entre si para melhorar continuamente os resultados. Outro fator importante para o desenvolvimento da IA Gen foi o desenvolvimento da arquitetura *Transformer* (VASWANI et al., 2017) que revolucionou o campo de PLN e outros campos de aprendizado de máquina –, que, juntamente com os aprimoramentos nos algoritmos de redes neurais profundas, deram origem aos Modelos de Linguagem de Grande Escalas (LLMs) (sigla do inglês *Large Language Models*), os quais revolucionaram o PLN. Os LLMs são modelos de inteligência artificial baseados em redes neurais profundas, projetados para compreender e gerar linguagem humana de forma fluida e contextual, aprendendo padrões, estruturas e semânticas da linguagem a partir de um grande volume de texto (RAMOS, 2023), conforme descrito em seção posterior.

A partir do *Transformer* surge então o chamado Transformador Generativo Pré-treinado (GPT), baseado em aprendizado de máquina pré-treinado, em que o treinamento é feito inicialmente em uma grande quantidade de dados não rotulados para aprender representações gerais e convincentes da linguagem (LIMA; SERRANO, 2024). Dentre os destaques à época, o desenvolvimento de três grandes tecnologias de LLMs tinha o *Transformer* como base comum, sendo eles: O Bert (2018) e o T5 (2019), da *Google* e ChatGPT-3 (2022) da OpenAI (RAMOS, 2023). O funcionamento da arquitetura *Transformer* é detalhado no próximo capítulo deste trabalho.

Após o lançamento e sucesso do ChatGPT em 2022, avanços significativos ocorreram no campo da IA Gen e chamaram a atenção do grande público para as possibilidades da tecnologia dado, principalmente, ao alto nível de conversações de *chatbots*. Nesse período, também se destacou o lançamento do PaLM2[2] pela Google, oferecendo grandes avanços no desempenho de tarefas complexas de linguagem. A DeepMind apresentou o modelo Gato[3], projetado para realizar múltiplas tarefas de forma generalista, como jogar e controlar robôs. Além disso, em 2022, a OpenAI também lançou o DALL-E 2[4], que revolucionou a geração de imagens a partir de texto, permitindo edições específicas nas imagens geradas. O MidJourney[5], também lançado em 2022, se destacou por sua capacidade de criar arte visual de alta qualidade a partir de descrições textuais. Em 2023, a OpenAI lançou o GPT-4[6], uma versão multimodal que ampliou a capacidade de processar tanto texto quanto imagens, aprimorando ainda mais a qualidade das respostas e o contexto das interações. Esses marcos, junto com o Stable Diffusion (2022) e o GitHub Copilot[7] (2021), foram essenciais para expandir as aplicações de IA generativa, impactando áreas como criatividade, programação e design, e consolidando a presença dessas tecnologias no mercado.

As melhorias em técnicas como o uso de modelos pré-treinados têm permitido que soluções de IA se adaptem mais rapidamente a necessidades específicas, como análise de sentimentos ou tradução de idiomas, geração de conteúdo convincente e análises de dados realizados em linguagem natural através de *chatbots*, aumentando a eficácia desses sistemas em ambientes corporativos e criativos. Esses desenvolvimentos estão configurando um futuro em que a IA Gen se torna uma ferramenta ainda mais integrada e essencial nas práticas diárias, impactando desde a criação de conteúdo até a automação e inovação em processos industriais.

A figura 1 apresenta, de maneira resumida e com datas aproximadas, uma linha do tempo da evolução da IA, destacando os principais marcos históricos que contribuíram para o avanço dessa tecnologia até a recente ascensão da IA Gen.

[2] PaLM2 <https://ai.google/discover/palm2/>
[3] Gato <https://deepmind.google/discover/blog/a-generalist-agent/>
[4] DALL-E 2 <https://openai.com/index/dall-e-2/>
[5] MidJourney <https://www.midjourney.com/>
[6] GPT-4 <https://openai.com/index/gpt-4/>
[7] Github Copilot <https://github.com/features/copilot>

Figura 1 – Linha do Tempo da Inteligência Artificial

Fonte: Próprio autor

2.2 A Inteligência Artificial Generativa

Nas subseções a seguir são explorados os conceito de IA Gen, com introdução aos fundamentos que a definem e suas principais características. Em seguida, é abordado o seu funcionamento, detalhando os principais mecanismos e algoritmos que possibilitam a criação de conteúdo, bem como os principais modelos e as metodologias utilizadas para treino. A compreensão dessas questões é essencial para entender o impacto das IA Gen diversas áreas de aplicação e os potenciais riscos e benefícios associados a sua adoção, os quais se debruçam este trabalho.

2.2.1 O Conceito de Inteligência Artificial Generativa

A Inteligência Artificial Generativa (do ingês "*Generative Artificial Intelligence*") é uma subárea da IA que se concentra no desenvolvimento de *softwares* capazes de criar conteúdo novo e original – diferente dos modelos clássicos de IA focados em analisar e interpretar dados – através da entrada (*prompt*) do usuário (WESSEL et al., 2023) e processamento a partir de uma vasta base de dados conhecida como Modelos de Linguagem de Grande Escala (LLM).

Além de gerar textos, essa tecnologia também tem sido aplicada na criação de arte visual e auditiva, conseguindo produzir imagens e até sons que imitam a produção humana. Exemplos de seu uso vão desde obras de arte até a criação de rostos humanos sintéticos, passando por descobertas científicas, como a geração de magnetogramas solares (HUSSAIN, 2023).

O lançamento do ChatGPT intrigou a sociedade como um todo e lançou holofotes sobre as capacidades da IA Gen, gerando discussões sobre seu impacto em diversas áreas,

desde a educação até a indústria. Para Ramos (2023), o sucesso da ferramenta se deu por ser de fácil acesso e uso, uma vez que não requer que o usuário seja especialista em IA para obter resultados satisfatórios e de alta qualidade. Essa simplicidade e eficácia despertaram um interesse generalizado e possibilitaram que um público mais amplo interagisse com a tecnologia de maneira intuitiva. Ainda, a alta adoção dessa tecnologia é diretamente influenciada por fatores como prontidão organizacional, infraestrutura disponível e vantagens observadas durante seu uso (ALI; MUSTAFA; AYSAN, 2024).

Deste modo, sua principal característica é a capacidade de usar informações já existentes para gerar novos conteúdos. Essa habilidade é alimentada por enormes conjuntos de dados e algoritmos sofisticados que permitem à IA identificar padrões, fazer associações e criar material que simula, ou até mesmo expande, a criatividade humana. Com isso, ferramentas de IA generativa têm o potencial de revolucionar diversas áreas, como marketing, educação, design, e até mesmo a produção artística, oferecendo uma nova forma de automação que desafia as fronteiras tradicionais da criatividade e inovação (WESSEL et al., 2023).

Dados estes conceitos, a próxima subseção aborda o funcionamento da IA Gen, com foco nos principais modelos e métodos que possibilitam a geração de novos conteúdos. Serão discutidos os processos de treinamento dessas IAs, que permitem a identificação de padrões e a criação de resultados originais. No entanto, considerando que este trabalho se concentra nos riscos do compartilhamento de dados empresariais, o enfoque será direcionado às tecnologias voltadas para o processamento e geração de texto, sem aprofundar-se em modelos específicos de criação de imagens, áudio ou outras formas de mídia.

2.2.2 Modelos de Linguagem de Grande Escala

Os Modelos de Linguagem de Grande Escala atualmente representam um dos avanços mais significativos no campo da Inteligência Artificial Generativa. Desenvolvidos a partir de técnicas avançadas de aprendizado de máquina, especialmente no PLN, esses modelos utilizam *deep learning* para compreender e gerar textos de forma coerente e contextualizada, aprendendo com vastos volumes de dados textuais obtidos na internet.

O avanço das Redes Neurais Artificiais e a introdução da arquitetura *Transformer* foram marcos fundamentais para a evolução dos LLMs. Diferentemente de abordagens anteriores, que limitavam o contexto analisado, os *Transformers* possibilitaram um processamento mais eficiente e paralelo, permitindo que os modelos identificassem padrões complexos em grandes quantidades de dados.

Atualmente, os LLMs são a base de diversas ferramentas de Inteligência Artificial Generativa, como os modelos da família GPT, sendo amplamente empregados em assistentes de escrita, *chatbots* e sistemas de automação. As próximas subseções abordarão com mais detalhes as principais estruturas, treinamentos e os desafios associados a esses modelos e ao seu uso.

2.2.2.1 *Transformer*: A Principal Arquitetura dos LLMs

A arquitetura *Transformer*, que dá origem ao "T" de *ChatGPT*, marca um avanço significativo no Processamento de Linguagem Natural e no desenvolvimento dos LLM. Esta abordagem, introduzida por Vaswani et al. (2017) substitui, em muitos casos, métodos tradicionais baseados em Rede Neural Recorrente (RNN) e Rede Neural Convolucional (CNN) (MERRITT, 2022), que enfrentam dificuldades no processamento de sequências longas. Sua principal inovação está no mecanismo de *self-attention*, que permite ao modelo avaliar a relevância de cada palavra no contexto geral, independente de sua posição no texto.

Enquanto as RNNs e CNNs processam palavras de forma sequencial e unilateral, os *Transformers* analisam toda a cadeia de palavras simultaneamente, tornando o treinamento mais eficiente e paralelo. Isso é possível graças ao mecanismo de *self-attention*, que calcula e atribui pesos matemáticos de cada palavra em relação a todas as outras dentro da sentença. Essa característica permite capturar grandes relações textuais, resultando em uma compreensão mais precisa.

Além, sua estrutura é composta por camadas de codificadores (*encoders*) e decodificadores (*decoders*), que trabalham em conjunto para gerar linguagem de forma precisa e sofisticada. Neste modelo, os *encoders* recebem o texto de entrada e aplicam múltiplas operações de *self-attention* e normalização, capturando padrões e relações semânticas, e os *decoders* são responsáveis pela geração textual, utilizando as informações processadas pelos *encoders* para prever a próxima palavra levando em consideração o contexto global da sentença.

A capacidade dos *Transformers* em capturar relações de longo alcance em textos e lidar, paralelamente, com grande volume de dados, impulsionou o desenvolvimento dos LLMs modernos, com usos que vão além de simples geradores de texto, como o processamento de cadeias de aminoácidos, conforme descrito por Merritt (2022). Para gerar respostas mais naturais e coerentes, esses modelos passam por um treinamento em larga escala, utilizando bilhões de parâmetros (ESSEL et al., 2024) e sendo expostos a grandes volumes de dados. A próxima subseção abordará o processo de treinamento e as principais técnicas empregadas para otimização de desempenho.

