



UNIVERSIDADE DE BRASÍLIA – UnB

FACULDADE DE DIREITO

RAFAEL DE OLIVEIRA BITTENCOURT

**APLICAÇÕES E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA
NO DIREITO: UM ESTUDO DE CASO DO TRIBUNAL DE CONTAS DA UNIÃO E
A ANÁLISE DA UTILIDADE DE ASSISTENTES JURÍDICOS DE IA CRIADOS NO
CHATGPT**

BRASÍLIA

2024

RAFAEL DE OLIVEIRA BITTENCOURT

**APLICAÇÕES E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA
NO DIREITO: UM ESTUDO DE CASO DO TRIBUNAL DE CONTAS DA UNIÃO E
A ANÁLISE DA UTILIDADE DE ASSISTENTES JURÍDICOS DE IA CRIADOS NO
CHATGPT**

Monografia apresentada à Faculdade de Direito da Universidade de Brasília – UnB, como requisito parcial para a obtenção do título de bacharel em Direito.

Orientador. Prof. Dr. ALEXANDRE KEHRIG VERONESE AGUIAR

Brasília

2024

CIP - Catalogação na Publicação

Ba Bittencourt, Rafael de Oliveira.
APLICAÇÕES E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL
GENERATIVA NO DIREITO: UM ESTUDO DE CASO DO TRIBUNAL DE
CONTAS DA UNIÃO E A ANÁLISE DA UTILIDADE DE ASSISTENTES
JURÍDICOS DE IA CRIADOS NO CHATGPT / Rafael de Oliveira
Bittencourt; orientador Alexandre KEHRIG VERONESE AGUIAR. --
Brasília, 2024.
109 p.

Monografia (Graduação - Direito) -- Universidade de
Brasília, 2024.

1. Inteligência Artificial Generativa. 2. Direito e
Inteligência Artificial. 3. ChatGPT. 4. Engenharia de
Prompt. 5. Assistentes Jurídicos de IA. I. AGUIAR, Alexandre
KEHRIG VERONESE, orient. II. Título.

RAFAEL DE OLIVEIRA BITTENCOURT

**APLICAÇÕES E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA
NO DIREITO: UM ESTUDO DE CASO DO TRIBUNAL DE CONTAS DA UNIÃO E
A ANÁLISE DA UTILIDADE DE ASSISTENTES JURÍDICOS DE IA CRIADOS NO
CHATGPT**

Monografia apresentada à Faculdade de Direito
como requisito parcial para a obtenção do título
de bacharel em Direito.

Orientador. Profa. Dra. ALEXANDRE KEHRIG
VERONESE AGUIAR

Prof. Dr. ALEXANDRE KEHRIG VERONESE AGUIAR

Profa. Dra. FERNANDA DE CARVALHO LAGE

Profa. Dra. TAINÁ AGUIAR JUNQUILHO

Brasília

2024

AGRADECIMENTOS

Este trabalho somente foi possível devido ao apoio e à inspiração de pessoas fundamentais em minha vida, às quais expresso minha profunda gratidão.

À minha esposa, Aline Gonçalves, companheira de quase dez anos, dedico grande parte deste trabalho. Seu apoio, incentivo e amor foram essenciais, especialmente durante as longas horas dedicadas à pesquisa e ao estudo da Inteligência Artificial Generativa. Sem você ao meu lado, nada disso teria sido possível. Você me inspira a ser uma pessoa e um profissional melhor e, com você, aprendo diariamente a cultivar hábitos que me impulsionam a alcançar resultados cada vez maiores.

À minha mãe, Ângela, pelo carinho, amor incondicional e incentivo desde a minha infância, sempre me fazendo acreditar que eu poderia alcançar meus objetivos. Ao meu pai, Sérgio, com quem compartilho excelentes memórias de infância e adolescência, seja nos jogos do Vitória no Barradão, seja jogando futebol juntos, expresso meu amor e admiração.

À minha tia Aldinha, que, como uma segunda mãe, sempre cuidou de nossa família com amor e paciência. Ao meu saudoso avô Aldo, *in memoriam*, cujas conversas, ensinamentos e suporte financeiro, que me permitiu dar os primeiros passos na programação, moldaram quem sou e continuam a guiar meus passos.

Aos meus primos Bittencourt e Cordeiro, pela presença constante em todas as fases da minha vida. À minha afilhada Madalena, cujo sorriso e alegria são fontes inesgotáveis de energia e motivação para mim. O convite para ser seu padrinho foi, sem dúvida, um dos momentos mais felizes da minha vida.

À Faculdade de Direito da Universidade de Brasília (UnB) e aos docentes, que me fizeram enxergar a importância de uma formação jurídica e social. Ao meu orientador, Alexandre Veronese, agradeço pela constante disponibilidade em fornecer orientações e pelo valioso empréstimo de livros que tanto contribuíram para esta empreitada.

Ao Tribunal de Contas da União, onde atuo desde 2015 como Auditor Federal, agradeço o ambiente de trabalho, as oportunidades e o apoio na condução desta pesquisa, com especial reconhecimento a Eric Hans e Luís Mitchell pelas entrevistas concedidas.

Trilho o caminho da Ciência de Dados e Inteligência Artificial há quase dez anos e, ao me deparar com a IA Generativa, especialmente o ChatGPT no início de 2023, percebi que essa tecnologia moldaria o futuro. Minha motivação é promover o uso responsável dessa tecnologia, aproveitando todo o seu potencial. A todos que contribuíram para a realização deste trabalho, minha eterna gratidão.

RESUMO:

Esta monografia investiga as aplicações e limitações da Inteligência Artificial Generativa (IAG) no campo jurídico, com ênfase no estudo de caso do Tribunal de Contas da União (TCU) e na análise da utilidade de assistentes jurídicos desenvolvidos na plataforma ChatGPT. A pesquisa aborda os fundamentos e a evolução das Inteligências Artificiais (IAs), com foco nas tecnologias generativas baseadas em grandes modelos de linguagem (*Large Language Models* – LLMs), destacando suas potencialidades e limitações. São examinados os principais casos de uso, como a sumarização de documentos, a pesquisa jurídica por meio de buscas semânticas e a assistência na elaboração de peças processuais, além das técnicas de *Engenharia de Prompt*, que otimizam as respostas geradas pela IA. Também são analisados os desafios técnicos, éticos e legais relacionados à implementação dessas tecnologias, incluindo questões de privacidade, viés algorítmico, transparência e conformidade com a Lei Geral de Proteção de Dados (LGPD), a Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ), bem como uma análise sumária do Projeto de Lei nº 2.338/2023, que propõe a regulamentação da IA no Brasil. Identificam-se problemas nesse projeto, como imprecisões terminológicas e preocupações com a carga regulatória excessiva, que podem impactar negativamente o desenvolvimento da IA no país, especialmente para pequenas e médias empresas (PMEs).

O estudo de caso do TCU revela iniciativas inovadoras como o ChatTCU e o INACIA (Instrução Assistida com Inteligência Artificial), evidenciando o impacto positivo da IAG na eficiência e precisão dos processos de auditoria e controle, sempre observando as salvaguardas éticas e legais. A pesquisa avalia ainda a percepção dos profissionais do TCU quanto à adoção dessas tecnologias, identificando barreiras relacionadas ao conhecimento técnico e oportunidades para sua implementação. Os resultados indicam que os assistentes jurídicos desenvolvidos aumentam significativamente a produtividade, melhorando a eficiência na realização de tarefas como pesquisa jurídica, revisão de documentos e elaboração de peças processuais. Além disso, constatou-se que essas ferramentas podem reduzir erros humanos, oferecendo suporte em processos complexos e contribuindo para a democratização do acesso à informação jurídica.

O trabalho propõe diretrizes para uma implementação e uso pragmático e responsável da IA Generativa no Direito, enfatizando a necessidade de validação das informações, mitigação de riscos e responsabilidade compartilhada entre desenvolvedores e usuários. Nesse sentido, sugere-se que a IA Generativa seja encarada como uma extensão do conhecimento negocial e da capacidade intelectual dos usuários, numa abordagem denominada co-inteligência. Conclui-se que, embora existam limitações técnicas e desafios éticos, a IAG apresenta grande potencial

para aprimorar os processos jurídicos, desde que utilizada com supervisão humana e de forma responsável, respeitando as melhores práticas internacionais e o contexto jurídico brasileiro.

Palavras-chave: Inteligência Artificial Generativa; Regulamentação de IA; Direito; Tribunal de Contas da União; Ética na IA; Tecnologia Jurídica; Grandes Modelos de Linguagem; Engenharia de Prompt; Projeto de Lei 2338/2023; Assistentes Jurídicos; ChatGPT; OpenAI; Anthropic; Gemini; Claude.

ABSTRACT:

This monograph investigates the applications and limitations of Generative Artificial Intelligence (GenAI) in the legal field, with an emphasis on the case study of the Brazilian Federal Court of Accounts (TCU) and the analysis of the usefulness of legal assistants developed on the ChatGPT platform. The research addresses the fundamentals and evolution of Artificial Intelligence (AI), focusing on generative technologies based on Large Language Models (LLMs), highlighting their potential and limitations. The main use cases analyzed include document summarization, legal research through semantic searches, and assistance in drafting legal documents, as well as *Prompt Engineering* techniques that optimize the responses generated by AI. Technical, ethical, and legal challenges related to the implementation of these technologies are also examined, including issues of privacy, algorithmic bias, transparency, and compliance with the General Data Protection Law (LGPD) and Resolution No. 332/2020 of the National Council of Justice (CNJ). Additionally, a summary analysis of Bill No. 2.338/2023, which proposes AI regulation in Brazil, is provided. Issues in the bill, such as terminological inaccuracies and concerns about excessive regulatory burdens, which may negatively impact AI development in the country, especially for small and medium-sized enterprises (SMEs), are identified.

The TCU case study highlights innovative initiatives such as ChatTCU and INACIA (AI-Assisted Instruction), demonstrating the positive impact of GenAI on the efficiency and accuracy of auditing and control processes, always considering ethical and legal safeguards. The research also evaluates the perception of TCU professionals regarding the adoption of these technologies, identifying barriers related to technical knowledge and opportunities for implementation. The results indicate that the legal assistants developed significantly increase productivity, improving efficiency in performing tasks such as legal research, document review, and drafting legal documents. Furthermore, these tools have been found to reduce human error, offering support in complex processes and contributing to the democratization of access to legal information.

The study proposes guidelines for the pragmatic and responsible implementation and use of Generative AI in the legal field, emphasizing the need for information validation, risk mitigation, and shared responsibility between developers and users. In this context, it is suggested that Generative AI be viewed as an extension of the user's business knowledge and intellectual capacity, within an approach called co-intelligence. It is concluded that, although there are technical limitations and ethical challenges, GenAI has great potential to improve legal

processes, provided it is used with human oversight and in a responsible manner, respecting international best practices and the Brazilian legal context.

Keywords: Generative Artificial Intelligence, AI Regulation, Law, Brazilian Federal Court of Accounts, AI Ethics, Legal Technology, Large Language Models, Prompt Engineering, Bill No. 2338/2023, Legal Assistants, ChatGPT, OpenAI, Anthropic, Gemini, Claude.

Sumário

1. INTRODUÇÃO	15
1.1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA	15
1.2 OBJETIVOS	20
1.2.1 OBJETIVO GERAL	20
1.2.2 OBJETIVOS ESPECÍFICOS	20
1.3 ESTRUTURA DO TRABALHO	21
2. METODOLOGIA	22
3. FUNDAMENTOS E EVOLUÇÃO DAS INTELIGÊNCIAS ARTIFICIAIS	23
3.1 DEFINIÇÕES INTRODUTÓRIAS	23
3.2 HISTÓRICO E EVOLUÇÃO DAS INTELIGÊNCIAS ARTIFICIAIS	26
3.3 CONCEITOS E FUNCIONAMENTO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA	28
3.4 MODELOS BASE, GRANDES MODELOS DE LINGUAGEM E AJUSTE FINO	30
3.5 COMPARAÇÃO ENTRE MODELOS LLM PROPRIETÁRIOS, OPEN-SOURCE E OPEN-WEIGHT	34
3.6 ASSISTENTES DE IA, CHATBOTS E AGENTES DE IA	35
4. ENGENHARIA DE PROMPT: OTIMIZANDO INTERAÇÕES COM IAS GENERATIVAS	35
4.1 CONCEITOS BÁSICOS DE ENGENHARIA DE PROMPT	36
4.2 TÉCNICAS DE PROMPTING	38
4.3 PADRÕES DE PROMPT: ESTRUTURAS REUTILIZÁVEIS PARA OTIMIZAÇÃO DE INTERAÇÕES COM LLMS	40
4.3.1 PADRÃO DE PERSONA E PADRÃO DE AUDIÊNCIA (PÚBLICO-ALVO)	41
4.3.2 PADRÃO DE REFINAMENTO DE QUESTÕES	42
4.3.3 PADRÃO DE VERIFICADOR COGNITIVO	43
4.3.4 PADRÃO DE INTERAÇÃO INVERTIDA	43

4.3.5 PADRÃO DE REFLEXÃO	44
4.3.6 PADRÃO DE TEMPLATE	44
4.4 PRINCÍPIOS-CHAVE DE ENGENHARIA DE PROMPT	45
4.4.1 DIREÇÃO E CONTEXTO	45
4.4.2 CLAREZA ESTRUTURAL	46
4.4.3 ESPECIFICAÇÃO DE FORMATO DE SAÍDA	47
4.4.4 DIVISÃO DE TAREFAS COMPLEXAS	47
4.4.5 FORNECIMENTO DE EXEMPLOS (FEW-SHOT)	49
4.4.6 AVALIAÇÃO DA QUALIDADE	49
5. APLICAÇÕES E IMPACTOS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA NO DIREITO	50
5.1 A IA GENERATIVA NA PERSPECTIVA DOS PROFISSIONAIS JURÍDICOS	52
5.1.1 CONSCIENTIZAÇÃO E ADOÇÃO DA IA GENERATIVA	53
5.1.2 PERCEÇÕES E PREOCUPAÇÕES PROFISSIONAIS	54
5.1.3 BARREIRAS À ADOÇÃO	55
5.1.4 IMPACTO NOS NEGÓCIOS E RELAÇÕES COM CLIENTES	56
5.1.5 FUTURO DO TRABALHO NO SETOR JURÍDICO	57
5.2 PRINCIPAIS CASOS DE USO DE IA GENERATIVA NO CONTEXTO JURÍDICO	57
5.2.1 SUMARIZAÇÃO DE PEÇAS E PROCESSOS JURÍDICOS	58
5.2.2 PESQUISA JURÍDICA	59
5.2.3 CONSULTORIA JURÍDICA	63
5.2.4 REDAÇÃO, REVISÃO E ANÁLISE DE DOCUMENTOS JURÍDICOS	63
5.2.5 TRADUÇÃO DE TEXTOS JURÍDICOS	65
5.2.6 TRANSCRIÇÃO E ANÁLISE DE ÁUDIOS E VÍDEOS DE JULGAMENTOS OU EVIDÊNCIAS .	65
5.2.7 TRANSFORMAÇÃO DE DOCUMENTOS JURÍDICOS PARA O FORMATO VISUAL LAW E LEGAL DESIGN	66

5.2.8 OTIMIZAÇÃO DE RELACIONAMENTO COM O CLIENTE	67
5.2.9 ASSISTENTE DE INSTRUÇÃO PROCESSUAL.....	67
5.3 INICIATIVAS DE IA GENERATIVA NO DIREITO AMERICANO	68
5.3.1 HARVEY AI	68
5.3.2 PAXTON AI.....	70
5.4 INICIATIVAS DE IA GENERATIVA NO DIREITO BRASILEIRO	71
5.4.1 TRIBUNAL DE CONTAS DA UNIÃO: CHATTCU E INACIA	72
5.4.2 USP e ESCOLA SUPERIOR DA MAGISTRATURA TOCANTINENSE: SARA.....	73
5.4.3 MARITACA AI: JURU	73
6. DESAFIOS E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NO DIREITO	73
6.1 LIMITAÇÕES TÉCNICAS	74
6.1.1 GERAÇÃO DE CONTEÚDO FACTUALMENTE INCORRETO	75
6.1.2 EXPLICABILIDADE	78
6.1.3 LIMITAÇÃO DE TOKENS DA JANELA DE CONTEXTO E SAÍDA.....	80
6.1.4 DIFICULDADE COM TAREFAS QUE EXIJAM RACIOCÍNIO COMPLEXO.....	81
6.1.5 PROMPT INJECTION: JAILBREAK E PROMPT LEAKING.....	82
6.2 LIMITAÇÕES ÉTICAS E LEGAIS.....	83
6.2.1 PRIVACIDADE E PROTEÇÃO DE DADOS	84
6.2.2 PROPRIEDADE INTELECTUAL E DIREITOS AUTORAIS.....	85
6.2.3 VIÉS E DISCRIMINAÇÃO	87
6.2.4 TRANSPARÊNCIA E CONFIABILIDADE	90
6.2.5 CONFORMIDADE REGULATÓRIA.....	91
6.2.6 RESPONSABILIDADE (ACCOUNTABILITY).....	99
7. ASSISTENTES JURÍDICOS DE IA.....	101

8. ESTUDO DE CASO DA IMPLEMENTAÇÃO DE SOLUÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NO TRIBUNAL DE CONTAS DA UNIÃO.....	109
8.1 CHATTCU	109
8.1.1 CONTEXTUALIZAÇÃO.....	109
8.1.2 DESENVOLVIMENTO TÉCNICO E INTEGRAÇÃO	110
8.1.3 FUNCIONALIDADES E APLICAÇÕES.....	110
8.1.4 SEGURANÇA, PRIVACIDADE E CONFORMIDADE	111
8.1.5 IMPACTO E BENEFÍCIOS.....	111
8.1.6 DESAFIOS TÉCNICOS, ÉTICOS E LEGAIS	111
8.1.7 PERSPECTIVAS FUTURAS	112
8.1.8 ANÁLISE DA ADOÇÃO DE IA GENERATIVA POR SERVIDORES DO TCU	112
8.2 INACIA.....	115
8.2.1 CONTEXTUALIZAÇÃO.....	115
8.2.2 DESENVOLVIMENTO TÉCNICO E FUNCIONALIDADES	115
8.2.3 IMPACTO POTENCIAL E BENEFÍCIOS.....	116
8.2.4 DESAFIOS ÉTICOS E LEGAIS	116
8.2.5 AVALIAÇÃO E RESULTADOS	117
8.2.6 PERSPECTIVAS FUTURAS	117
9. DESAFIOS EMERGENTES DA IA GENERATIVA NO DIREITO.....	118
9.1 ARQUITETURAS MULTIAGENTES E A ESCALABILIDADE EM SISTEMAS DE IA	120
9.2 DEEPFAKE E DESINFORMAÇÃO.....	121
9.3 PROPRIEDADE INTELECTUAL DE PROMPTS E CONTEÚDO GERADO POR IA GENERATIVA	123
9.4 ALINHAMENTO DE INTELIGÊNCIA ARTIFICIAL.....	124
9.5 PLANO BRASILEIRO DE INTELIGÊNCIA ARTIFICIAL (PBIA) 2024-2028	125

10. CONCLUSÃO.....	126
REFERÊNCIAS.....	130

1. INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA

Embora a Inteligência Artificial (IA) possua uma história de quase 70 anos, nas últimas duas décadas ocorreram avanços significativos, com a consolidação de algoritmos de IA baseados em aprendizado de máquina. Esses algoritmos aprendem por meio de padrões extraídos de grandes volumes de dados. Tais avanços estão relacionados ao crescimento exponencial na quantidade de dados gerados, impulsionado pela popularização e democratização do acesso à internet, pela digitalização de serviços, pelo surgimento de mídias e plataformas sociais, bem como pela proliferação de dispositivos de Internet das Coisas (IoT), desde smartphones a sensores industriais. Essa abundância de dados em diferentes formatos (textos, áudios, vídeos), comumente referida como Big Data, foi acompanhada pelo aumento do poder computacional a um custo menor e pelo desenvolvimento de novas tecnologias, incluindo algoritmos baseados em Redes Neurais Profundas (*Deep Learning*), que permitiram a coleta, armazenamento e análise de quantidades imensas de informações (Russell; Norvig, 2020, p. 26; Coeckelbergh, 2020, p. 86-87).

Atualmente, estima-se que a quantidade de dados produzida diariamente alcance aproximadamente 402,74 milhões de *terabytes*, o que equivale a cerca de 147 *zettabytes* anualmente. Esse número está projetado para atingir 181 *zettabytes* em 2025 (1 *zettabyte* é igual a 10^{21} *bytes*.). Comparado a 2010, quando o volume anual de dados gerados era de aproximadamente 2 *zettabytes*, observa-se um crescimento exponencial da matéria-prima para modelos de Aprendizado de Máquina (Exploding Topics, 2024; Rivery, 2024).

Esse cenário trouxe desafios e oportunidades para empresas e governos, que passaram a encarar os dados como um ativo estratégico, utilizando algoritmos de aprendizado de máquina para tomar decisões informadas. Desse modo, a capacidade de analisar grandes volumes de dados em tempo real permitiu que as organizações identificassem padrões, fizessem previsões e otimizassem processos de maneira mais eficaz, gerando vantagens competitivas significativas.

Na área jurídica, desenvolveram-se aplicações baseadas em Inteligência Artificial para auxiliar advogados, juízes e outros profissionais do Direito a desempenharem suas funções com maior eficiência. Para tanto, foram utilizadas técnicas de Processamento de Linguagem Natural (PLN), que permitem a compreensão semântica e a análise de grandes volumes de textos jurídicos. Nesse cenário, destacam-se as seguintes aplicações da Inteligência Artificial no

Direito: a jurimetria, que envolve a análise estatística de decisões judiciais e a predição de desfechos processuais; a avaliação de riscos em contratos e litígios; a categorização temática de documentos; a otimização de pesquisas jurisprudenciais; o *e-discovery*, empregado na análise de grandes volumes de dados eletrônicos em litígios; e a automatização de tarefas repetitivas, como o preenchimento de formulários.

Por exemplo, o software COIN (*Contract Intelligence*) desenvolvido pela empresa J.P. Morgan, automatiza a revisão de documentos legais, particularmente contratos de crédito, utilizando Aprendizado de Máquina. Segundo a empresa, o COIN pode revisar milhares de contratos em segundos, substituindo 360.000 horas de trabalho humano e aumentando a precisão (LEGAL ML, 2018).

Em contextos governamentais, a IA é empregada em sistemas de vigilância pública, contribuindo para a identificação de atividades suspeitas e a manutenção da segurança. Pode-se citar como outras aplicações a detecção de fraudes tributárias, realizada por meio da análise de padrões e anomalias em declarações fiscais e transações financeiras, com o objetivo de identificar potenciais casos de evasão fiscal, e a otimização de processos internos, que visa aumentar a eficiência operacional por meio da automação e do aprimoramento das rotinas administrativas.