2.2.2.2 Treinamento e Ajuste de LLMs

Pré-treinamento

O desenvolvimento de LLMs ocorre em duas etapas principais: o pré-treinamento e o ajuste fino (*fine-tuning).* O pré-treinamento corresponde à fase inicial do processo, onde o modelo aprende padrões linguísticos a partir de vastos conjuntos de dados, sem a necessidade de supervisão humana direta. Essa etapa é fundamental para que o modelo

adquira conhecimento amplo sobre a linguagem, permitindo-lhe gerar textos coerentes e compreender e responder solicitações de diferentes tipos.

O pré-treinamento é realizado com base em técnicas de aprendizado não supervisionado, em que o modelo analisa grandes volumes de texto para prever palavras ou preencher lacunas dentro de sentenças. Dentre os métodos mais comuns utilizados nesse processo, destacam-se os Modelagem de Linguagem Mascarada (MLM) e o Modelagem de Linguagem Causal (CLM). No MLM, adotado em arquiteturas como o BERT (DEVLIN et al., 2018), parte das palavras de uma sentença é ocultada, e o modelo deve inferir os termos corretos com base no contexto restante. Já o CLM, utilizado em modelos como a família GPT, consiste na previsão da próxima palavra em uma sequência textual considerando apenas palavras anteriores como referência.

Para que um LLM atinja um desempenho satisfatório, seu pré-treinamento exige o processamento de bilhões de palavras (ou *tokens*), abrangendo textos extraídos de diversas fontes, como livros, artigos científicos e páginas da *web*. Desta forma, tais modelos podem reproduzir vieses presentes nos dados e apresentar limitações na coerência de suas respostas. Para mitigar essas questões e adaptar os LLMs à aplicações específicas, é realizada uma segunda etapa, conhecida como ajuste fino ou *fine-tuning*.

Fine-tuning

Após o pré-treinamento, os LLMs passam por uma segunda etapa chamada ajuste fino, na qual são refinados para tarefas específicas e alinhados a objetivos determinados. Diferentemente do pré-treinamento, o ajuste fino normalmente envolve dados rotulados e diretrizes específicas para que o modelo aprenda a responder dentro de um determinado contexto, reduzindo imprecisões e aprimorando sua utilidade prática.

O *fine-tuning* pode ocorrer de diferentes formas, dependendo da aplicação desejada. No ajuste fino supervisionado, o modelo é treinado em um conjunto de dados específico, que contém exemplos de entrada e saída esperados, o que lhe permite se especializar em domínios como atendimento ao cliente, redação técnica ou análise jurídica. Já no aprendizado em contexto (*in-context learning*), o modelo é exposto a poucos exemplos durante a própria interação, sem que haja uma modificação permanente em seus pesos internos, como corre no *fine-tuning* convencional. Esse método permite maior flexibilidade, pois permite ao modelo a adaptação de suas respostas de acordo com as instruções fornecidas no momento da consulta.

Além dessas abordagens, uma técnica amplamente utilizada no ajuste fino moderno é o Reforço com *Feedback* Humano (RLHF) (do inglês "*Reinforcement Learning from Human Feedback*"). Neste método, o modelo recebe avaliações humanas sobre a qualidade de suas respostas e, com base nesse *feedback* ajusta seus parâmetros para gerar saídas mais alinhadas às expectativas do usuário (CHRISTIANO et al., 2023). Essa técnica

foi fundamental no desenvolvimento de modelos como o ChatGPT, permitindo que eles fossem aprimorados para gerar respostas mais naturais e adequadas ao contexto.

O ajuste fino também pode ser realizado com abordagens mais eficientes em termos computacionais, como o Adaptação de Baixa Classificação (LoRA), que diferentemente dos outros modelos citados, permite modificar apenas pequenas porções de parâmetros do modelo base ou invés de reajustar toda a base neural (FERREIRA, 2024). Essa técnica é particularmente útil para empresas que desejam customizar modelos de IA sem a necessidade de treinar do zero redes neurais extremamente complexas.

Apesar das vantagens do *fine-tuning*, sua aplicação também apresenta desafios, como o risco da superespecialização do modelo, que pode perder a generalização adquirida no pré-treinamento ao focar excessivamente em um conjunto restrito de dados. Além disso, há preocupações relacionadas à ética e à segurança, pois ajuste inadequados podem reforçar vieses ou permitir que o modelo seja manipulado para gerar informações prejudiciais.

2.2.3 Outras Arquiteturas de IA Generativa

Embora os LLMs, citados anteriormente, sejam a principal tecnologia voltada para a geração de textos, outras abordagens de IA Gen desempenham um papel fundamental na criação de diferentes tipos de conteúdo. Em particular, as GANs e Modelos de Difusão revolucionaram a geração de imagens, vídeos e síntese de voz, demonstrando aplicações significativas em setores como *design*, entretenimento e pesquisa científica.

As GANs introduzidas por Goodfellow et al. (2014), consistem em um modelo de aprendizado profundo baseado em duas redes neurais que competem entre si: um gera dor, que cria dados sintéticos, e um descriminador, que avalia se os dados gerados são reais ou falsos. Esse processo permite que as GANs produzam imagens (ou outro tipo de conteúdo) com alta fidelidade, sendo amplamente utilizada para criação de rostos humanos sintéticos, restauração de imagens e geração de cenários realistas em ambientes virtuais.

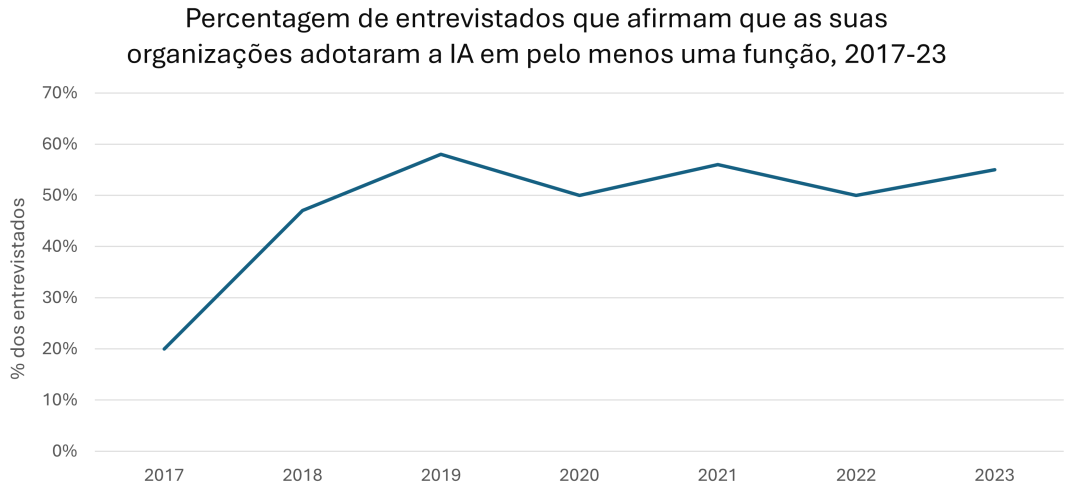
Por outro lado, os Modelos de Difusão representam uma evolução na geração de imagens, baseando-se em um processo probabilístico que transforma ruído aleatório em representações visuais coerentes (HO; JAIN; ABBEEL, 2020). Essa técnica demonstra ótimos resultados na criação de imagens hiper-realistas e tem sido amplamente aplicada em ferramentas como o DALL · E e o Stable Diffusion. Além disso, há pesquisas que exploram sua aplicação na síntese de áudio e geração de dados estruturados.

Apesar do impacto dessas arquiteturas ser diretamente na criação de conteúdo visuais, o compartilhamento de dados com esses modelos também levanta preocupações quanto à segurança e privacidade. Assim como os LLMsm tanto as GANs quanto os Modelos de Difusão dependem de vastos conjuntos de dados para treinamento, o que pode incluir informações sensíveis ou protegidas por direitos autorais. Além, o avança dessas tecnologias te impulsionado a geração de *deepfakes* e outras formas de manipulação digital, ampliando os riscos de desinformação e uso indevido de dados empresariais.

2.2.4 A Adoção da IA Generativa no Ambiente Empresarial

A crescente popularização da IA Gen não se restringe ao meio acadêmico ou à industria tecnológica. Nos últimos anos, empresas de diferentes portes e setores têm explorado essas ferramentas para otimizar processos, aumentar produtividade e aprimorar a experiência do cliente – HAI (2024) apresenta que nos últimos anos, empresas de diferentes portes e setores têm explorado essas ferramentas para otimizar processos, aumentar a produtividade e aprimorar a experiência do cliente, onde 55% das organizações entrevistadas em 2023 afirmaram já utilizar inteligência artificial em pelo menos uma unidade ou função de negócio, em comparação aos 50% de 2022 e apenas 20% em 2017 2. Modelos de linguagem como ChaGPT, por exemplo, vêm sendo utilizando para automatizar atendimentos, redigir relatórios, analisar contratos e documentos e até apoiar a processos criativos, como a geração de conteúdo publicitário.

Figura 2 – Adoção de IA ao longo dos anos



Fonte: AI Index Report 2024, HAI (2024) – adaptado

A busca por inovação e eficiência operacional tem sido o principal motor dessa adoção. Em um cenário de transformação digital constante, organizações veem nas ferramentas de IA Gen uma forma de ganhar vantagem competitiva, reduzindo custos e acelerando decisões. Essa corrida tecnológica, no entanto, muitas vezes ocorre de forma descentralizada, com colaboradores testando ferramentas baseadas em modelos generativos sem que existem diretrizes claras sobre seu uso e proteção de dados.

Grandes empresas de tecnologia, como Google, Microsoft, Meta e OpenAI, têm influenciado diretamente essa tendência ao disponibilizarem modelos generativos por meio de Interface de Programação de Aplicaçõess (APIs), integrações com sistemas corporativos e plataformas acessíveis ao público. Além disso, provedores de soluções empresariais passaram a incorporar funcionalidades baseadas em IA Gen em seus produtos, como revisores de texto, planilhas e ferramentas de produtividade, facilitando ainda mais o uso

por profissionais de diferentes áreas. Um exemplo disso é a Microsoft, que oferece, em um de seus planos, a ferramenta de Inteligência Artificial Generativa "*Copilot*", integrada a produtos consagrados como Excel, Word e PowerPoint.

Ainda que os benefícios sejam amplamente discutidos e vivenciados, a adoção indiscriminada de ferramentas de IA Gen pode trazer implicações significativas para segurança, privacidade e governança de dados, conforme mencionado por KPMG (2023). Nesse sentido, compreender como a tecnologia tem sido introduzida no contexto empresarial é fundamental para refletir sobre seus potenciais riscos e lacunas existentes na regulamentação e nas práticas de Uso. A próxima subseção abordará o papel dos dados nesse cenário, destacando sua relevância como insumo de treinamento de IAs e os desafios associados à sua gestão.

2.2.5 Os Dados como Insumo de Modelos de IA Generativa

O avanço das ferramentas baseadas em IA Gen se apoia, em grande maioria, na disponibilidade massiva de dados. Para que esses modelos sejam capazes de gerar textos, imagens ou análises de alto grau de coerência e relevância, eles precisam ser treinados com conjuntos extensos de informações – vindas de livros, artigos, sites, fóruns e outros repositórios digitais. Quanto maior e mais diversificado o volume de dados, maior tende a ser a capacidade do modelo de compreender contextos, identificar padrões e gerar respostas satisfatórias.

No ambiente empresarial, essa lógica volta à reflexão de até que ponto os dados compartilhados durante o uso dessas ferramentas podem ser incorporados como insumo para seu aperfeiçoamento. Aplicações baseadas em IA Gen ao processar as solicitações dos usuários, podem registrar partes dessas interações – dependendo da política da plataforma – com a finalidade de refinar seus algoritmos e gerar atualizações futuras. Isso significa que, ao utilizar dessas ferramentas para redigir um *e-mail*, analisar contratos ou gerar relatórios, uma empresa pode estar, inadvertidamente, fornecendo dados valiosos ao sistema.

Essas informações, embora aparentemente inofensivas em um primeiro momento, podem conter termos estratégicos, práticas internas, fragmentos de documentos confidenciais ou até mesmo insumos que remetam ao *know-how* da organização. Quando repetidas e variadas interações alimentam os sistemas com esse tipo de conteúdo, abre-se espaço para que o modelo aprenda, armazene e reutilize esse conhecimento de formas não previstas – inclusive em reposta a outros usuários (KPMG, 2023).

Com isso, dados empresariais deixam de ser apenas um recursos operacional e passam a ocupar um lugar central na dinâmica de uso das IA Gens. Entender como essas informações trafegam, são processadas e, eventualmente, reaproveitadas pelos modelos é essencial para que se compreenda os reais impactos da sua adoção.

2.3 Segurança, Governança e Privacidade de Dados

2.3.1 Fundamentos de Segurança de Informação

A segurança da informação consiste no conjunto de práticas, políticas e controles destinados a proteger dados e sistemas contra acessos não autorizados, vazamentos, alterações indevidas e destruição. Sua aplicação não se limita apenas ao ambiente tecnológico, mas envolve também aspectos físicos, organizacionais e humanos que, em conjunto, compõem a base de defesa dos ativos informacionais de uma organização.

Os fundamentos de segurança de informação são organizados em torno dos princípios de confidencialidade, integridade, disponibilidade, autenticidade e não-repúdio, conforme mencionado por Fontes (2017). A confidencialidade visa garantir que somente pessoas autorizadas tenham acesso a determinadas informações. A disponibilidade refere-se à capacidade de garantir que dados e sistemas estejam acessíveis sempre que necessário. Ainda, a autenticidade está relacionada à verificação da identidade de quem acessa o altera dados, enquanto o não-repúdio, por sua vez, assegura que autores de ações não possam negar sua responsabilidade posteriormente.

Esses pilares são fundamentais para a operação segura e confiável de sistemas corporativos, especialmente em ambientes que dependem fortemente do uso de dados para tomada de decisão. Empresas que não adotam medidas básicas de segurança estão, fundamentalmente, mais suscetíveis a incidentes de vazamento de dados sensíveis, fraudes, espionagem industrial e interrupções de serviços críticos.

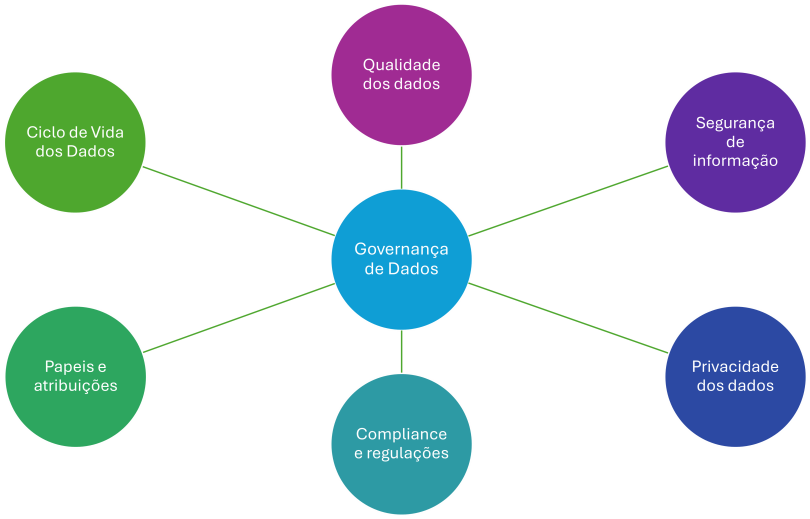
Além de medidas técnicas, como criptografia, autenticação multifator, *backup* e *firewalls*, a segurança da informação exige uma abordagem organizacional robusta, que envolve a criação de políticas internas, treinamentos periódicos com colaboradores, auditorias e definição clara de responsabilidades. A gestão de riscos também é uma parte central dessa abordagem, permitindo que a empresa identifique vulnerabilidades, estime impactos potenciais e adote medidas preventivas.

Em um cenário marcado pela crescente complexidade dos sistemas e pela mudança constante de ameaças digitais, os fundamentos da segurança da informação servem como um guia estratégico para preservação da confiabilidade, confidencialidade e a sustentabilidade dos processos empresariais baseados em dados. Neste contexto, a governança de dados surge não como um conceito isolado, mas como uma aliada essencial às práticas de segurança – atuando de forma integrada para garantir que, além de protegidos, os dados sejam gerenciados com transparência, responsabilidade e alinhamento aos objetivos estratégicos da organização.

2.3.2 O Conceito de Governança de Dados

Com o avanço da transformação digital e o crescimento exponencial da geração de dados, a governança de dados tornou-se um pilar estratégico para organizações. Com a intensificação da transformação digital e o crescimento exponencial na geração de dados, as organizações passaram a reconhecer a governança de dados como uma necessidade estrutural para garantir o uso inteligente e responsável dessas informações. Mais do que um conjunto de boas práticas, a governança atua como um sistema de diretrizes e tomadas de decisão que organiza, padroniza e orienta o ciclo de vida dos dados dentro da empresa. A The Data Governance Institute (2025) descreve a governança como "(...) O exercício de tomada de decisão e autoridade para assuntos relacionados a dados." (em tradução livre). Essa abordagem diz respeito ao conjunto de práticas, políticas, processos e estruturas organizacionais que visam assegurar que os dados corporativos seja gerenciados de forma eficiente, segura e conforme os objetivos do negócio ou instituição (Governo Digital, 2024), conforme ilustrado na figura 3.

Figura 3 – Principais aspectos que compõem a governança de dados



Fonte: Próprio autor

Deste modo, a governança de dados envolve tomadas de decisão e o exercício de autoridade sobre os dados. As práticas e processos dizem respeito à garantia da qualidade dos dados e à gestão de seu ciclo de vida. As políticas englobam as normas internas e conformidade com legislações aplicáveis, como a LGPD. A segurança da informação e a proteção da privacidade de dados asseguram a confidencialidade e integridade dos ativos informacionais. Por fim, a estrutura organizacional define papéis e responsabilidades, estabelecendo quem é o responsável pela gestão, controle e uso adequado dos dados dentro da organização.

A importância da governança vai além do aspecto técnico. Em um ambiente empresarial altamente competitivo e regulamentado, decisões fundamentadas em dados confiáveis

podem representar um diferencial significativo. Ao mesmo tempo, falhas na gestão e controle de dados podem gerar riscos operacionais, perdas financeiras, sanções legais e dados à reputação. Desta forma, a governança de dados atua como uma interface entre os ativos informacionais, isto é, ferramentas onde transitam informações, e os objetivos estratégicos da instituição, criando condições para que os dados sejam utilizados de maneira ética, responsável e alinhada aos marcos regulatórios internos ou públicos.

Ainda, a governança de dados vai além de um simples conceito teórico, pois envolve a definição clara de papéis e responsabilidades dentro da organização. Entre os principais papéis, desatacam-se o *Data Owner*, responsável por tomar decisões estratégicas sobre os dados, o *Data Steward*, encarregado de garantir que os dados sejam gerenciados de acordo com as políticas estabelecidas e com qualidade, e o *Data Custodian*, que se ocupa da manutenção física dos dados, incluindo o armazenamento e a segurança técnica (Anmut, 2024). Quando se trata de IA Gen, a ausência de uma definição clara de responsabilidades pode agravar riscos significativos, como vazamento de dados ou uso indevido de informações sensíveis, uma vez que dados envolvidos frequentemente transitam por múltiplos sistemas e interações.

A aplicação de governança, no entanto, não é idêntica entre as empresas. Sua maturidade e arquitetura estão condicionadas a diversos de fatores, como cultura organizacional, a disponibilidade de recursos, o nível de conhecimento técnico, as exigências legais e apoio da alta gestão. Marcondes (2021) apresenta em seu estudo alguns *frameworks* consolidados, como o DAMA-DMBOK, desenvolvido pela *Data Management Association* (DAMA), e modelo proposto pelo *Data Governance Institute* (DGI), ambos oferecendo abordagens estruturadas para a implementação e gestão da governança de dados.

Com a popularização da Inteligência Artificial Generativa, o conceito de governança de dados passa a ser ainda mais desafiador – isso porque modelos generativos demandam de grandes volume de dados como insumo para seu funcionamento, muitas vezes operando em ambientes descentralizados, com múltiplos atores internos e externos operando e interagindo com os sistemas. Ainda, nem sempre há transparência quanto ao que é feito com os dados introduzidos em ferramentas baseadas em IA, o que pode acender alertas sobre o controle, rastreabilidade e proteção das informações inseridas.

Singla et al. (2025) aborda o aumento do gerenciamento de riscos relacionados ao uso de Inteligência Artificial Generativa em comparação ao último ano. Entre as categorias descritas no trabalho, destacam-se aquelas com maior aumento percentual na busca de mitigação: “imprecisão”, “cibersegurança” e “violação de propriedade intelectual”. Essas categorias refletem as áreas onde as organizações têm centrado seus esforços para reduzir os riscos associados à IA Gen, evidenciando uma crescente preocupação com a precisão dos resultados, a segurança contra ameaças e a proteção de *know-how*, principalmente em empresas maiores.

Portanto, compreender os fundamentos da governança de dados é essencial para que

empresas preparem não apenas para atender às exigências legais e normativas (como a LGPD), mas também para lidar com transformações impostas por tecnologias futuras.