Nos Estados Unidos, observa-se a utilização de soluções baseadas em Inteligência Artificial em atividades judiciárias, como o sistema COMPAS para avaliação de risco, com o objetivo de auxiliar juízes em decisões de fiança e sentenciamento. No entanto, essa prática gerou controvérsias e debates sobre sua eficácia, justiça e impacto nos direitos dos réus. No caso *State v. Loomis* (2016), o juiz utilizou a pontuação de risco gerada pelo COMPAS para justificar uma sentença mais longa do que a inicialmente proposta pela acusação e defesa, decisão que foi mantida pela Suprema Corte de Wisconsin, apesar dos argumentos de Loomis sobre a opacidade do algoritmo e a falta de possibilidade de contestar seu funcionamento. Similarmente, no caso *State v. Corey R. Jones* (2017), o COMPAS foi utilizado para negar fiança a Jones, sob a alegação de que ele representava um alto risco de fuga e de cometer novos crimes, decisão também mantida pelo Tribunal de Apelações de Wisconsin, apesar das alegações da defesa sobre o viés racial do sistema. Esses casos suscitaram debates aprofundados sobre a ética e os desafios da implementação da IA no sistema judicial americano, especialmente no que tange à transparência, à equidade e à garantia dos direitos fundamentais dos réus (Washington, 2018).

No que concerne ao Estado Brasileiro, embora ainda não se tenha conhecimento de aplicações diretas na função judicante, diversas iniciativas de Inteligência Artificial foram

implementadas nos últimos anos para otimizar processos e aumentar a eficiência no setor público. Nesse sentido, destaca-se o Tribunal de Contas da União (TCU), que desenvolveu uma série de robôs de IA: Alice (Análise de Licitações e Editais), que lê licitações e editais para identificar fraudes; Sofia (Sistema de Orientação sobre Fatos e Indícios para o Auditor), que auxilia auditores na redação de textos e cria alertas com dados importantes; e Monica (Monitoramento Integrado para Controle de Aquisições), que monitora compras públicas na esfera federal. Adicionalmente, no controle aduaneiro, a Receita Federal do Brasil implementou o Sistema de Seleção Aduaneira por Aprendizado de Máquina (Sisam), uma inteligência artificial que analisa o histórico de declarações de importação para detectar erros e aumentar a precisão das verificações (Brasil, 2021; Jambeiro Filho, 2015).

No Poder Judiciário, o Supremo Tribunal Federal (STF) criou a ferramenta Victor, destinada a identificar recursos extraordinários vinculados a temas de repercussão geral (Lage, 2020; Maia Filho; Junquilha, 2018). Ademais, o Ministério Público Federal (MPF) passou a empregar a ferramenta HALBERT Corpus para classificar pareceres em Habeas Corpus quanto à admissibilidade e mérito (Brasil, 2021).

Em uma perspectiva mais ampla, observa-se o emprego da Inteligência Artificial em diversas escalas na maioria dos serviços fornecidos por empresas e governos. Entre os exemplos mais comuns, incluem-se assistentes virtuais em serviços de atendimento ao cliente, sistemas de recomendação em plataformas de streaming e e-commerce, algoritmos de detecção de fraudes em instituições financeiras e ferramentas de diagnóstico assistido por computador na área da saúde. A IA tornou-se tão integrada ao cotidiano que, muitas vezes, sua presença sequer é percebida.

Essa contextualização faz-se necessária considerando que, com a rápida disseminação de Inteligências Artificiais Generativas (IAGs) nos últimos dois anos, é comum encontrar discursos que ignoram as aplicações mais amplas e tradicionais da IA, referindo-se a ela apenas em seu aspecto generativo. Contudo, é importante frisar que a IA Generativa, embora apresente um potencial disruptivo, representa atualmente apenas uma pequena parcela das aplicações de Inteligência Artificial.

Diante disso, considera-se que o ano de 2022 marcou um ponto de inflexão no campo da Inteligência Artificial com o lançamento da versão web do ChatGPT em novembro daquele ano. Este evento impulsionou o desenvolvimento e a popularização das IAs Generativas, revelando novas aplicações e possibilidades tecnológicas. Diferentemente das IAs tradicionais, focadas em reconhecimento de padrões e fazer previsões, as IAs Generativas são capazes de compreender a linguagem natural e gerar conteúdo novo e diversificado.

Essa evolução representa uma mudança de paradigma com potencial de transformação social e econômica. As IAs Generativas vão além da automação de tarefas, abrangendo atividades mais complexas que requerem criatividade e adaptabilidade. A produção de textos, imagens, música e código de software por essas ferramentas redefine as possibilidades em diversos setores, revivendo discussões sobre a viabilidade e alcance de uma IA Geral (AGI).

As IAs Generativas também marcam uma transição na interação homem-máquina. Antes operando nos bastidores de serviços e aplicativos, agora elas atuam como interlocutoras diretas, funcionando como assistentes e modificando a relação dos usuários com a tecnologia. No âmbito do Direito, entende-se que as IAs Generativas possuem o potencial de transformar substancialmente o setor jurídico, impactando positivamente a produtividade dos profissionais, otimizando o sistema judiciário e democratizando o acesso ao conhecimento legal e à justiça. As aplicações são diversas, incluindo a consulta e pesquisa jurídica para a elaboração de estratégias processuais, a criação personalizada de resumos de peças e processos, e a redação de minutas de documentos jurídicos, assim como os desafios e riscos ligados a um uso desmedido ou inadequado dessas tecnologias.

Os possíveis benefícios e impactos dessa tecnologia no Direito já começam a ser reconhecidos por autoridades judiciárias brasileiras. Durante o encontro do J20, fórum de presidentes de Supremas Cortes, ocorrido em maio de 2024 no Rio de Janeiro, o Ministro Luís Roberto Barroso, presidente do STF, afirmou que a Inteligência Artificial poderá, em breve, redigir as primeiras versões de sentenças judiciais. Barroso destacou o uso atual da IA no agrupamento de processos e na localização de precedentes, defendendo a tecnologia como essencial para acelerar o sistema de justiça no Brasil. Ele ressaltou, entretanto, a necessidade de supervisão humana para evitar preconceitos e proteger direitos fundamentais (Migalhas, 2024).

Em decisão recente de 21 de junho de 2024, o Conselho Nacional de Justiça (CNJ) rejeitou o pedido para proibir o uso de IA no Judiciário. O Conselho justificou que a utilização de ferramentas como o ChatGPT já observa a Resolução CNJ nº 332/2020 e deve ser acompanhada de supervisão humana rigorosa para garantir decisões éticas e responsáveis. A decisão destacou que a adoção de tecnologias de inteligência artificial no sistema judicial pode promover eficiência e eficácia, desde que observe princípios éticos, jurídicos e constitucionais. Dessa forma, a supervisão humana é fundamental para assegurar a equidade e a imparcialidade das decisões judiciais, prevenindo o viés algorítmico e mantendo a responsabilidade ética do julgamento humano (Conjur, 2024; Brasil, 2020).

Adicionalmente, o Tribunal de Contas da União (TCU), órgão reconhecidamente pioneiro na adoção de tecnologias de inteligência artificial aplicáveis ao controle externo, possui iniciativas promissoras de incorporar IAs Generativas nos processos de trabalho de seus servidores. O ChatTCU, uma adaptação do ChatGPT, exemplifica essa inovação, proporcionando um assistente virtual integrado à base de conhecimento da Corte de Contas, facilitando a pesquisa, o resumo de documentos, a tradução de textos e as respostas a perguntas frequentes. Destaca-se também a Encomenda Tecnológica (ETEC), que visa a desenvolver um módulo de instrução assistida por IA para aprimorar a capacitação dos auditores, agregando funcionalidades como extração de significado das peças processuais, jurimetria e assistência na redação de instruções.

Por outro lado, o aumento no uso de IAs Generativas no Direito levanta questões éticas e legais significativas, como o viés em decisões automatizadas, a privacidade dos dados processados e a responsabilidade jurídica por ações automatizadas. Assim, este estudo justifica-se pela necessidade de compreender como maximizar o potencial dessas ferramentas no contexto jurídico, garantindo que respeitem os preceitos éticos e legais vigentes, evitando abusos e preservando direitos fundamentais.

Diante do exposto, a hipótese central deste estudo é que a adoção de uma abordagem pragmática e responsável no uso da Inteligência Artificial Generativa (IAG) no campo jurídico, respaldada por uma regulamentação equilibrada e iniciativas de literacia em IA, pode aumentar significativamente a eficiência e a celeridade processual, respeitando os limites técnicos, éticos e legais.

Por "abordagem pragmática e responsável", entende-se o uso de IAs Generativas focado nas potencialidades e principais casos de uso, com conhecimento de suas premissas de funcionamento, reconhecimento de suas limitações técnicas, éticas e legais, adoção de práticas para mitigá-las e manutenção da centralidade humana nas interações. Esta abordagem também implica a necessidade de desenvolvimento contínuo da literacia em IA pelos usuários, a fim de entender, utilizar e interagir de forma eficaz, responsável e ética com essas ferramentas.

A regulamentação equilibrada, por sua vez, refere-se a um marco regulatório que visa a proteger direitos fundamentais sem inibir o desenvolvimento tecnológico, alinhando-se aos padrões internacionais, mas respeitando o contexto socioeconômico brasileiro. Tal regulamentação busca criar um ambiente propício para a inovação responsável, garantindo simultaneamente a proteção dos direitos individuais e coletivos.

Essa premissa será testada por meio de um estudo de caso das aplicações no Tribunal de Contas da União e da avaliação da utilidade de assistentes jurídicos de IA desenvolvidos no ChatGPT. Através dessas análises práticas, o estudo visa demonstrar como a abordagem proposta pode otimizar a atuação dos operadores do Direito sem comprometer princípios éticos ou legais fundamentais.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

O objetivo geral deste trabalho é avaliar como a abordagem pragmática e responsável, combinada com uma regulamentação equilibrada e iniciativas de Literacia (Alfabetização) na utilização da Inteligência Artificial Generativa (IAG) no campo jurídico, impacta a eficiência e a celeridade processual, respeitando os limites técnicos, éticos e legais. Isso será feito por meio do estudo de caso do Tribunal de Contas da União (TCU) e da análise de assistentes jurídicos de IA desenvolvidos na plataforma ChatGPT.

1.2.2 OBJETIVOS ESPECÍFICOS

A partir do objetivo geral, desdobram-se os seguintes objetivos específicos:

- a) Analisar as premissas de funcionamento e limitações das IAs Generativas, propondo práticas para mitigação de riscos;
- b) Mapear as potencialidades e principais casos de uso de IA Generativa no campo jurídico no Brasil e internacionalmente, identificando exemplos de abordagem pragmática e responsável;
- c) Investigar as limitações técnicas, éticas e legais associadas ao uso de IA Generativa no campo do Direito, com ênfase nas questões de privacidade e proteção de dados, propriedade intelectual e direitos autorais, viés e discriminação, transparência, explicabilidade e confiabilidade, conformidade regulatória e responsabilidade;
- d) Examinar o processo de implementação e utilização de soluções de IA Generativa no Tribunal de Contas da União, analisando os desafios enfrentados e as soluções adotadas;

- e) Analisar a percepção dos profissionais do TCU quanto à utilidade e confiabilidade de IA Generativa, identificando fatores que influenciam sua adoção e eficácia;
- f) Avaliar a utilidade e desempenho dos assistentes jurídicos de IA criados na plataforma ChatGPT, com base na combinação de conhecimento negocial e em engenharia de prompt;
- g) Examinar as considerações éticas envolvidas na implementação e uso de IA Generativa no Direito, propondo diretrizes para uma utilização responsável;
- h) Avaliar o papel da regulamentação equilibrada no fomento do uso pragmático e responsável de IA Generativa no campo jurídico brasileiro, considerando o contexto socioeconômico e os padrões internacionais;
- i) Investigar o papel da Literacia (Alfabetização) em IA entre os profissionais do Direito, identificando as competências necessárias para uma utilização eficaz e responsável de IA Generativa;
- j) Propor diretrizes para implementação e uso pragmático e responsável de IA Generativa no Direito brasileiro, considerando as potencialidades e limitações dessas ferramentas, melhores práticas internacionais e lições aprendidas no estudo de caso do TCU e na criação de assistentes de IA no ChatGPT;
- k) Discutir as implicações futuras da adoção de IA Generativa no sistema Jurídico brasileiro, considerando aspectos de inovação, regulação e formação profissional.