2.3.3 Privacidade de Dados no Contexto Empresarial

A privacidade de dados é um dos pilares centrais quando se discute segurança da informação e governança de dados no ambiente corporativo. Em um contexto onde ferramentas baseadas em inteligência artificial generativa têm ganhado espaço nas rotinas empresariais, cresce também a preocupação com a exposição inadvertida de informações sensíveis. Isso inclui não apenas dados pessoais protegidos por legislações como a LGPD do Brasil, mas também dados estratégicos de uma organização, como segredos industriais, processos internos, dados operacionais e outros ativos que, se compartilhados com ferramentas externas, podem comprometer a competitividade ou integridade da empresa.

Diferentemente de outras tecnologias, os modelos generativos frequentemente operam com base de dados extensas e, muitas vezes, de natureza opaca – ou seja, o usuário final nem sempre sabe como, onde e por quanto tempo os dados inseridos são armazenados ou utilizados, conforme descrito pela BBC News Mundo (2023). Nesse cenário, a linha que separa o uso legítimo de dados e sua exposição indevida torna-se tênue, ampliando os risos associados à privacidade.

A privacidade, portanto, deixa de ser apenas um requisito legal e passa a representar uma medida de proteção estratégica. Estabelecer limites claros sobre o que pode ou não ser compartilhado com soluções de IA Gen faz-se essencial. Além, promover uma cultura organizacional orientada à proteção da informação ajuda a reduzir comportamentos de risco, como o compartilhamento de dados confidenciais em ferramentas sem o devido respaldo técnico ou jurídico, conforme descrito por Lovatto (2024).

2.3.4 Riscos Mapeados em Diretrizes Técnicas

O uso de Inteligência Artificial Generativa no ambiente corporativo vem despertando crescente atenção de organizações, especialistas e órgãos reguladores enquanto aos riscos associados à segurança da informação, à privacidade de dados e à conformidade legal. Ainda que essas ferramentas ofereçam benefícios como automação, aumento de produtividade e suporte à criatividade, elas também impõem desafios substanciais relacionados à governança dos dados compartilhados consigo.

Várias instituições vêm desenvolvendo estudos técnicos, diretrizes e alertas sobre os perigos inerentes ao uso da IA, com foco especial nos modelos generativos por operarem, muitas vezes, de maneira obscura. Dentre os documentos de destaque neste cenário, estão os órgãos reconhecidos internacionalmente como a ENISA, a OECD e o NIST. Cada uma dessas entidades contribui com perspectivas complementares sobre os riscos mais relevantes observados na adoção de tecnologias baseadas em IA.

O relatório Threat Landscape for Artificial Intelligence, publicado pela Agência da União Europeia para a Cibersegurança (ENISA) (ENISA, 2024), apresenta um mapeamento abrangente das ameaças relacionadas ao uso de Inteligência Artificial. As ameaças são categorizadas em diferentes frentes, incluindo vulnerabilidades inerentes aos modelos, ataques cibernéticos direcionados, falhas nos processos de integração e uso indevido das funcionalidades da IA. A publicação também ressalta os esforços de monitoramento contínuo do ecossistema europeu de cibersegurança aplicado à IA, conduzidos com o apoio do Grupo de Trabalho Ad-Hoc em Cibersegurança para Inteligência Artificial.

O relatório da ENISA destaca ainda a importância da construção de um ecossistema seguro e confiável para IA, cobrindo toda a cadeia de suprimentos — desde os dados de entrada até os modelos e sistemas finais. Nesse contexto, são enfatizadas ações voltadas à proteção de dados, à segurança cibernética e à promoção de inovação, qualificação profissional e desenvolvimento científico.

No caso da IA Gen, o relatório aponta riscos específicos, como a inserção involuntária de dados sensíveis por parte dos usuários, que podem ser absorvidos pelos modelos sem garantias de anonimização ou descarte. Além disso, alerta-se para ameaças como o data poisoning, em que dados manipulados são inseridos maliciosamente no sistema, e o risco de vazamentos involuntários de informações, sobretudo em ferramentas que mantêm histórico de interações.

A OECD tem se destacado como uma das principais instituições internacionais na proposição de diretrizes para o uso responsável da IA. Entre suas iniciativas, destaca-se o relatório (Organisation for Economic Co-operation and Development (OECD), 2025) que discute a convergência entre os Princípios de IA da OECD e suas diretrizes de privacidade, com o objetivo de alinhar estratégias que promovam tanto a inovação tecnológica quanto a proteção de dados pessoais. O documento apresenta um mapeamento de abordagens regulatórias adotadas por diferentes países e propõe oportunidades de colaboração internacional para harmonizar políticas públicas voltadas à IA. Um dos focos centrais do relatório está na necessidade de maior integração entre governança de dados e desenvolvimento de sistemas inteligentes, a fim de garantir que os direitos fundamentais dos usuários sejam respeitados ao longo de todo o ciclo de vida dos modelos.

Ainda, a OECD reforça a importância de princípios como transparência, supervisão humana, robustez e responsabilização no uso de dados por sistemas de IA, chamando atenção especial para riscos característicos da IA Gen, como a falta de clareza sobre o tratamento das informações inseridas e a dificuldade de rastreamento em casos de uso indevido. Além disso, destaca-se o papel dos reguladores na criação de ambientes normativos que estimulem a inovação sem comprometer a segurança e a privacidade dos indivíduos.

O NIST elaborou o *AI Risk Management Framework* (AI RMF) (STANDARDS; (NIST), 2024), ferramenta com o objetivo de apoiar organizações no gerenciamento res-

ponsável dos riscos associados a sistemas de Inteligência Artificial. O documento apresenta diretrizes que visam tornar a IA mais confiável, com foco em aspectos como segurança, privacidade, explicabilidade, equidade e robustez técnica.

O desenvolvimento do *framework* foi conduzido de forma aberta e colaborativa, contando com contribuições de organizações de diferentes setores – incluindo governo, academia, setor privado e sociedade civil. A estrutura proposta adota um ciclo contínuo de avaliação e mitigação de riscos, encorajando as instituições a implementarem políticas claras para o tratamento de dados, com mecanismos de revisão periódica, a fim de evitar o uso indevido de informações sensíveis.

O AI RMF enfatiza ainda a necessidade de aplicação de controles técnicos e administrativos que reduzam a exposição a vazamentos de dados e minimizem impactos legais, regulatórios e de compliance. Sua arquitetura está organizada em quatro funções principais – Governar, Mapear, Mensurar e Gerenciar –, que permitem o monitoramento contínuo do ciclo de vida de sistemas de IA.

Ao comparar os documentos das três instituições, nota-se uma forte convergência em torno de cinco categorias de riscos principais:

- **Coleta e retenção indevida de dados**: usuários podem inserir, voluntariamente ou não, dados sensíveis em ferramentas de IA Gen, os quais podem ser coletados, armazenados e reutilizados sem consentimento explícito;

- **Falta de transparência sobre o uso de dados**: muitas ferramentas não deixam claro como as informações fornecidas nos prompts são tratadas, por quanto tempo ficam retidas, e se são utilizadas para o re-treinamento de modelos;

- **Compartilhamento com terceiros**: políticas pouco detalhadas ou permissivas podem permitir que dados corporativos sejam repassados a outros sistemas, entidades ou prestadores de serviço vinculados às empresas desenvolvedoras;

- **Ausência de mecanismos de controle e exclusão**: não são garantidas ao usuário final formas claras de deletar ou auditar o uso posterior dos dados inseridos;

- **Conformidade legal limitada**: embora muitas ferramentas operem globalmente, suas políticas nem sempre estão alinhadas à legislação local de proteção de dados, como a LGPD no Brasil ou o Regulamento Geral sobre a Proteção de Dados (União Europeia) (RGPD) na União Europeia.

Esses riscos, embora também presentes em outras tecnologias, são potencializados na IA Gen devido ao seu funcionamento autônomo, à escala de uso e à dificuldade de rastreamento das operações internas dos modelos. Diante desse cenário, organizações que desejam adotar ferramentas de IA Gen devem considerar essas categoriais de risco como ponto de partida para a construção de uma governança de dados sólida e bem desenhada.

Tais medidas são fundamentais para garantir a integridade dos ativos informacionais, mitigar vulnerabilidades e assegurar conformidade com os marcos legais em vigor.

Capítulo 3

Experimentos e Análise dos Resultados

Este capítulo apresenta o método adotado para avaliar a hipótese central do trabalho, bem como os procedimentos realizados para coleta e análise de dados. A intenção é compreender de que forma o compartilhamento de informações empresariais com ferramentas de Inteligência Artificial Generativa pode representar riscos à segurança da informação, à privacidade e à governança de dados corporativos.

3.1 Método para a Avaliação

Este trabalho adota uma abordagem qualitativa e exploratória, com base em análise de conteúdo proposta por Bardin (1977). A análise foi aplicada sobre documentos públicos oficiais de ferramentas de IA Gen gratuitas, como termos de uso, políticas de privacidade e guias técnicos. Além disso, foram utilizadas diretrizes de segurança da informação e governança de dados de órgãos internacionais, como ENISA, NIST e OECD, como quadro referencial normativo.

A metodologia de Análise de Conteúdo, conforme proposta por Bardin (1977), constitui um conjunto de técnicas sistemáticas de descrição e interpretação de conteúdos, com o objetivo de obter indicadores que permitam a inferência de conhecimentos relativos às condições de produção e recepção das mensagens. Essa abordagem é utilizada em pesquisas qualitativas, especialmente quando se busca analisar documentos, discursos ou textos a partir de categorias temáticas.

Segundo a autora, a Análise de Conteúdo desenvolve-se em três fases principais: a pré-análise, a exploração do material e o tratamento dos resultados, inferência e interpretação. Essas fases são interdependentes e visam garantir o rigor metodológico na categorização e interpretação dos dados. A pré-análise corresponde ao momento de organização do material. Na fase de exploração, realiza-se a codificação dos dados, identificando unidades de registro e contexto, que serão classificadas segundo categorias previamente definidas ou emergentes. Por fim, o tratamento dos resultados envolve a análise interpretativa dos dados categorizados, permitindo inferências fundamentadas no material analisado.

Neste trabalho, a aplicação da Análise de Conteúdo seguiu as etapas fundamentais descritas pela autora, adaptadas ao contexto da pesquisa:

1. Pré-análise: Consistiu na leitura flutuante e exploratória dos documentos coletados, visando a familiarização com os conteúdos e a identificação preliminar de temas recorrentes relacionados à segurança, privacidade e governança de dados em ferramentas de IA Gen;

2. Exploração do material (codificação): Foram delimitadas as unidades de contexto — trechos significativos como frases ou parágrafos — que tratavam diretamente do tratamento de dados, além de unidades de registro, dadas como termos-chave de maior relevância no *corpus* analisado. Esses fragmentos foram classificados a partir de categorias temáticas formuladas com base nos referenciais normativos (ENISA, NIST, OECD) e na fundamentação teórica da pesquisa.

3. Tratamento dos resultados, inferência e interpretação: Após a organização das unidades em categorias, realizou-se o agrupamento dessas categorias em eixos analíticos mais amplos (segurança da informação, privacidade de dados e governança). A análise interpretativa final permitiu comparar as práticas descritas nos documentos com as diretrizes técnicas estabelecidas, gerando uma avaliação do grau de aderência de cada ferramenta às normas analisadas.

Essa estrutura metodológica permite uma análise comparativa consistente entre diferentes ferramentas, revelando padrões, fragilidades e boas práticas no tratamento de dados sensíveis em contextos de uso de Inteligência Artificial Generativa.

3.2 Experimentos

A presente seção descreve a aplicação da metodologia proposta com foco na interpretação sistemática de documentos públicos de ferramentas de IA Gen gratuitas. O objetivo foi identificar a forma como essas ferramentas lidam com dados dos usuários, e compará-las a diretrizes técnicas sobre segurança, privacidade e governança da informação. A análise, portanto, foi conduzida seguindo sete etapas principais, conforme descrito no seção de método, apresentadas a seguir.

3.2.1 Leitura Geral dos Documentos

Inicialmente foi realizada uma leitura exploratória das políticas de privacidade das seguintes ferramentas: ChatGPT, Copilot (Microsoft) e Gemini (Google). A seleção considerou a popularidade das ferramentas e a disponibilidade pública de suas documentações. A lista completa dos documentos analisados é apresentada no Apêndice A.

3.2.2 Codificação Inicial e Formulação das Categorias

Com base na leitura geral e na literatura técnica (ENISA, NIST e OECD), foram definidos os principais eixos de análise, que se tornaram as categorias temáticas: coleta de dados; uso de informações; compartilhamento de dados com terceiros; medidas de segurança; conformidade com legislações de proteção de dados.

Cada uma dessas categorias foi fundamentada por princípios e recomendações extraídos das diretrizes técnicas e da LGPD, conforme descrito no referencial teórico e evidenciados no Apêndice B.

3.2.3 Recorte em Unidades de Registro e Contexto

Foram então selecionados trechos dos documentos analisados (frases, parágrafos, expressões) que mencionavam práticas relacionadas às categorias definidas. Cada trecho foi agrupado e tratado como uma unidade de registro, isto é, o segmento textual mínimo que contém uma informação relevante para os objetivos da pesquisa. Já a unidade de contexto corresponde a um trecho mais amplo (como o parágrafo completo ou a seção do documento) que permite compreender melhor o significado da unidade de registro, oferecendo os elementos necessários para interpretar adequadamente o conteúdo.

Cada unidade de registro e seu respectivo contexto foram associados à categoria correspondente, conforme mostra o Apêndice C. Deste modo, as unidades de registro foram agrupadas segundo sua correspondência temática, formando subconjuntos homogêneos de dados por categoria. A estruturação seguiu os princípios da exclusão mútua entre categorias, homogeneidade interna e clareza semântica.

3.2.4 Construção das Categorias Finais e Eixos Temáticos

As cinco categorias iniciais foram organizadas sob três eixos maiores de análise, correspondente aos objetivos de estudo: segurança de informação; privacidade de dados; e, governança de dados;

Esse agrupamento permitiu uma análise integrada das práticas observadas em cada ferramenta e a comparação com os pilares das diretrizes técnicas, conforme representado na tabela 1.

Tabela 1 – Eixos Temáticos

Eixo Temático	**Categorias Associadas**
Segurança da Informação	Medidas de Segurança
Privacidade de Dados	Coleta de Dados; Uso das Informações; Compartilhamento com Terceiros
Governança de Dados	Conformidade Legal

3.2.5 Critérios de Avaliação

Para possibilitar uma análise estruturada das ferramentas de IA Gen, foram definidos critérios que contemplam aspectos fundamentais relacionados à segurança da informação, privacidade e governança de dados. Esse critérios foram organizados frente aos cinco pilares definidos anteriormente: coleta de dados, uso dos *prompts*, compartilhamento, segurança e conformidade legal, conforme detalhado na tabela 2. Cada ferramenta foi avaliada a partir desses parâmetros, buscando identificar boas práticas, possíveis lacunas e riscos associados ao uso dessas soluções.

Tabela 2 – Critérios de Avaliação

Pilar	**Requisitos**
Coleta de Dados	Informações claras sobre quais dados são coletados, com opção de consentimento ou controle pelo usuário.
Uso dos *prompts*	Indicação de que os *prompts* não são usados para treinamento por padrão, ou opção clara de exclusão.
Compartilhamento	Declaração explícita de que os dados não serão compartilhados com terceiros, ou só mediante consentimento.
Segurança	Citação de mecanismos como criptografia, acesso restrito, anonimização, etc.
Conformidade Legal	Referência direta à LGPD ou outras normas, com explicação de adequação.

3.3 Interpretação Resultados

Esta seção apresenta a análise dos resultados obtidos a partir das unidades de registro e contexto coletadas nos documentos oficiais das ferramentas ChatGPT, Copilot e Gemini. A avaliação foi conduzida com base nas três categorias definidas previamente: Segurança da Informação, Privacidade de Dados e Governança de Dados, considerando critérios objetivos relacionados à coleta, uso e compartilhamento de dados, além da observância às normas legais aplicáveis.

Coleta de Dados

Todas as ferramentas analisadas deixam clara a amplitude da coleta de dados. ChatGPT, Copilot e Gemini especificam que coletam informações diretamente fornecidas pelo usuário (como dados de conta, conteúdo inserido e interações), além de dados coletados automaticamente (como IP, localização aproximada, dispositivo, entre outros).

Entretanto, o requisito de oferecer opções de consentimento ou controle sobre a coleta não é amplamente atendido de forma evidente nas unidades de registro. Há menção de que

o usuário pode recusar o fornecimento de dados, mas isso pode impedir o funcionamento de alguns serviços.

Sendo assim, as ferramentas demonstram certo grau de transparência na descrição dos dados coletados, mas são vagas quanto à autonomia do usuário sobre o controle da coleta, falhando parcialmente no requisito.

Uso dos *Prompts*

Embora este pilar exija a indicação clara de que os *prompts* não são usados para treinamento por padrão – ou que haja uma forma de exclusão clara, nenhuma das ferramentas demonstra cumprir integralmente esse requisito nos trechos analisados.

O ChatGPT afirma explicitamente que os conteúdos fornecidos pelos usuários podem ser usados para treinar os modelos. O Copilot apresenta um enunciado semelhante, ao afirmar que pode usar os dados para treinar os modelos de IA e serviços e ferramentas da Microsoft, inclusive com análise manual. O Gemini também confirma o uso dos dados para aprimorar tecnologias de aprendizado de máquina e serviços da Google.

Todas as ferramentas sugerem uso dos prompts para fins de treinamento, e indicam, de forma discreta, a possibilidade de exclusão ou controle por parte do usuário, o que representa baixa aderência ao pilar.

Compartilhamento

Neste quesito, o requisito é que os dados não sejam compartilhados com terceiros, salvo mediante consentimento explícito.

Copilot e Gemini destacam que o compartilhamento ocorre com consentimento ou quando necessário para a prestação de serviços, o que sugere um nível razoável de conformidade com o requisito. Gemini, em especial, é a única a mencionar explicitamente que solicita consentimento explícito para o compartilhamento de dados sensíveis.

O ChatGPT, por sua vez, cita diversas possibilidades de compartilhamento com prestadores de serviços, afiliadas, autoridades legais e terceiros em situações específicas (como falência ou fusão), mas não há menção clara sobre exigência de consentimento prévio do usuário em todos os casos, o que enfraquece a transparência nesse aspecto.

Neste ponto, o Gemini apresenta melhor alinhamento ao requisito, ao passo que Copilot está parcialmente aderente. Já o ChatGPT revela maior fragilidade em garantir o controle do usuário sobre o compartilhamento de seus dados.

Segurança

Todas as ferramentas descrevem o uso de mecanismos técnicos e organizacionais para proteção dos dados, cumprindo, em boa medida, o requisito deste pilar.

O ChatGPT menciona medidas comerciais razoáveis, mas também transfere parte da responsabilidade ao usuário ao alertar sobre os riscos de compartilhamento de dados, indo de encontro à nota emitida pela OpenAI após o episódio de vazamento de dados, onde afirma que o vazamento se deu por conta de dispositivos infectados e não uma violação da ferramenta (PIRES, 2023). Copilot apresenta detalhes técnicos como criptografia, desidentificação de dados e sistemas com acesso controlado. O Gemini se destaca por listar práticas robustas: criptografia, autenticação em duas etapas, restrição de acesso a funcionários e medidas físicas de proteção.

Assim, as três ferramentas demonstram boa aderência ao pilar de segurança, com destaque positivo para Gemini, que apresenta a abordagem mais completa e detalhada. Cabe destacar ainda que Gemini e Copilot utilizam os mesmos mecanismos de segurança de suas desenvolvedoras no que tange o controle de acessos.

Conformidade Legal

Este pilar exige referência direta à LGPD (ou outras leis equivalentes) e explicação da adequação legal.

O ChatGPT aborda a transferência internacional de dados e afirma seguir mecanismos legais válidos, ainda que não mencione explicitamente a LGPD. O Copilot, embora com trecho cortado, declara cumprir leis de proteção de dados aplicáveis, o que provavelmente inclui legislações como RGPD e LGPD. Gemini apresenta nas unidades de registro analisadas citação à legislação vigente ou à conformidade jurídica, assegurando controle sobre o tratamento das informações.

Todas as ferramentas demonstram um bom grau de aderência ao pilar, mesmo que sem detalhamento e aderência explícita à LGPD.

3.4 Avaliação dos Resultados

A análise comparativa dos documentos investigados permitiu identificar um panorama quanto à aderência às boas práticas de segurança da informação, privacidade e governança de dados. Observou-se que, embora todas as ferramentas apresentem algum grau de comprometimento com a proteção de dados, existem pontos de atenção importantes, como uso de *prompts* do usuário para treino do modelo e a ausência de referências específicas à legislação brasileira (LGPD), o que pode representar riscos adicionais para o uso dessas ferramentas em ambientes corporativos.