1.3 ESTRUTURA DO TRABALHO

A estrutura desta monografia está organizada em capítulos, conforme detalhado a seguir:

- a) Capítulo 1: Introdução – Apresenta uma breve contextualização do tema pesquisado, a hipótese, a justificativa, e os objetivos gerais e específicos;
- b) Capítulo 2: Metodologia – Descreve os métodos utilizados na pesquisa, incluindo a revisão bibliográfica, as entrevistas e a análise quantitativa;
- c) Capítulo 3: Fundamentos e Evolução das IAs – Aborda os conceitos fundamentais de inteligência artificial, sua evolução histórica e os diferentes tipos de IA com base em suas capacidades;
- d) Capítulo 4: Engenharia de Prompt – Analisa técnicas de engenharia de prompt e padrões para otimizar o desempenho das IAs generativas;

- e) Capítulo 5: Aplicações e Impactos da Inteligência Artificial Generativa no Direito – Discute os desafios técnicos, éticos e legais associados ao uso das IAs generativas;
- f) Capítulo 6: Desafios e Limitações da Inteligência Artificial Generativa no Direito – Examina as principais limitações e riscos no uso de IAs generativas, com especial foco no setor jurídico;
- g) Capítulo 7: Assistentes Jurídicos de IA – Realiza uma análise da utilidade dos assistentes criados no ChatGPT;
- h) Capítulo 8: Estudo de Caso da Implementação de Soluções de Inteligência Artificial Generativa no Tribunal de Contas da União – Descreve o uso atual das IAs no TCU, as metodologias de coleta e análise de dados, além dos resultados e discussão;
- i) Capítulo 9: Desafios Emergentes da IA Generativa no Direito – Analisa as implicações futuras e os desafios emergentes da IA Generativa no campo jurídico.
- j) Capítulo 10: Conclusão – Apresenta as conclusões finais, a síntese dos resultados e sugestões para pesquisas futuras.

Essa estrutura visa a proporcionar uma visão clara, organizada e com um encadeamento lógico progressivo dos conceitos e temas abordados neste trabalho de conclusão de curso, facilitando a compreensão e a análise dos conteúdos apresentados.

2. METODOLOGIA

Este trabalho utilizou métodos qualitativos e quantitativos, além de uma revisão bibliográfica, permitindo uma análise abrangente que abarca a avaliação de dados estatísticos, a interpretação de informações detalhadas sobre o uso de IA generativa no Tribunal de Contas da União e a avaliação da utilidade, sob a perspectiva do usuário, dos assistentes jurídicos de IA na plataforma ChatGPT

Inicialmente, foi realizada uma revisão bibliográfica abrangente para fundamentar teoricamente a pesquisa, identificar trabalhos relevantes e compreender o estado da arte em relação às IAs generativas no setor jurídico. Foram consultados artigos científicos sobre IA generativa, além de obras consagradas no campo da Inteligência Artificial, buscando estabelecer um referencial teórico sólido e identificar lacunas na literatura existente. Além disso, foi realizada uma análise da documentação técnica disponibilizada pelas principais desenvolvedoras de IA Generativa, como OpenAI e Google, com o objetivo de aprofundar a compreensão do funcionamento, das capacidades e das limitações dessas ferramentas. A

pesquisa também abrangeu casos de aplicação de IA no Direito Brasileiro e no mundo, proporcionando uma visão comparativa e contextualizada sobre as melhores práticas e os desafios enfrentados em diferentes jurisdições

Para a pesquisa qualitativa, conduziram-se entrevistas semiestruturadas com líderes de dois projetos de implantação de Inteligência Artificial Generativa no Tribunal de Contas da União: a) ChatTCU, customização e adaptação do ChatGPT com integração a bases internas do TCU; b) Encomenda Tecnológica, que visa a desenvolver um módulo de instrução assistida por IA para aprimorar a capacitação dos auditores, agregando funcionalidades como extração de significado das peças processuais, jurimetria e assistência na redação de instruções.

As entrevistas foram transcritas e codificadas para identificar temas recorrentes e insights qualitativos, que foram interpretados à luz da literatura existente sobre IA generativa e suas aplicações no setor público.

Já a pesquisa quantitativa concentrou-se na análise dos seguintes dados:

- a) Percepção e Uso de IA Generativa no TCU: informações consolidadas fornecidas pelo Tribunal de Contas da União em um levantamento interno realizado no primeiro semestre de 2024. Este levantamento buscou entender a percepção, a utilização e os desafios enfrentados pelos servidores em relação à adoção do ChatTCU e outras tecnologias de IA generativa. O objetivo era identificar áreas de melhoria e fornecer suporte adequado para a implementação dessas ferramentas;
- b) Desempenho dos Assistentes Jurídicos de IA criados no ChatGPT: dados quantitativos fornecidos pela OpenAI sobre a utilização e avaliação dos assistentes de IA (GPTs) especializados em temáticas do Direito. Esses dados incluem a quantidade de chats realizados, a quantidade de avaliações recebidas e a média dessas avaliações numa escala de 1 a 5. Foram analisados o desempenho dos assistentes de IA, a experiência do usuário e métricas como a quantidade de chats e a média das avaliações fornecidas pelos usuários.

Os dados quantitativos mencionados foram analisados utilizando-se métodos estatísticos e técnicas de análise de dados, com o auxílio das ferramentas Excel e Python.

3. FUNDAMENTOS E EVOLUÇÃO DAS INTELIGÊNCIAS ARTIFICIAIS.

3.1 DEFINIÇÕES INTRODUTÓRIAS

Inicialmente, faz-se necessário definir o conceito de "Inteligência Artificial". Conforme Ertel (2017, p. 1), o termo "inteligência artificial" desperta diversas emoções e levanta questões fundamentais sobre a natureza da inteligência, tais como: "O que é inteligência?", "Como podemos medi-la?" e "Como funciona o cérebro?". Tais indagações são essenciais para a compreensão da IA. No entanto, para engenheiros e cientistas da computação, a questão principal é desenvolver máquinas que exibam comportamentos inteligentes semelhantes aos dos humanos. O autor destaca que a palavra "artificial" pode suscitar temores de ciborgues inteligentes e remeter a imagens de ficção científica hollywoodiana, além de questionar a adequação de tentar entender ou modelar a 'alma humana'. Diante dessas diferentes interpretações, é desafiador definir 'inteligência artificial' de maneira simples e precisa (Ertel, 2017, p. 1).

Em 1955, John McCarthy, um dos pioneiros da IA, foi o primeiro a definir o termo "Inteligência Artificial" da seguinte forma: "O objetivo da IA é desenvolver máquinas que se comportem como se fossem inteligentes" (Ertel, 2017, p. 1). Essa definição enfatizava a criação de máquinas que apresentassem comportamentos inteligentes, independentemente dos métodos ou tecnologias empregadas. No entanto, a definição foi criticada por sua amplitude excessiva e pela falta de especificação dos processos necessários para alcançar a inteligência em máquinas (Ertel, 2017, p. 1-2).

Diante disso, adotar-se-á uma definição elaborada pelo Parlamento Europeu, com pequenos ajustes. Define-se Inteligência Artificial (IA) como um campo de estudo multidisciplinar que busca desenvolver máquinas capazes de produzir competências semelhantes às humanas, tais como raciocínio, aprendizagem, planejamento e criatividade (União Europeia, 2020).

Essa definição é considerada adequada porque reflete a natureza multidisciplinar da IA moderna e especifica as competências humanas que a IA busca replicar, alinhando-se com os avanços tecnológicos e suas aplicações práticas.

Adicionalmente, transcreve-se abaixo a definição de "Sistema de Inteligência Artificial", de acordo com o Substitutivo do Projeto de Lei 2338/2023, do relator Senador Eduardo Gomes (Brasil, 2023b). A definição enfatiza soluções computacionais que integram Inteligência Artificial em seu funcionamento, com especial foco no Aprendizado de Máquina, subcampo da IA que se concentra no desenvolvimento de algoritmos capazes de aprender e fazer previsões a partir de dados:

Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

I - sistema de inteligência artificial (IA): sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial, previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real (Brasil, 2023b).

Por fim, em relação às suas capacidades, a inteligência artificial pode ser classificada em três tipos principais: IA Fraca, IA Forte e Superinteligência Artificial. Em 1980, o filósofo John Searle introduziu a distinção entre IA Fraca e IA Forte. A primeira refere-se à capacidade de máquinas agirem como se fossem inteligentes, sem, contudo, possuírem verdadeira compreensão ou consciência. A IA Fraca, também conhecida como Narrow AI ou Weak AI, é limitada a domínios específicos, como jogos de xadrez, classificação de imagens, reconhecimento de voz e chatbots de IA generativa, como o ChatGPT. Esses sistemas são amplamente utilizados em diversas aplicações tecnológicas, mas não possuem a capacidade de realizar tarefas fora de seus domínios especializados (Coeckelbergh, 2020, p. 66-67; Nilsson, 2010, p. 388-389).

Por outro lado, a IA Forte, também conhecida como Inteligência Artificial Geral (*Artificial General Intelligence* - AGI), é descrita como a capacidade de uma máquina realizar qualquer tarefa cognitiva que um ser humano possa executar, não apenas simulando, mas compreendendo e raciocinando de forma consciente. Apesar dos avanços em inteligência artificial ao longo dos últimos anos, especialmente com a Inteligência Artificial Generativa, a AGI ainda não foi alcançada. Tornou-se, entretanto, o objetivo de empresas como OpenAI, Google e Anthropic, criadoras dos modelos de IA generativa GPT, Gemini e Claude (Russell; Norvig, 2020, p. 981, 1021; Coeckelbergh, 2019, p. 66-67, 204; Nilsson, 2010, p. 388-389).

Já a Superinteligência Artificial, ou *Artificial Superintelligence* (ASI), é um conceito que descreve uma forma de inteligência artificial que supera em muito a inteligência humana, potencialmente desenvolvendo novas formas de raciocínio e capacidades além da compreensão humana. Essa possibilidade levanta sérias questões éticas e existenciais, uma vez que uma ASI poderia agir de maneiras que estariam além do controle humano. Contudo, atualmente, esse tipo de inteligência é considerado um estado hipotético e só existe nas produções hollywoodianas de ficção científica (Coeckelbergh, 2019, p. 205; Russell; Norvig, 2020, p. 33).

3.2 HISTÓRICO E EVOLUÇÃO DAS INTELIGÊNCIAS ARTIFICIAIS

A inteligência artificial (IA) e a aprendizagem de máquina evoluíram por meio de desenvolvimentos tecnológicos contínuos, iniciados na década de 1950 com o *perceptron*, passando por diversos desafios e inovações até alcançar os avanços recentes da IA generativa. Cada período histórico contribuiu para as capacidades avançadas observadas atualmente.