Capítulo 4

Conclusão

Este trabalho teve como objetivo principal analisar os riscos associados ao compartilhamento de dados empresariais com ferramentas de Inteligência Artificial Generativa, com foco nos aspectos de segurança da informação, privacidade e governança de dados. A partir da análise de documentos públicos de ferramentas amplamente utilizadas – como ChatGPT, Copilot e Gemini – foi possível observar como essas soluções tratam os dados dos usuários e em que medida suas práticas se alinham às diretrizes técnicas internacionais e à legislação brasileira.

Os resultados obtidos indicam que, embora existam políticas de privacidade e termos de uso detalhados, ainda há pontos de atenção importantes no que diz respeito ao compartilhamento de dados organizacionais com as ferramentas. O uso dos dados para fins de melhoria dos modelos e personalização, mesmo que justificado pelas empresas, levanta preocupações adicionais quando se trata de ambientes corporativos, onde os dados manipulados muitas vezes envolvem informações sensíveis, estratégicas ou sigilosas – o que vai de encontro com a hipótese deste trabalho.

Entre os pilares analisados, as ferramentas demonstraram níveis similares de aderência. As medidas de segurança e a governança em geral são mais bem descritas, enquanto os aspectos ligados à conformidade com a LGPD apresentaram os maiores pontos de atenção. Nenhuma das ferramentas mencionou diretamente a legislação brasileira, o que vai de encontro a hipótese deste trabalho de que, mesmo com marcos regulatórios já estabelecidos, ainda há pouco alinhamento com o contexto normativo local por parte dessas tecnologias, ao menos no que se refere à citação clara.

Ao refletir sobre o avanço recente da IA, é possível associar os resultados deste estudo à ideia dos chamados "invernos da IA" – momentos históricos em que o desenvolvimento da área sofreu desaceleração por conta de limitações técnicas, falta de aplicabilidade prática ou desconfiança social. A análise conduzida aqui sugere que um novo período de resistência pode surgir não por falhas tecnológicas, mas por ausência de confiança, de transparência e de controle sobre o uso de dados. Se as organizações não souberem como seus dados estão sendo utilizados por essas ferramentas, ou se enfrentarem vazamentos, a

consequência pode ser uma postura mais cautelosa, retração no uso e, consequentemente, desaceleração do entusiasmo em torno da Inteligência Artificial Generativa.

Portanto, este trabalho reforça a importância de iniciativas de governança de dados voltadas especificamente para o uso de IA Gen no contexto organizacional. Mais do que uma tendência tecnológica, essas ferramentas representam um novo modelo de relação entre humanos, máquinas e informação – e, por isso, exigem atenção redobrada das empresas quanto à proteção de seus ativos digitais.

Como contribuição prática, esta pesquisa apresentou uma matriz de avaliação baseada em diretrizes consolidadas, que pode servir como instrumento de apoio para profissionais de tecnologia, segurança da informação e *compliance*. Embora limitada à análise documental, a abordagem utilizada oferece uma base sólida para reflexões futuras e destaca a necessidade de novos estudos que envolvam também a percepção dos usuários e o contexto prático de uso dentro das empresas.

Por fim, espera-se que este trabalho contribua com o debate sobre o uso responsável de IA Gen e incentive a construção de políticas mais claras, seguras e alinhadas à realidade prática e ética das organizações.

4.1 Principais Contribuições

Este trabalho apresenta como principais contribuições a sistematização dos riscos associados inerentes ao uso de ferramentas de IA Gen não se restringindo ao contexto empresarial e a síntese de recomendações que visam mitigar esses riscos. Sendo assim, contribui não apenas para a compreensão de riscos existentes, mas também abre discussão sobre a necessidade de um posicionamento mais claro e estruturado por parte das organizações em relação ao uso seguro dessas tecnologias.

4.2 Trabalhos Futuros

Durante a realização deste trabalho foram identificadas algumas limitações, bem como oportunidades para investigações futuras: um aspecto que merece maior aprofundamento é a realização de estudos de caso em organizações que já implementaram políticas específicas para o uso de IA Gen, com o objetivo de avaliar a efetividade dessas medidas na prática. Além disso, a pesquisa sobre a construção de *frameworks* de governança voltados para IA Gen representa um campo promissor, especialmente diante da rápida evolução dessas tecnologias e das constantes mudanças no cenário regulatório e de segurança. Outro ponto relevante para futuras investigações é o estudo de projetos de Inteligência Artificial Generativa proprietários, focando na maneira como as empresas estão desenvolvendo suas soluções internas para mitigar riscos relacionados à privacidade, segurança da informação e conformidade legal.

Investigar essas iniciativas pode fornecer insights valiosos sobre práticas mais seguras e eficazes no uso de IA Gen em ambientes corporativos, ajudando a moldar as melhores abordagens para o futuro do setor.

Referências

ALI, H.; MUSTAFA, A. ul; AYSAN, A. F. Global adoption of generative ai: What matters most? **Journal of Economy and Technology**, 2024. ISSN 2949-9488. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2949948824000520>. Citado na página 26.

Anmut. **Data Governance Roles Explained: Data Owner, Steward & Custodian**. 2024. Disponível em: <https://www.anmut.co.uk/data-governance-roles-and-responsibilities/>. Citado na página 34.

BARBOSA, X. de C.; BEZERRA, R. F. **Breve introdução à história da inteligência artificial**. Rio Branco: [s.n.], 2020. Acesso em: 15 nov. 2024. Disponível em: <https://periodicos.ufac.br/index.php/jamaxi/article/view/4730/2695>. Citado 4 vezes nas páginas 17, 18, 21 e 22.

BARDIN, L. **L'analyse de contenu**. [S.l.]: Presses universitaires de France Paris, 1977. v. 69. Citado 2 vezes nas páginas 39 e 56.

BAZZAN, A. L. C. et al. **"A Nova Eletricidade: Aplicações, Riscos e Tendências da IA Moderna – "The New Electricity": Applications, Risks, and Trends in Current AI**. 2023. Disponível em: <https://arxiv.org/abs/2310.18324>. Citado 4 vezes nas páginas 13, 17, 18 e 19.

BBC News Mundo. **O que é a misteriosa 'caixa preta' da inteligência artifical que preocupa os especialistas**. 2023. Disponível em: <https://www.bbc.com/portuguese/articles/c870xmd2dv0o>. Citado na página 35.

CHRISTIANO, P. et al. **Deep reinforcement learning from human preferences**. 2023. Disponível em: <https://arxiv.org/abs/1706.03741>. Citado na página 28.

COSTA, W. S.; SILVA, S. C. M. Aquisição de conhecimento: O grande desafio na concepção de sistemas especialistas. **Holos**, v. 2, p. 37–46, 2005. Citado na página 20.

COZMAN, F. G. Inteligência artificial: Uma utopia, uma distopia. **TECCOGS – Revista Digital de Tecnologias Cognitivas**, Programa de Pós-graduação em Tecnologias da Inteligência e Design Digital (TIDD), PUC-SP, n. 17, p. 32–43, Jan-Jun 2018. ISSN 1984-3585. Citado 2 vezes nas páginas 12 e 20.

COZMAN, F. G.; NERI, H. O que, afinal, é inteligência artificial? In: **Inteligência Artificial: Avanços e Tendências**. [S.l.]: Universidade de São Paulo, 2021. p. 19–29. Citado na página 22.

DEVLIN, J. et al. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. 2018. Disponível em: <https://arxiv.org/abs/1810.04805>. Citado na página 28.

ENISA. **Threat Landscape 2024**. 2024. Disponível em: <https://www.enisa.europa.eu/sites/default/files/2024-11/ENISA%20Threat%20Landscape%202024_0.pdf>. Citado 2 vezes nas páginas 36 e 55.

ESSEL, H. B. et al. Chatgpt effects on cognitive skills of undergraduate students: Receiving instant responses from ai-based conversational large language models (llms). **Computers and Education: Artificial Intelligence**, v. 6, p. 100198, 2024. ISSN 2666-920X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666920X23000772>. Citado na página 27.

FAVARON, G. **Os Invernos da IA: Ciclos de Ascensão e Queda na História da Inteligência Artificial**. 2024. Acesso em: 15 nov. 2023. Disponível em: <https://www.guilhermefavaron.com.br/post/os-invernos-da-ia-ciclos-de-ascensao-e-queda-na-historia-da-inteligencia-artificial>. Citado na página 22.

FERREIRA, A. L. C. A. **Fine-tuning de LLMs para geração de código Mojo**. Trabalho de Conclusão de Curso — Universidade Federal de Pernambuco, Recife, agosto 2024. Disponível em: <https://repositorio.ufpe.br/handle/123456789/57488>. Citado na página 29.

FONTES, E. L. G. Segurança da informação. **Saraiva Educação SA**, p. 2, 2017. Citado 2 vezes nas páginas 32 e 55.

GOODFELLOW, I. J. et al. **Generative Adversarial Networks**. 2014. Disponível em: <https://arxiv.org/abs/1406.2661>. Citado na página 29.

Google. **Central de Privacidade dos apps do Gemini**. 2025. Acesso em 25 de maio de 2025. Disponível em: <https://support.google.com/a/answer/15706919?hl=pt-BR>. Citado na página 54.

_____. **IPolítica de Privacidade do Google**. 2025. Acesso em 25 de maio de 2025. Disponível em: <https://policies.google.com/privacy>. Citado na página 54.

Governo Digital. **Governança de Dados**. 2024. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/governancadedados>. Citado na página 33.

HAI, S. I. for H.-C. A. I. **AI Index Report 2024**. 2024. Disponível em: <https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf>. Citado 2 vezes nas páginas 13 e 30.

HO, J.; JAIN, A.; ABBEEL, P. **Denoising Diffusion Probabilistic Models**. 2020. Disponível em: <https://arxiv.org/abs/2006.11239>. Citado na página 29.

HUSSAIN, M. When, where, and which?: Navigating the intersection of computer vision and generative ai for strategic business integration. **IEEE Access**, v. 11, 2023. Citado na página 25.

KPMG. **Inteligência artificial generativa: riscos e oportunidades**. 2023. Acesso em: 27 out. 2024. Disponível em: <https://kpmg.com/br/pt/home/insights/2023/07/kpmg-inteligencia-artificial-generativa-riscos-oportunidades.html>. Citado 4 vezes nas páginas 13, 14, 23 e 31.

LIGHTHILL, J. Artificial intelligence: A general survey. In: **Artificial Intelligence: a paper symposium**. [S.l.]: Science Research Council, 1973. Citado na página 21.

LIMA, C. B.; SERRANO, A. Inteligência artificial generativa e chatgpt: uma investigação sobre seu potencial na educação. **Transinformação**, v. 36, p. e2410839, 2024. Disponível em: <https://doi.org/10.1590/2318-0889202436e2410839>. Citado na página 24.