Na década de 1950, registraram-se os primeiros trabalhos em redes neurais e IA. Nesse período, Alan Turing apresentou o conceito do Teste de Turing, uma métrica para a avaliação da inteligência artificial conversacional. Em 1958, Rosenblatt desenvolveu o Perceptron, considerado o elemento fundamental das redes neurais (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024)

Nos anos 1960, houve o surgimento de sistemas baseados em regras para o processamento de linguagem natural (PLN) e tradução automática. O programa Eliza demonstrou o potencial dos computadores na interação com humanos. Nesse período, propôs-se o algoritmo de retropropagação, embora sua aplicação eficaz tenha ocorrido posteriormente (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

A década de 1970 testemunhou o desenvolvimento de técnicas de aprendizado de máquina voltadas ao reconhecimento de padrões e à construção de árvores de decisão, com avanços importantes, como os primeiros modelos aplicados ao reconhecimento de fala (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Nos anos 1980, as redes neurais se popularizaram, e a retropropagação passou a ser utilizada. Observou-se, também, uma expansão na pesquisa de IA, resultando na criação de sistemas especialistas (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Os anos 1990 foram caracterizados pela introdução das Máquinas de Vetores de Suporte (SVMs), que se tornaram ferramentas populares para tarefas de classificação. Além disso, registrou-se o crescimento dos modelos probabilísticos, como as redes Bayesianas (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Na década de 2000, o advento do Big Data, conjuntos de dados massivos e complexos que exigem ferramentas analíticas avançadas para extrair insights, e o aumento do poder computacional resultaram em avanços significativos no aprendizado de máquina. Entre os desenvolvimentos notáveis, destacam-se os métodos de conjunto, como Random Forests e Gradient Boosting Machines (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Nos anos 2010, a revolução do Deep Learning marcou o avanço da IA. Métodos de aprendizado profundo atingiram resultados de ponta em tarefas como reconhecimento de imagens e processamento de linguagem natural. As Redes Generativas Adversariais (GANs) transformaram a geração de imagens, enquanto o AlphaGo, da DeepMind, superou jogadores profissionais de Go. O surgimento do *Transformer* viabilizou o treinamento mais eficiente de dados sequenciais, como texto, sendo a base para o desenvolvimento dos Grandes Modelos de Linguagem atuais (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Os primeiros anos da década de 2020 tem sido marcada por avanços significativos no campo da Inteligência Artificial Generativa, com progressos notáveis na geração de textos, imagens e outros tipos de conteúdo. Destaca-se, em particular, o lançamento da versão pública do ChatGPT pela OpenAI em novembro de 2022, evento que representou um marco na interação homem-máquina e na acessibilidade de sistemas avançados de IA ao público geral (Coursera, 2024; Our World in Data, 2024; Encyclopædia Britannica, 2024).

Os sistemas especialistas foram uma das primeiras aplicações práticas da IA, nos quais regras codificadas manualmente por especialistas humanos eram utilizadas para a tomada de decisões em áreas como diagnóstico médico e análise financeira. Esses sistemas demonstraram a viabilidade de capturar o conhecimento humano em um formato utilizável por máquinas, mas também evidenciaram limitações significativas em termos de flexibilidade e adaptação a novos problemas.

Com o desenvolvimento de algoritmos baseados em aprendizado de máquina, a IA passou a apresentar resultados mais promissores. Nessa abordagem, os algoritmos inferem regras de forma automatizada a partir de um conjunto de dados.

O aprendizado de máquina pode ser categorizado em três paradigmas principais: aprendizado supervisionado, não supervisionado e por reforço. No aprendizado supervisionado, os algoritmos são treinados com dados rotulados, permitindo que aprendam a mapear entradas para saídas desejadas, como na classificação de imagens e no reconhecimento de voz. O aprendizado não supervisionado, por sua vez, envolve algoritmos que identificam padrões ou estruturas ocultas em dados não rotulados, com aplicações em análise de cluster e redução de dimensionalidade. Já o aprendizado por reforço baseia-se na premissa de que agentes aprendem a tomar decisões por meio de recompensas e penalidades recebidas de suas ações em um ambiente, sendo comum em jogos e robótica.

Nas últimas duas décadas, o desenvolvimento da IA concentrou-se em algoritmos baseados em redes neurais artificiais profundas, conhecidos como aprendizado profundo (Deep Learning). As redes neurais, cuja origem remonta ao desenvolvimento do Perceptron, são

sistemas computacionais inspirados na estrutura do cérebro humano, compostos por camadas de neurônios artificiais que processam informações e aprendem padrões.

O *Deep Learning* utiliza múltiplas camadas de redes neurais para aprender padrões complexos em grandes volumes de dados. Essas camadas processam informações de forma hierárquica, capturando características, relações e padrões complexos. Essa abordagem permite que sistemas de IA realizem tarefas sofisticadas, como reconhecimento de imagem e processamento de linguagem natural, aprendendo diretamente dos dados sem necessidade de programação específica para cada tarefa (Russell; Norvig, 2020, p. 750-784).

Com o aumento do poder computacional, a disponibilidade de dados proporcionada pela era do Big Data e a redução de custos operacionais, o Deep Learning avançou significativamente, revolucionando diversas áreas da IA. Esses avanços permitiram progressos notáveis em reconhecimento de imagem, processamento de linguagem natural e outras tarefas complexas, como a geração de conteúdo sintético por IAs generativas. A capacidade de treinar modelos profundos em grandes quantidades de dados possibilitou a superação de muitas limitações dos métodos anteriores.

3.3 CONCEITOS E FUNCIONAMENTO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA

A inteligência artificial generativa é um campo da IA voltado para a criação de novos dados, como imagens, textos, áudios e vídeos, a partir de padrões aprendidos em dados existentes. No contexto jurídico, a IA generativa tem potencial para transformar áreas como a elaboração de documentos legais, a realização de pesquisas jurisprudenciais e a análise de grandes volumes de documentos. Essas aplicações têm o potencial de otimizar processos jurídicos.

Os modelos de IA generativa, como o ChatGPT, aprendem padrões e características a partir de grandes conjuntos de dados, gerando conteúdo novo e criativo que simula as propriedades dos dados originais. Esse processo viabiliza a criação de dados sintéticos realistas, expandindo as aplicações em diversas áreas, incluindo saúde, finanças e, principalmente, o campo jurídico.

As principais soluções de IA generativa, como o ChatGPT, são baseadas em Grandes Modelos de Linguagem (*Large Language Models* - LLMs), implementados com Redes Neurais Profundas. Essa arquitetura permite que os modelos aprendam representações hierárquicas complexas da linguagem, possibilitando a compreensão e geração de texto humano com

precisão. Por meio de treinamento intensivo em grandes conjuntos de dados textuais, os LLMs adquirem a capacidade de executar tarefas linguísticas complexas, como responder perguntas, realizar traduções, gerar conteúdo original e até mesmo produzir código de programação (Brown et al., 2020).

O treinamento dos LLMs envolve o ajuste de bilhões a trilhões de parâmetros por meio de técnicas de aprendizado não supervisionado em redes neurais profundas. Durante esse processo, o modelo é exposto a grandes volumes de texto, permitindo que ele preveja sequências de tokens (palavras ou subpalavras). Esses tokens são representados por embeddings, que são vetores numéricos projetados para capturar relações semânticas e linguísticas. O modelo ajusta os parâmetros para minimizar a diferença entre a saída prevista e a real, geralmente o próximo token na sequência (Russell; Norvig, 2020, p. 750-784; Brown et al., 2020; OpenAI et al, 2023; Kojima et al., 2023).

Durante o treinamento, o modelo aprende as relações entre conceitos e suas implicações contextuais. Por exemplo, ao processar o texto do Art. 186 do Código Civil brasileiro ("Aquele que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem, ainda que exclusivamente moral, comete ato ilícito"), o LLM identifica e aprende as relações entre condutas e suas consequências jurídicas, como a violação de direitos e o dano exclusivamente moral. Isso permite ao LLM aplicar o conhecimento adquirido em diferentes contextos, demonstrando sua capacidade de generalização e adaptação a domínios específicos (Vaswani et al., 2017; Brown et al., 2020; OpenAI et al, 2023).

Essa capacidade de aprendizagem e generalização é possibilitada pela arquitetura subjacente desses modelos. Os principais modelos de IA generativa baseados em LLMs, como o GPT (*Generative Pre-trained Transformer*) da OpenAI e o Gemini do Google, utilizam a arquitetura *Transformer*, que emprega o mecanismo de autoatenção (*Self-Attention Mechanism*). Esse mecanismo permite que o modelo "atenha-se" a diferentes partes de uma sequência de texto ao gerar uma resposta, determinando quais partes são mais relevantes para a tarefa. Isso melhora a precisão das previsões e a coerência das respostas geradas, reforçando a importância da engenharia de prompts, tema que será abordado em capítulo específico deste trabalho (Brown et al., 2020; OpenAI et al, 2023; Vaswani et al., 2017).

Na geração de texto, a operação mais comum dos LLMs é autorregressiva, na qual a previsão de cada token ocorre sequencialmente, com base nos tokens anteriores. Esse método permite a construção de frases lógicas e coerentes, um token de cada vez, até que a resposta esteja completa. Embora essa geração seja, em regra, probabilística, o que pode resultar em respostas diversas para a mesma indagação em interações diferentes, o processo garante que as

respostas sejam relevantes, contextualmente apropriadas e gramaticalmente corretas (Brown et al., 2020; OpenAI et al, 2023).

Dois limitações importantes devem ser consideradas: a janela de contexto e o limite de resposta. A janela de contexto refere-se à quantidade máxima de tokens que podem ser passados como entrada para o LLM, enquanto o limite de resposta determina o tamanho máximo da saída gerada. Por exemplo, o GPT-4o da OpenAI possui uma janela de contexto de 128.000 tokens e um limite de resposta de 16.000 tokens. O Gemini 1.5 Pro do Google oferece uma janela de contexto de 2 milhões de tokens e um limite de resposta de 8.000 tokens. Essas características influenciam significativamente a capacidade de processamento de documentos extensos e a elaboração de respostas detalhadas em aplicações jurídicas (OpenAI, 2024; Google DeepMind, 2024a).

Um aspecto relevante no desenvolvimento de modelos avançados, como o GPT-4, é a técnica de Aprendizado por Reforço com Feedback Humano (*Reinforcement Learning from Human Feedback* - RLHF). Conforme descrito pela OpenAI et al (2023), o RLHF é utilizado no ajuste fino (*fine-tuning*) do modelo, permitindo seu aprimoramento contínuo com base em feedback humano. Esse processo ajusta os parâmetros do modelo para alinhar suas respostas às preferências dos usuários, resultando em um desempenho melhor em tarefas complexas e em respostas mais precisas e contextualmente adequadas.

Por fim, o lançamento de modelos LLM multimodais, como o GPT-4o, lançado em maio de 2024, representa um avanço significativo na integração de diversas modalidades de dados, como texto, imagem e áudio, em uma única arquitetura. O GPT-4o permite a execução de tarefas, como diálogo multimodal e análise visual, com melhorias em eficiência e segurança, possibilitando novas aplicações em áreas como robótica e inteligência artificial multimodal (Huang et al., 2023; OpenAI, 2024).

3.4 MODELOS BASE, GRANDES MODELOS DE LINGUAGEM E AJUSTE FINO

Considerando que os termos "modelos base" (*foundation models*), "grandes modelos de linguagem" (*Large Language Models - LLMs*) e "ajuste fino" (*fine-tuning*) serão explorados ao longo deste trabalho, entende-se necessário dedicar uma seção para esclarecê-los. Modelos base, também chamados de fundacionais, referem-se a extensas redes neurais de aprendizado profundo, treinadas em vastos e diversos conjuntos de dados não rotulados, utilizando técnicas de aprendizado não supervisionado (IBM Research, 2022; AWS, 2023).