LOVATTO, M. B. A. InteligÊncia artificial: GovernanÇa e transparÊncia? **Revista Ibmec de Direito - ISSN 3085-704X**, v. 1, n. 1, 2024. Disponível em: <https://ibmec.periodicoscientificos.com.br/index.php/cienciajuridica/article/view/245>. Citado na página 35.

LUGER, G. F. **Artificial Intelligence: Structures and Strategies for Complex Problem Solving**. 6th. ed. [S.l.]: Pearson, 2008. ISBN 978-0-321-54589-3. Citado na página 16.

MACHADO, A. d. O. B. **A Inteligência Artificial Generativa como Novo Agente Disruptor de Mercado**. 2023. Trabalho de Conclusão de Curso (Graduação) – Faculdade de Ciências Econômicas, Universidade Federal da Bahia, Salvador, 2023. Disponível em: <https://repositorio.ufba.br/handle/ri/39246>. Citado 4 vezes nas páginas 12, 13, 19 e 23.

MARCONDES, M. M. **Uma proposta de melhorias da governança de dados em uma startup do setor de software**. Trabalho de Conclusão de Curso — Universidade Federal de Ouro Preto, 2021. Disponível em: <http://www.monografias.ufop.br/handle/35400000/3180>. Citado na página 34.

MCCARTHY, J. **What is Artificial Intelligence?** Stanford, CA: [s.n.], 2007. Revised November 12, 2007. Disponível em: <http://www-formal.stanford.edu/jmc/>. Citado 2 vezes nas páginas 12 e 16.

MCCLELLAND, J. L.; RUMELHART, D. E.; HINTON, G. E. The appeal of parallel distributed processing. In: COLLINS, A. M.; SMITH, E. E. (Ed.). **Readings in cognitive science: A perspective from psychology and artificial intelligence**. [S.l.]: Morgan Kaufmann, 1988. p. 52–72. Citado na página 21.

MENDES, R. D. Inteligência artificial: sistemas especialistas no gerenciamento da informação. **Ciência da Informação**, v. 26, p. 39–45, 1997. Disponível em: <https://doi.org/10.1590/S0100-19651997000100006>. Citado na página 20.

MERRITT, R. **O que é um Modelo Transformer?** NVIDIA Blog, 2022. Disponível em: <https://blog.nvidia.com.br/blog/o-que-e-um-modelo-transformer/#:~:text=Como%20os%20Transformers%20Receberam%20Seu%20Nome&text=%E2%80%

9CO%20nome%20'Attention%20Net',quem%20criou%20o%20nome%20Transformer.> Citado na página 27.

Microsoft. **Política de Privacidade da Microsoft**. 2025. Acesso em 20 de abril de 2025. Disponível em: <https://www.microsoft.com/pt-br/privacy/privacystatement>. Citado na página 54.

MINSKY, M.; PAPERT, S. A. **Perceptrons**. Cambridge, MA: MIT Press, 1969. v. 6. 7 p. Citado na página 19.

NILSSON, N. **Introduction to Machine Learning—An Early Draft of a Proposed Textbook**. Stanford, CA: [s.n.], 1998. Disponível em: <https://ai.stanford.edu/~nilsson/MLBOOK.pdf>. Citado na página 19.

OpenAI. **Política de Privacidade**. 2024. Acesso em 20 de abril de 2025. Disponível em: <https://openai.com/pt-BR/policies/row-privacy-policy/>. Citado na página 54.

Organisation for Economic Co-operation and Development (OECD). **AI, data governance and privacy: Synergies and areas of international co-operation**. Paris: OECD Publishing, 2024. (OECD Artificial Intelligence Papers, 22). Disponível em: <https://doi.org/10.1787/2476b1a4-en>. Citado na página 14.

_____. **AI Principles**. 2025. Disponível em: <https://www.oecd.org/en/topics/ai-principles.html>. Citado 2 vezes nas páginas 36 e 55.

PIRES, F. **OpenAI Responds to ChatGPT User Account Credentials Found on Dark Web**. 2023. Acesso em 25 de maio de 2025. Disponível em: <https://www.tomshardware.com/news/over-100000-chatgpt-account-credentials-made-available-on-the-dark-web>. Citado na página 44.

Presidência da República (Brasil). **Lei nº 13.709, de 14 de agosto de 2018 — Lei Geral de Proteção de Dados Pessoais (LGPD)**. 2018. Acesso em 28 de abril de 2025. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>. Citado na página 55.

RAMOS, A. S. M. Inteligência artificial generativa baseada em grandes modelos de linguagem - ferramentas de uso na pesquisa acadêmica. **SciELO Preprints**, 2023. Acesso em: 16 nov. 2024. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/6105>. Citado 3 vezes nas páginas 23, 24 e 26.

RAUBER, T. W. Redes neurais artificiais. **Universidade Federal do Espírito Santo**, v. 29, 2005. Citado na página 19.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4. ed. [S.l.]: Pearson, 2013. ISBN 978-0136042594. Citado 7 vezes nas páginas 16, 17, 18, 19, 20, 21 e 22.

SHARMA, V.; GOYAL, M.; MALIK, D. An intelligent behaviour shown by chatbot system. **International Journal of New Technology and Research**, v. 3, n. 4, p. 263–312, 2017. Citado na página 20.

SINGLA, A. et al. **The state of AI: How organizations are rewiring to capture value**. 2025. Disponível em: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/>. Citado na página 34.

SMITH, C. et al. **The History of Artificial Intelligence**. 2006. Acesso em. Disponível em: <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>. Citado na página 23.

STANDARDS, N. I. of; (NIST), T. **AI Risk Management Framework (AI RMF) Resources**. 2024. Disponível em: <https://airc.nist.gov/airmf-resources/airmf/>. Citado 2 vezes nas páginas 36 e 55.

The Data Governance Institute. **Definitions of Data Governance**. 2025. Acesso em 1 de abril de 2025. Disponível em: <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/>. Citado na página 33.

VASWANI, A. et al. **Attention Is All You Need**. 2017. Disponível em: <https://arxiv.org/abs/1706.03762v1>. Citado 2 vezes nas páginas 23 e 27.

WESSEL, M. et al. Call for papers to the special issue: Generative ai and its transformative value for digital platforms. **Journal of Management Information Systems**, 2023. Disponível em: <https://www.jmis-web.org/cfps/JMIS_SI_CfP_Generative_AI.pdf>. Citado 2 vezes nas páginas 25 e 26.

Apêndices

APÊNDICE A

Documentação Analisada

Este apêndice apresenta os documentos oficiais das ferramentas de IA Generativa selecionadas (OpenAI, 2024; Google, 2025a; Google, 2025b; Microsoft, 2025) e as sintetiza na tabela 3.

Tabela 3 – Políticas de Ferramentas

Ferramenta	**Documentos Analisados**	**Atualização**
ChatGPT	<https://openai.com/pt-BR/policies/privacy-policy/>	nov, 2024
Copilot	<https://www.microsoft.com/pt-br/privacy/privacystatement>	mar, 2025
Gemini	<https://support.google.com/gemini/answer/13594961#privacy_notice> <https://policies.google.com/privacy>	mai, 2025

APÊNDICE **B**

Categorias de Análise e Fundamentos Teóricos

As categorias temáticas utilizadas neste trabalho foram definidas com base em diretrizes e recomendações de segurança e governança da informação, abrangendo o referencial teórico abordado, além de recomendações internacionais como (ENISA, 2024; STANDARDS; (NIST), 2024; Organisation for Economic Co-operation and Development (OECD), 2025) e, a LGPD (Presidência da República (Brasil), 2018), conforme tabela 4.

Tabela 4 – Categorias Temáticas

Categoria	**Fundamento Teórico**
Coleta de Dados	Fundamentos da LGPD (artigos 7º e 8º) e diretrizes do NIST sobre a minimização da coleta e necessidade de consentimento informado
Uso das Informações	Princípios de finalidade e transparência (LGPD, art. 6º), e recomendações da OECD quanto à limitação de uso e explicabilidade
Compartilhamento de Dados	Diretrizes da ENISA sobre compartilhamento seguro, e exigência de consentimento específico (LGPD, art. 7º §5º)
Medidas de Segurança	Princípios da segurança da informação, incluindo criptografia, controle de acesso e integridade (FONTES, 2017)
Conformidade Legal	Observância da LGPD e políticas de governança preconizadas pela OECD e pelo NIST para IA confiável

APÊNDICE C

Unidades de Registro e de Contexto

Este apêndice apresenta as unidades de registro e as unidades de contexto identificadas a partir da análise documental das políticas de privacidade e termos de uso de ferramentas IA Gen. A categorização foi conduzida com base na técnica de Análise de Conteúdo proposta por Bardin 1977, por meio da codificação temática de trechos que se repetem com frequência ou que expressam aspectos relevantes em relação aos objetivos do estudo.

As unidades foram agrupadas em categorias temáticas, com o objetivo de sintetizar e evidenciar padrões recorrentes entre os documentos analisados, desconsiderando elementos exclusivos de uma única ferramenta, a fim de garantir maior generalização e comparabilidade entre os dados.

A seguir, estão listadas as principais unidades de registro identificadas:

- Coleta de Dados
  - Coleta de dados pessoais fornecidos diretamente pelo usuário;
  - Coleta de dados gerados automaticamente durante o uso.
- Uso das Informações
  - Utilização de dados para fornecer, manter e melhorar os serviços e recomendações;
  - Aplicação dos dados no desenvolvimento e treinamento de modelos de IA;
  - Prevenção de fraudes, abusos e proteção dos sistemas.
- Compartilhamento de Dados
  - Compartilhamento com afiliadas, parceiros e prestadores de serviço para operação dos serviços;
  - Compartilhamento mediante consentimento do usuário ou obrigações legais.
- Medidas de Segurança

- Implementação de medidas técnicas, administrativas e organizacionais para proteção dos dados;
- Uso de criptografia na transmissão e armazenamento de dados sensíveis.

❑ Conformidade Legal

- Cumprimento das legislações de proteção de dados aplicáveis.

Nas tabelas de 5 a 9, são exibidas as unidades de contexto (trechos) extraídos das políticas de privacidade e termos de uso das ferramentas analisadas, classificadas de acordo com as categorias temáticas estabelecidas na metodologia.