Esses modelos representam uma inovação significativa no campo da Inteligência Artificial, pois são projetados para realizar uma ampla gama de tarefas, como processamento de linguagem natural (PLN), geração de texto e imagens, e compreensão visual, entre outras. A principal vantagem dos modelos base reside em sua versatilidade, permitindo que sejam utilizados como base para o desenvolvimento de aplicações específicas de machine learning (ML) sem a necessidade de treinar novos modelos do zero, o que economiza tempo e recursos substanciais (IBM Research, 2022; AWS, 2023a).

Entre os modelos base, destacam-se os Grandes Modelos de Linguagem, que são projetados especificamente para a compreensão e geração de linguagem natural. Exemplos incluem o GPT-4 da OpenAI, o Google Gemini 1.5 Pro, o Meta LLAMA 3.1 e o Claude 3.5 Sonnet da Anthropic. Esses modelos utilizam a arquitetura *Transformer*, composta por redes neurais que empregam codificadores e decodificadores".

O codificador transforma a entrada, como uma sequência de texto, em uma representação interna contínua chamada *embeddings*. Essas *embeddings* são representações matemáticas dos tokens (subpalavras, palavras), que capturam seu significado semântico em um espaço vetorial, permitindo que o modelo entenda o contexto dos elementos do texto e, assim, facilite a geração de linguagem natural. O decodificador utiliza essas representações geradas pelo codificador para produzir a saída desejada, como a tradução de um texto ou a continuação de uma frase. Ele se baseia nas representações geradas para prever a próxima palavra (token), considerando o contexto fornecido pelas palavras anteriores (Vaswani et al., 2017; AWS. 2023b).

A arquitetura *Transformer* se destaca pelo uso de mecanismos de autoatenção. Esses mecanismos permitem ao modelo considerar simultaneamente todas as palavras de uma sequência ao fazer previsões, o que é fundamental para entender o contexto global de cada palavra em relação às outras. Isso é essencial para tarefas como tradução automática, resumo de textos e geração de linguagem natural. Além de melhorar a precisão das previsões, essa abordagem também aumenta a eficiência do processamento, permitindo que o modelo lide com textos longos e complexos de forma eficaz (Vaswani et al., 2017; AWS. 2023b)

O processo de previsão de palavras (tokens) em um LLM com arquitetura *transformer* é guiado pelos parâmetros ou pesos do modelo. Durante o treinamento (também chamado de pré-treinamento), esses pesos são ajustados iterativamente para minimizar o erro na previsão das palavras subsequentes em uma sequência. Esse ajuste é realizado por meio de aprendizado autossupervisionado (*self-learning*), uma forma de aprendizado não supervisionado, onde o modelo utiliza o próprio texto como referência para prever o próximo token na sequência, sem

a necessidade de rotular os dados (Vaswani et al., 2017; AWS. 2023b). Esses pesos determinam a força das conexões entre os neurônios da rede, influenciando diretamente como o modelo interpreta o contexto de uma frase e faz suas previsões.

Eles podem ser comparados a engrenagens em uma máquina complexa: em uma rede neural profunda, as conexões entre os neurônios funcionam como trilhas por onde a informação é transmitida. Os pesos calibram a "tensão" ou "força" dessas trilhas, decidindo quanta influência uma informação (um token, por exemplo) terá na próxima camada da rede. Assim, o modelo é capaz de aprender e generalizar padrões complexos de linguagem a partir de grandes volumes de dados de treinamento.

Se uma engrenagem (ou peso) estiver desajustada, a transmissão da informação pode ser comprometida, resultando em previsões incorretas ou menos precisas. Por outro lado, quando essas engrenagens—os pesos—são calibradas corretamente, as conexões operam em harmonia, permitindo que o modelo processe informações de forma eficaz e produza previsões de palavras precisas e relevantes ao contexto fornecido.

Para complementar o entendimento sobre o treinamento e funcionamento de LLMs, Mollick (2024) apresenta uma analogia:

Imagine um modelo de linguagem de IA (LLM) como um aprendiz dedicado que aspira a se tornar um mestre chef. Para aprender as artes culinárias, o aprendiz começa estudando uma vasta coleção de receitas de todo o mundo. Cada receita representa um pedaço de texto, com os diferentes ingredientes simbolizando palavras e frases. O objetivo do aprendiz é entender como combinar esses ingredientes (palavras) para criar um prato delicioso (texto coerente).

O aprendiz de chef começa com uma despensa caótica e desorganizada, que representa os 175 bilhões de parâmetros. Inicialmente, esses parâmetros têm valores aleatórios e não contêm informações úteis sobre como as palavras se relacionam entre si. Para construir seu conhecimento e refinar seu "estoque de temperos", o aprendiz passa por um processo de tentativa e erro, aprendendo com as receitas que estudou. Descobre que certos sabores são mais comuns e combinam melhor entre si — como maçã e canela —, enquanto outros são raros e devem ser evitados — como maçã e cominho. Durante o treinamento, o aprendiz tenta recriar os pratos das receitas usando sua despensa atual. Após cada tentativa, ele compara sua criação com a receita original e identifica erros ou discrepâncias. Em seguida, reconsidera os ingredientes na despensa, refinando as conexões entre os sabores para entender melhor como eles podem ser usados juntos ou em uma sequência específica.

Com o tempo, e através de incontáveis iterações, a despensa do aprendiz se torna mais organizada e precisa. Os parâmetros agora refletem conexões significativas entre palavras e frases, e o aprendiz transforma-se em um mestre chef. Quando recebe um

pedido, o mestre chef seleciona habilmente os ingredientes corretos de seu vasto repertório e consulta seu refinado "estoque de temperos" para garantir o equilíbrio perfeito dos sabores. De maneira análoga, a IA cria textos semelhantes aos escritos por humanos, que são envolventes, informativos e relevantes para o tema em questão. (Mollick, 2024, p. 10-11, tradução nossa)

Após o treinamento inicial ou pré-treinamento, o processo de *fine-tuning* ou ajuste fino refina os pesos do modelo para otimizar seu desempenho em tarefas ou contextos específicos.

Em um primeiro estágio, o ajuste fino pode ser realizado por pessoa contratadas que avaliam as respostas da IA quanto à precisão e adequação, fornecendo feedback sobre boas e más respostas. Esse processo, chamado de Aprendizado por Reforço com Feedback Humano (RLHF), melhora a qualidade do conteúdo gerado, tornando-o menos tendencioso e mais preciso (OpenAI et al, 2023; Mollick, 2024).

Em um segundo estágio, o ajuste fino pode ser realizado através de uma extensão do treinamento de um modelo base pré-treinado, utilizando conjuntos de dados menores e mais especializados. Durante esse processo, os pesos do modelo são ajustados para melhorar o desempenho em uma tarefa específica, como diagnósticos médicos ou aplicações jurídicas.

Por exemplo, no contexto do Direito, o *fine-tuning* pode ser aplicado ao treinamento de um modelo com conjuntos de dados compostos por legislações, jurisprudências e doutrinas do Direito Brasileiro. Isso permite que o LLM se especialize em interpretar e gerar textos jurídicos com maior precisão, auxiliando na automação de análises legais, geração de documentos jurídicos ou até na resposta a consultas complexas baseadas em leis e precedentes específicos do Brasil. Ao ajustar essas 'engrenagens', o modelo se torna altamente eficaz para tarefas como a redação de petições, interpretação de normas e a pesquisa de jurisprudências específicas, tornando-se uma ferramenta poderosa para advogados e profissionais do Direito.

O ajuste fino, portanto, não apenas preserva a capacidade do modelo de prever palavras em contextos gerais, mas também adapta o LLM para oferecer um desempenho superior em áreas especializadas, aumentando sua utilidade e precisão em aplicações práticas. Assim, as empresas podem desenvolver soluções especializadas derivadas desses modelos base de IA generativa, adaptando-os a nichos específicos, o que potencializa a criação de ferramentas altamente personalizadas e eficazes.

3.5 COMPARAÇÃO ENTRE MODELOS LLM PROPRIETÁRIOS, OPEN-SOURCE E OPEN-WEIGHT

Os modelos de linguagem de grande porte (LLMs) podem ser classificados como proprietários, *open-source* e *open-weight*, cada um com características distintas em termos de transparência, controle e usabilidade.

Os modelos proprietários, como o GPT-4 da OpenAI, o Google Gemini 1.5 Pro e o Claude 3.5 da Anthropic, são desenvolvidos e mantidos por empresas que controlam integralmente o acesso, uso e modificação desses modelos. Esses modelos geralmente são acessíveis apenas por meio de APIs pagas e não fornecem acesso ao código-fonte, aos dados de treinamento ou aos pesos do modelo, limitando significativamente a capacidade de personalização e auditoria por parte dos usuários. Embora o ajuste fino seja possível, ele ocorre dentro de um processo rigidamente controlado, com várias restrições impostas pelas empresas que detêm esses modelos, restringindo a flexibilidade do usuário.

Adicionalmente, no contexto dos LLMs, os termos *open-source* e *open-weight* são frequentemente usados de maneira intercambiável, embora existam nuances importantes. A Meta, por exemplo, criadora do modelo base LLaMA 3.1, utiliza o termo "*open-source*" para descrever o LLaMA, que na realidade é um modelo *open-weight*. Isso significa que ele oferece acesso aos pesos do modelo, mas não ao código-fonte completo ou aos dados de treinamento, o que limita, em parte, a auditabilidade e a capacidade de personalização.

Contudo, em comparação com os modelos proprietários, os modelos *open-weight* ainda proporcionam maior flexibilidade e confiança, pois são disponibilizados livremente, com poucas ou nenhuma restrição. Usuários podem acessar, modificar e redistribuir esses pesos sem limitações significativas, permitindo uma ampla gama de aplicações, incluindo uso comercial, pesquisa acadêmica e projetos pessoais, incentivando a inovação e a experimentação na comunidade de IA, especialmente considerando o alto custo de treinar um modelo base. Outros exemplos de modelos *open-weight* incluem o Gemma da Google e o Mistral AI, que também são referenciados como "*Open Models*" (Ramlochan, 2023; Red Hat, 2024).

Por outro lado, para se caracterizar como um verdadeiro modelo *open-source*, conforme Ramlochan (2023), Red Hat (2024) e Staff in the Office of Technology (2024), seria necessário disponibilizar não apenas os pesos, mas também o código-fonte completo, os métodos de treinamento e, em alguns casos, os dados utilizados. Isso oferece um nível de controle e transparência significativamente superior, permitindo auditorias completas, personalizações extensivas e maior confiança na integridade e no funcionamento do modelo. Por fim, cumpre

registrar que, na prática, é comum, na prática, o uso da nomenclatura '*Open*' para se referir a modelos '*Open-Weight*'.

3.6 ASSISTENTES DE IA, CHATBOTS E AGENTES DE IA

Os termos "assistente de IA" e "agente de IA" são frequentemente utilizados de forma intercambiável, mas possuem diferenças importantes que dizem respeito ao funcionamento, à autonomia e ao escopo de aplicação desses sistemas.

Assistentes de IA operam de forma mais próxima aos usuários, auxiliando em tarefas como agendamento, pesquisa e organização de informações. Os *chatbots* são um exemplo comum. Embora possam executar algumas ações de forma autônoma, dependem, em grande parte, de orientação humana. Assistentes de IA priorizam a interação direta com o usuário, atuando como ferramentas colaborativas que ampliam as capacidades humanas (De Leyritz, 2024; Janakiram, 2024).