Tabela 5 – Unidades de Registro – Coleta de dados

Ferramenta	**Unidades de Registro**
ChatGPT	"Quando você cria uma conta conosco, nós coletamos informações associadas à sua conta, incluindo seu nome, informações de contato, credenciais da conta, data de nascimento, dados de pagamento e histórico de transações..."; "Nós coletamos Dados Pessoais que você fornece ao inserir informações em nossos Serviços"; Nós coletamos informações que o seu navegador ou dispositivo envia automaticamente quando você utiliza os nossos Serviços"; "Nós podemos determinar a área geral a partir da qual o seu dispositivo acessa os nossos Serviços".
Copilot	"A Microsoft coleta dados de você, por meio de nossas interações com você e por meio de nossos produtos"; "Você fornece alguns desses dados diretamente, enquanto alguns deles obtemos ao coletar dados sobre suas interações, uso e experiências com nossos produtos";" Os dados que coletamos dependem do contexto de suas interações com a Microsoft e das suas escolhas, incluindo suas configurações de privacidade e os produtos e recursos que você usa"; "Também obtemos dados sobre você de afiliadas da Microsoft, subsidiárias e terceiros"; "Quando pedimos que forneça os seus dados pessoais, você pode recusar"; "Muitos dos nossos produtos exigem alguns dados pessoais para que o serviço seja oferecido".
Gemini	"O Google coleta suas conversas, incluindo gravações de suas interações no Gemini Live"; "O que você envia aos apps do Gemini, como arquivos, imagens, telas e conteúdo de páginas do seu navegador"; "Informações relacionadas ao uso de produtos, seu feedback, dados de apps conectados e sua localização"; "As informações sobre localização incluem a área geral em que seu dispositivo se encontra, endereço IP ou endereço residencial ou de trabalho registrados na sua Conta do Google".

Tabela 6 – Unidades de Registro – Uso das Informações

Ferramenta	**Unidades de Registro**
ChatGPT	"Fornecer, analisar e manter os nossos Serviços, por exemplo, para responder às suas perguntas para o ChatGPT"; "Melhorar e desenvolver os nossos Serviços e realizar pesquisas, por exemplo, para desenvolver novas funcionalidades de produtos"; "Comunicar-nos com você, inclusive para enviar informações sobre nossos Serviços e eventos, como mudanças ou melhorias nos Serviços"; "Evitar fraudes, atividades ilegais ou utilizações indevidas dos nossos Serviços e proteger a segurança dos nossos sistemas e Serviços"; "Cumprir com obrigações legais e proteger os direitos, a privacidade, a segurança ou a propriedade dos nossos usuários, da OpenAI ou de terceiros"; "Nós podemos usar o Conteúdo que você nos fornece para melhorar nossos Serviços, por exemplo, para treinar os modelos que alimentam o ChatGPT".
Copilot	"A Microsoft usa os dados que coletamos para proporcionar experiências sofisticadas e interativas"; "Fornecer nossos produtos, incluindo a atualização, segurança e solução de problemas, bem como o fornecimento de suporte"; "Isso também inclui o compartilhamento de dados, quando ele é necessário, para fornecer o serviço ou realizar as transações que você solicitar"; "Melhorar e desenvolver nossos produtos"; "Personalizar nossos produtos e fazer recomendações"; "Anunciar e comercializar para você, incluindo o envio de comunicações promocionais, o direcionamento de anúncios e a apresentação de ofertas relevantes para você"; "Também usamos os dados para operar nossos negócios, incluindo a análise de nosso desempenho, o cumprimento de nossas obrigações legais, o desenvolvimento de nossa força de trabalho e a realização de pesquisas"; "Para criar, treinar e melhorar a precisão dos nossos métodos automatizados de processamento, examinamos manualmente algumas das saídas produzidas pelos métodos automatizados em relação aos dados subjacentes"; "Como parte dos nossos esforços para melhorar e desenvolver os nossos produtos, poderemos utilizar os seus dados para desenvolver e treinar os nossos modelos de IA".
Gemini	"O Google usa esses dados, de acordo com nossa Política de Privacidade, para fornecer, aprimorar, desenvolver e personalizar produtos e serviços, além de tecnologias de aprendizado de máquina, incluindo produtos empresariais, como o Google Cloud".

Tabela 7 – Unidades de Registro – Compartilhamento de Dados

Ferramenta	**Unidades de Registro**
ChatGPT	"Nós podemos divulgar Dados Pessoais a fornecedores e prestadores de serviços, incluindo provedores de serviços de hospedagem, prestadores de serviços de atendimento ao cliente, serviços de nuvem, serviços de entrega de conteúdo, serviços de monitoramento de suporte e segurança, software de comunicação por e-mail, serviços de análise da navegação na web, processadores de pagamentos e transações e outros provedores de tecnologia da informação; Transferências comerciais"; "Se estivermos envolvidos em transações estratégicas, reestruturações, falência, liquidação ou transição de serviço para outro fornecedor, seus Dados Pessoais poderão ser divulgados durante o processo de diligência com as partes envolvidas e outras pessoas que estão auxiliando na Transação e transferidas, juntamente com outros ativos, para um sucessor ou afiliado no âmbito da Transação; "Nós podemos compartilhar seus Dados Pessoais, incluindo informações sobre sua interação com nossos Serviços, com autoridades governamentais, pares do setor ou outros terceiros em conformidade com a lei"; "Nós podemos divulgar Dados Pessoais aos nossos afiliados, ou seja, uma entidade que controla, é controlada por, ou está sob controle comum da OpenAI".
Copilot	"Compartilhamos seus dados pessoais com o seu consentimento ou para concluir transações ou fornecer um determinado produto solicitado ou autorizado"; "Podemos também compartilhar dados com afiliados e subsidiárias controlados pela Microsoft, com fornecedores autorizados, quando exigido por lei ou para responder a um processo jurídico";"O "compartilhamento" também está relacionado ao fornecimento de dados pessoais a terceiros para fins de publicidade personalizados".
Gemini	"Compartilhamos informações pessoais fora do Google quando temos seu consentimento"; "Pediremos seu consentimento explícito antes de compartilhar quaisquer informações pessoais sensíveis"; "Se você estuda ou trabalha em uma organização que usa os Serviços do Google, seu administrador do domínio e os revendedores que gerenciam a conta terão acesso à sua Conta do Google"; "Fornecemos informações pessoais às nossas afiliadas ou outras empresas ou pessoas confiáveis para tratar tais informações por nós, de acordo com nossas instruções e em conformidade com nossa Política de Privacidade";" Vamos compartilhar informações pessoais fora do Google se acreditarmos, de boa-fé, que a divulgação das informações seja razoavelmente necessária para cumprir legislação, proteger direitos e evitar fraudes"; "Podemos compartilhar informações de identificação não pessoal publicamente e com nossos parceiros, como editores, anunciantes, desenvolvedores ou detentores de direitos"; "Se o Google for envolvido em uma fusão, aquisição ou venda de ativos, continuaremos a garantir a confidencialidade das suas informações pessoais".

Tabela 8 – Unidades de Registro – Medidas de Segurança

Ferramenta	**Unidades de Registro**
ChatGPT	"Nós implementamos medidas técnicas, administrativas e organizacionais comercialmente razoáveis concebidas para proteger os Dados Pessoais contra perda, uso indevido, acesso, divulgação, alteração ou destruição não autorizados"; "Você deve ter cuidado ao decidir quais informações fornecer aos Serviços"; "Nós não somos responsáveis pela fraude de quaisquer configurações de privacidade ou medidas de segurança contidas no serviço ou em sites de terceiros".
Copilot	"Utilizamos uma vasta gama de tecnologias e procedimentos de segurança para ajudar a proteger seus dados pessoais contra acesso, utilização ou divulgação não autorizados"; "Armazenamos seus dados pessoais fornecidos em sistemas informáticos com acesso limitado e localizados em instalações controladas"; "Quando transmitimos dados altamente confidenciais (como um número de cartão de crédito ou senha) pela Internet, nós os protegemos com o uso de criptografia";'Nós também podemos agregar ou desidentificar Dados Pessoais para que eles deixem de identificar você e utilizá-los para os fins descritos acima, como para analisar a forma como os nossos Serviços são utilizados, melhorar e adicionar funcionalidades e realizar pesquisas"; "Nós manteremos e utilizaremos as informações desidentificadas no formato desidentificado e não tentaremos reidentificá-las, exceto se exigido por lei".
Gemini	"Utilização de criptografia para manter os seus dados privados enquanto estão em trânsito"; "A oferta de uma variedade de recursos de segurança, como a Navegação segura, Verificação de segurança e Verificação em duas etapas para ajudar você a proteger sua conta"; "A análise de nossa coleta, práticas de armazenamento e processamento de informações, o que inclui medidas de segurança física, para evitar acesso não autorizado aos nossos sistemas"; "A restrição ao acesso a informações pessoais por parte de funcionários, contratados e representantes do Google que necessitam dessas informações para processá-las. Toda pessoa com esse acesso está sujeita a rigorosas obrigações contratuais de confidencialidade, podendo ser disciplinada ou dispensada se deixar de cumprir tais obrigações."

Tabela 9 – Unidades de Registro – Conformidade Legal

Ferramenta	**Unidades de Registro**
ChatGPT	"A OpenAI seus Dados Pessoais para as finalidades descritas nesta Política de Privacidade em servidores localizados em vários territórios, incluindo o tratamento e o armazenamento dos seus Dados Pessoais em nossas instalações e servidores nos Estados Unidos"; "Embora a legislação de proteção de dados varie de país para país, aplicamos as proteções descritas nesta política aos seus Dados Pessoais, independentemente do local onde eles são tratado";" Apenas transferimos esses dados de acordo com mecanismos de transferência legalmente válidos."
Copilot	"A Microsoft cumpre as leis de proteção de dados aplicáveis, incluindo leis de notificação de violação de segurança aplicáveis";"Cumprir a lei ou responder a processos legais, incluindo os provenientes de autoridades ou órgãos governamentais."; "Você pode ter esses direitos pelas leis aplicáveis, incluindo o Regulamento Geral de Proteção de Dados da União Europeia (GDPR), mas nós os oferecemos independentemente de sua localização. Em alguns casos, sua capacidade de acessar ou controlar seus dados pessoais poderá ser limitada, conforme necessário ou permitido pela lei aplicável";" Observe que a capacidade de um responsável de acessar e/ou excluir as informações pessoais de uma criança no painel de privacidade varia de acordo com as leis de onde você está localizado";"A Microsoft está sujeita aos poderes de investigação e cumprimento da lei da Federal Trade Commission (FTC) dos EUA".
Gemini	"Revisamos regularmente esta Política de Privacidade e nos certificamos de que processamos suas informações de formas que estão em conformidade com ela"; "Alteramos esta Política de Privacidade periodicamente". "Nós não reduziremos seus direitos nesta Política de Privacidade sem seu consentimento explícito"; "Se a legislação de proteção de dados do Brasil se aplicar ao tratamento das suas informações, forneceremos os controles descritos nesta política para que você possa exercer seu direito de: receber confirmação sobre o tratamento de suas informações; atualizar, corrigir, anonimizar, remover e solicitar acesso às suas informações; restringir ou se opor ao tratamento das suas informações; exportar suas informações para outro serviço".