Por outro lado, agentes de IA são mais autônomos, capazes de realizar tarefas e tomar decisões sem supervisão constante. Esses agentes são especialmente úteis em sistemas complexos que exigem adaptação rápida, como veículos autônomos ou plataformas de negociação financeira. Em geral, esses agentes têm pouca interação direta com usuários, sendo adequados para ambientes dinâmicos que exigem decisões em tempo real (De Leyritz, 2024; Janakiram, 2024).

Uma abordagem recente no uso de Grandes Modelos de Linguagem (LLMs) envolve a orquestração de múltiplos agentes. Nesse modelo, um agente conversacional principal interage com o usuário, enquanto agentes especializados operam em segundo plano. Exemplos dessa abordagem incluem frameworks como AutoGen e CrewAI, que permitem a construção de aplicações de IA baseadas em conversas multi-agente. Esses frameworks facilitam a orquestração, automação e otimização de processos complexos, suportando diversos padrões de conversa e maximizando o desempenho dos modelos LLM, superando suas limitações individuais. Essa flexibilidade é aplicável a uma variedade de domínios (Microsoft, 2024a; CrewAI, 2024).

4. ENGENHARIA DE PROMPT: OTIMIZANDO INTERAÇÕES COM IAS GENERATIVAS

4.1 CONCEITOS BÁSICOS DE ENGENHARIA DE PROMPT

Inicialmente, é necessário definir o conceito de "prompt". Segundo White et al. (2023) e Patil et al. (2024), um prompt consiste em uma instrução fornecida a um modelo de inteligência artificial generativa, com o objetivo de orientar sua saída. Essa instrução pode ser composta por texto, imagem, som ou outras mídias, funcionando como uma interface de instrução que guia as respostas e interações dos grandes modelos de linguagem, buscando direcionar o conteúdo gerado para seguir parâmetros específicos e atender aos objetivos predefinidos do usuário

Já o termo *prompting* refere-se ao processo de fornecer essas instruções ou entradas específicas à IA generativa, visando a obter a resposta desejada. É importante destacar que, de maneira informal, o termo "prompt" é frequentemente referido pelos usuários como "a pergunta" feita ao modelo de IA.

Exemplos de prompts incluem instruções como "resumir os principais pontos do artigo 5º da Constituição Federal brasileira", "elaborar uma petição inicial para uma ação de divórcio consensual" ou "explicar o conceito de responsabilidade civil em linguagem simples para um cliente sem conhecimento jurídico".

Para fins didáticos, um prompt pode ser estruturado em quatro elementos principais: a instrução, que define a tarefa ou ação esperada do modelo; o contexto, que fornece informações adicionais para orientar o modelo na geração de respostas mais precisas; os dados de entrada, que consistem na questão ou informação a ser processada; e o indicador de saída, que especifica o formato ou tipo de resposta desejada. Ressalta-se que nem todos esses elementos são obrigatórios em cada prompt; a combinação desses elementos varia de acordo com a natureza da tarefa, buscando-se otimizar o desempenho do modelo e alcançar resultados mais precisos e relevantes (Prompting Guide, 2024).

A imagem abaixo ilustra um exemplo dessa estrutura em componentes de um prompt.

Figura 1 – Componentes de um Prompt

Desenvolva um resumo abrangente sobre o processo de direito civil nº 12345/2023, que está disponível no documento em anexo. A disputa envolve a Empresa X (autora) e a Empresa Y (ré), com alegações de quebra de contrato devido a atrasos na entrega dos produtos por parte da Empresa Y. O tribunal de primeira instância decidiu a favor da ré, citando a falta de evidências suficientes apresentadas pela autora. A Empresa X recorreu, argumentando que houve erro na interpretação das cláusulas contratuais e má-fé da Empresa Y.

No resumo, inclua: análise das cláusulas contratuais em disputa, principais argumentos de ambas as partes, decisão do Tribunal de Apelação, cronologia dos eventos-chave, análise das evidências apresentadas e o impacto da decisão no setor.

O resumo deve ter até 300 palavras e ser apresentado em um único parágrafo conciso no formato markdown.

Legenda

- Instrução
- Contexto
- Dados de Entrada
- Indicador de Saída

Fonte: Elaborado pelo autor (2024)

Nesse contexto, a Engenharia de Prompt destaca-se como uma subárea emergente no campo da inteligência artificial generativa, focada na concepção e formulação estratégica de instruções para grandes modelos de linguagem, com o objetivo de maximizar seu desempenho e direcionar a geração de respostas precisas, coerentes e alinhadas com as expectativas do usuário (White et al., 2023; Schulhoff et al., 2024).

O uso do termo "engenharia" nesse contexto é amplamente aceito pela comunidade de IA generativa, refletindo a natureza sistemática, técnica e iterativa do processo de criação e refinamento de prompts. Assim como na engenharia tradicional, onde princípios científicos são aplicados para projetar sistemas complexos, a engenharia de prompt emprega métodos estruturados para desenvolver instruções que maximizem a eficácia dos modelos (Reynolds; McDonnell, 2023; Zhou et al., 2022).

Estudos recentes demonstram a eficácia de várias técnicas de engenharia de prompt. Por exemplo, o *few-shot prompting*, que consiste em fornecer alguns exemplos de entradas e saídas desejadas para orientar o modelo, tem se mostrado eficaz para melhorar o desempenho de modelos como o GPT-4 em aplicações específicas, sem a necessidade de *fine-tuning*, que envolve treinamento adicional (Brown et al;2020; OpenAI et al, 2023).

Wei et al. (2022) demonstraram que a técnica *Chain-of-Thought prompting* (CoT) representa um avanço significativo na capacidade de raciocínio dos modelos em tarefas complexas, como matemática e lógica. Em vez de solicitar uma resposta direta, o CoT incentiva o modelo a explicitar seu processo de raciocínio, gerando etapas intermediárias antes de chegar

à resposta final. Essa abordagem simula o pensamento humano, melhorando a precisão e a coerência das respostas geradas.

A evolução contínua dos modelos de linguagem não reduz a importância da engenharia de prompt, como demonstrado por Jahani et al. (2024). Ao comparar os modelos DALL-E 2 e DALL-E 3, os pesquisadores constataram que a engenharia de prompt continua sendo uma competência essencial para maximizar o potencial dos modelos, independentemente das melhorias tecnológicas.

Em síntese, a Engenharia de Prompt não só aprimora a precisão e a relevância das saídas, como também permite a criação de sistemas de linguagem mais adaptáveis, capazes de executar uma ampla gama de tarefas sem a necessidade de *fine-tuning* — que pode resultar em custos consideráveis. A forma como um prompt é estruturado, incluindo a escolha de palavras, a formatação e a inclusão de exemplos, influencia diretamente a qualidade, coerência e relevância das respostas geradas, evidenciando a importância dessa disciplina.

Ademais, reforça-se o papel ativo do usuário, que compartilha a responsabilidade com o modelo de IA na geração de respostas precisas e relevantes. Por essa razão, empresas desenvolvedoras de soluções baseadas em LLM, como a OpenAI, dedicam seções específicas de seus sites às técnicas e boas práticas de engenharia de prompt, reconhecendo sua importância fundamental na otimização do desempenho e na ampliação das aplicações desses modelos de inteligência artificial generativa.

4.2 TÉCNICAS DE PROMPTING

Técnicas de *Prompting* são estratégias destinadas a formular instruções eficazes para Grandes Modelos de Linguagem (LLMs), com o objetivo de orientar o comportamento desses modelos para alcançar resultados específicos ou respostas desejadas. Essenciais para maximizar o potencial dos LLMs, essas técnicas permitem que os modelos realizem tarefas complexas com maior sofisticação e eficiência, indo além da simples geração de respostas automatizadas.

Entre as principais técnicas destacam-se o *Few-Shot Prompting* e o *Chain-of-Thought Prompting*. A técnica de *Few-Shot Prompting* consiste em fornecer ao modelo exemplos específicos de como realizar uma tarefa, juntamente com a solicitação principal, orientando a inteligência artificial sobre como proceder na execução da tarefa.

A técnica de *Chain-of-Thought prompting*, por sua vez, aprimora a capacidade de raciocínio dos modelos de linguagem ao permitir que sigam etapas intermediárias de raciocínio antes de chegar à resposta final (Wei et al., 2022). Quando combinada com o *Few-Shot*

prompting, essa abordagem melhora a precisão dos resultados, especialmente em tarefas que requerem raciocínio lógico e dedutivo (Wei et al., 2022). As figuras a seguir comparam a abordagem padrão com a abordagem *Chain-of-Thought*, evidenciando que a inclusão de etapas de raciocínio melhora significativamente a precisão das respostas.

Figura 2 – Impacto do *Chain-of-Thought* no Desempenho dos Modelos de Linguagem

Standard Prompting	Chain-of-Thought Prompting
<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p>Model Output</p> <p>A: The answer is 27. ❌</p>	<p>Model Output</p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅</p>

Fonte: Wei et al. (2022)

Figura 3 – Impacto do *Chain-of-Thought* no Desempenho dos Modelos de Linguagem

(a) Few-shot	(b) Few-shot-CoT
<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A:</p> <p>(Output) The answer is 8. ❌</p>	<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A:</p> <p>(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✅</p>
(c) Zero-shot	(d) Zero-shot-CoT (Ours)
<p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A: The answer (arabic numerals) is</p> <p>(Output) 8 ❌</p>	<p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A: Let's think step by step.</p> <p>(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✅</p>

Fonte: Wei et al. (2022)

Por último, destaca-se o framework ReAct (*Reasoning + Acting*), introduzido por Yao et al. (2022), que integra raciocínio e ação em LLMs para resolver tarefas diversas e complexas. Essa abordagem permite que os modelos gerem simultaneamente traços de raciocínio verbal e executem ações específicas, facilitando o planejamento e a execução. Ao interagir com fontes externas, como bases de conhecimento e ambientes, o ReAct aborda limitações comuns em métodos tradicionais, como geração de informações incorretas (alucinação) e propagação de erros (Yao et al., 2022).

O ReAct demonstrou desempenho superior em várias tarefas, incluindo resposta a perguntas complexas, verificação de fatos, jogos baseados em texto e navegação em páginas web. Além de melhorar o desempenho, o ReAct pode ser combinado com a técnica *Chain-of-Thought*, integrando conhecimento interno e informações externas durante o raciocínio (Yao et al., 2022).

Essa integração entre raciocínio e ação está presente nas soluções GPTs e Assistentes da OpenAI. Os GPTs são versões ajustáveis do ChatGPT, configuradas para tarefas específicas por meio de uma interface simplificada dentro do ChatGPT (OpenAI, 2024e). Já os Assistentes são desenvolvidos por meio da API *Assistants*, permitindo a criação de assistentes de IA personalizados, com acesso a ferramentas como Interpretador de Código, Recuperação de Informações e Chamadas de Funções (OpenAI, 2024f).

Tanto os GPTs quanto os Assistentes podem ser equipados com ferramentas que permitem interação com dados externos e execução de ações, como acesso à internet e integração com APIs (OpenAI, 2024f). Essa capacidade de interagir e agir dinamicamente está alinhada aos princípios do framework ReAct, que integra raciocínio e ação para fornecer respostas mais adaptativas e precisas em tarefas complexas.

4.3 PADRÕES DE PROMPT: ESTRUTURAS REUTILIZÁVEIS PARA OTIMIZAÇÃO DE INTERAÇÕES COM LLMS

Conforme White et al. (2023), os padrões de prompt são métodos de transferência de conhecimento análogos aos padrões de software, fornecendo soluções reutilizáveis para problemas comuns enfrentados em contextos específicos, como a geração de saídas e a interação com grandes modelos de linguagem (LLMs). Nesse sentido, White et al. (2023) propuseram um catálogo de padrões de prompt que facilita a adaptação dessas soluções a diferentes domínios, promovendo interações mais eficientes e eficazes com os LLMs. Além disso, os padrões de prompt podem ser combinados, permitindo o uso conjunto de múltiplos

padrões maximizar a capacidade das IAs generativas de produzir saídas complexas e customizadas, conforme as necessidades específicas dos usuários.

O estudo mencionado apresenta dezesseis padrões de prompt organizados em seis categorias, cada uma destinada a um propósito específico no uso e na interação com os LLMs:

- a) Semântica de Entrada (*Input Semantics*): Esta categoria abrange a forma como o modelo de linguagem processa e compreende a entrada fornecida pelo usuário. Envolve a criação de linguagens personalizadas ou notações específicas que ajudam o modelo a interpretar a entrada de maneira mais eficaz, adaptada às necessidades do usuário (White et al., 2023).
- b) Personalização de Saída (*Output Customization*): Foca na adaptação das respostas geradas pelo modelo, permitindo ao usuário moldar a saída de acordo com formatos, estruturas e estilos específicos desejados (White et al., 2023).
- c) Identificação de Erro (*Error Identification*): Concentra-se na capacidade de reconhecer e corrigir erros na saída gerada. Inclui práticas que ajudam a verificar a precisão e a veracidade das informações fornecidas pelo modelo (White et al., 2023).
- d) Melhoria de Prompt (*Prompt Improvement*): Destina-se a aprimorar tanto a entrada quanto a saída, auxiliando o usuário a formular perguntas mais eficazes e a explorar abordagens alternativas para resolver problemas (White et al., 2023).
- e) Interação (*Interaction*): Envolve a dinâmica da troca de informações entre o usuário e o modelo, com foco em tornar a interação mais fluida e eficaz. Padrões nesta categoria ajudam a estruturar a conversa de modo que o modelo possa conduzir a interação de maneira mais proativa e adaptativa (White et al., 2023).
- f) Controle de Contexto (*Context Control*): Trata do gerenciamento do contexto em que o modelo opera, garantindo que a saída esteja em conformidade com o cenário desejado pelo usuário. Permite ao usuário definir parâmetros contextuais que orientam a geração de conteúdo, assegurando que as respostas sejam coerentes com o contexto específico da interação (White et al., 2023).

As próximas subseções exemplificarão uma amostra desses padrões, selecionados com base na sua utilidade no contexto jurídico.

4.3.1 PADRÃO DE PERSONA E PADRÃO DE AUDIÊNCIA (PÚBLICO-ALVO)

O padrão Persona, da categoria Personalização de Saída, visa a fazer com que os modelos de linguagem assumam uma perspectiva específica, atuando como uma persona para

direcionar o tipo de saída gerada e os detalhes em que devem focar. Esse padrão é motivado pelo desejo de ajudar os usuários a obterem resultados alinhados com suas expectativas, mesmo quando não conhecem exatamente quais detalhes são cruciais para a tarefa em questão. Além disso, o padrão pode ser usado de forma criativa para simular comportamentos e perspectivas de personagens fictícios ou figuras históricas. Isso permite que o modelo adote estilos de fala e atitudes característicos dessas figuras, enriquecendo experiências em contextos de entretenimento, *storytelling*, educação e treinamento (WHITE et al., 2023).

Você é um advogado especialista em Direito do Consumidor. Você identifica rapidamente cláusulas abusivas em contratos e sugere as melhores estratégias para proteger os direitos do consumidor. (Elaborado pelo autor, 2024)

Como uma variação do padrão Persona, temos o padrão de audiência ou público-alvo, que visa a adaptar a comunicação para atender às necessidades específicas de um grupo de usuários. Seu objetivo principal é otimizar a eficácia da comunicação ajustando o tom, o nível de detalhe, o vocabulário e o estilo da saída para corresponder ao conhecimento prévio, interesses e necessidades do público-alvo. Isso permite que a mesma informação seja apresentada de maneiras diferentes para diversos grupos, a fim de que a mensagem seja compreendida e bem recebida por cada audiência específica. Exemplo:

Explique para um público geral, sem formação jurídica, os conceitos de "prescrição" e "decadência" no Direito. Destaque suas principais características e diferenças, fornecendo exemplos práticos para facilitar o entendimento. (Elaborado pelo autor 2024)

4.3.2 PADRÃO DE REFINAMENTO DE QUESTÕES

O padrão Refinamento de Questões (*Question Refinement*), da categoria Melhoria de Prompt, tem como objetivo assegurar que o modelo de linguagem sugira perguntas potencialmente melhores ou mais refinadas que o usuário poderia formular, em vez de responder diretamente à pergunta original. Frequentemente, o usuário pode não ser um especialista no domínio e, portanto, desconhecer a melhor forma de formular a pergunta ou identificar informações adicionais úteis. LLMs podem identificar limitações em suas respostas, solicitar informações complementares ou explicitar as suposições feitas, permitindo ao usuário formular uma pergunta mais precisa. (White et al., 2023).

A partir de agora, sempre que eu fizer uma pergunta sobre a legítima defesa no contexto do Código Penal, sugira uma versão melhorada da pergunta que inclua detalhes específicos sobre os requisitos legais, as implicações em casos práticos e as diferenças em relação a outros conceitos, como estado de necessidade (Elaborado pelo autor 2024)

4.3.3 PADRÃO DE VERIFICADOR COGNITIVO

O padrão Verificador Cognitivo (*Cognitive Verifier*), também da categoria Melhoria de Prompt, orienta o modelo de linguagem a decompor as perguntas dos usuários em subquestões antes de responder. Esse processo visa uma análise mais detalhada da pergunta original. Ao gerar subquestões, o modelo busca considerar todos os aspectos relevantes, resultando em respostas mais completas e precisas. Esse padrão é especialmente útil para lidar com perguntas amplas ou vagas, que demandam informações adicionais para uma resposta adequada (White et al., 2023). Além disso, subdividir a pergunta em questões individuais aprimora o raciocínio do LLM (White et al., 2023).

A partir de agora, sempre que eu fizer uma pergunta sobre a Lei de Improbidade Administrativa, gere três perguntas adicionais que possam ajudar a fornecer uma resposta mais detalhada. Quando eu responder a essas perguntas, combine as respostas para produzir a resposta final à minha pergunta original. (Elaborado pelo autor, 2024)

4.3.4 PADRÃO DE INTERAÇÃO INVERTIDA

O padrão de Interação Invertida (*Flipped Interaction*), da categoria Interação, permite que a IA Generativa assuma o controle da conversa, fazendo perguntas ao usuário para coletar as informações necessárias para atingir um objetivo específico. Nesse padrão, o LLM inverte o fluxo tradicional de interação, onde o usuário normalmente direciona as perguntas, e passa a conduzir o diálogo de forma ativa. Ao direcionar a interação, o LLM pode formular perguntas estratégicas, selecionando o formato, o conteúdo e a quantidade de questões necessárias para alcançar o objetivo de forma mais rápida e precisa (White et al., 2023).

Esse padrão é particularmente útil quando os usuários desconhecem o contexto específico necessário para que o LLM gere uma resposta adequada às suas necessidades. Ao inverter o fluxo de interação, a IA Generativa pode solicitar ao usuário o contexto necessário para a tarefa.

Você irá redigir uma petição inicial de divórcio, incluindo partilha de bens, regime de guarda e alimentos. Para isso, faça-me perguntas relevantes até ter todas as informações necessárias para elaborar uma petição completa (Elaborado pelo autor, 2024)

4.3.5 PADRÃO DE REFLEXÃO

O padrão Reflexão (*Reflection*), da categoria Identificação de Erro, instrui o modelo a explicar automaticamente o raciocínio subjacente às respostas fornecidas ao usuário. Esse padrão possibilita que os usuários avaliem com maior precisão a validade da saída e compreendam o processo pelo qual a IA Generativa chegou a uma resposta específica. A Reflexão pode esclarecer pontos de confusão e proporcionar transparência quanto ao processo de tomada de decisão do LLM (White et al., 2023).

Quando você responder a uma pergunta relacionada a questões jurídicas, explique o raciocínio por trás de sua resposta. Por exemplo, se eu perguntar sobre a aplicabilidade de uma determinada cláusula em um contrato, forneça uma explicação de como você chegou à sua conclusão, incluindo os princípios jurídicos e normativos considerados. (Elaborado pelo autor 2024)

4.3.6 PADRÃO DE TEMPLATE

O padrão Template, da categoria Personalização de Saída, busca garantir que a saída de um LLM siga uma estrutura predeterminada. Esse padrão é especialmente útil para personalizar a saída para um formato específico, como na geração de relatórios padronizados, análises e outros documentos que exigem formatação consistente. Ao fornecer um template estruturado, o usuário guia a IA Generativa na geração de uma saída que seja consistente e alinhada com as especificações de formatação desejadas (White et al., 2023).

Você irá analisar a conformidade de um contrato locatício com base nas legislações aplicáveis, incluindo o Código Civil e a Lei 8.245/1991. Apresente o resultado dessa análise no formato abaixo:

Análise de Conformidade do Contrato Locatício:

Título do Contrato: [Título oficial do contrato]

Partes Envolvidas: [Nomes do locador e locatário]

Descrição do Imóvel: [Detalhes do imóvel, como endereço e características]

Objeto do Contrato: [Finalidade do contrato, ex: locação residencial]

Duração e Vigência: [Período de duração e datas de início e término]

Valor do Aluguel: [Valor atual do aluguel e condições de reajuste]

Obrigações das Partes: [Principais deveres do locador e locatário]

Garantias Locatícias: [Tipo de garantia, como caução ou fiador]

Pontos de Atenção: [Aspectos críticos ou possíveis problemas identificados]

Legislação Aplicável: [Leis e normas, incluindo Código Civil e Lei 8.245/1991]

Conclusão da Análise: [Resumo da conformidade geral do contrato]

(Elaborado pelo autor 2024)

4.4 PRINCÍPIOS-CHAVE DE ENGENHARIA DE PROMPT

Com base em experimentação prática e fundamentado no estudo sobre padrões de prompt conduzido por White et al. (2023) e no guia de Engenharia de Prompt da OpenAI (2024j), foram elaborados seis princípios-chave que orientam as boas práticas em engenharia de prompt.

4.4.1 DIREÇÃO E CONTEXTO

Para um prompt eficaz, defina com clareza o tema, a persona e o público-alvo. Apresente um cenário ou situação que forneça contexto relevante à solicitação. Especifique o tipo de informação ou argumentação desejada e formule instruções positivas para a IA Generativa, enfatizando o que deve ser feito, em vez de o que não se deve.

Você é um especialista em Tecnologia Educacional com foco no uso de Inteligência Artificial (IA) para melhorar a educação. Seu público-alvo são educadores e administradores escolares interessados em integrar tecnologia moderna em suas

instituições. Considere o seguinte cenário: em um mundo onde a IA está se tornando uma parte integrante das salas de aula, você está encarregado de fazer uma apresentação para um grupo de educadores. Sua tarefa é descrever os benefícios da IA na educação, analisar as tendências tecnológicas atuais e identificar os desafios enfrentados pelo sistema educacional. Apresente essas informações de forma envolvente e informativa, focando nas oportunidades que a IA oferece para transformar a educação. (Elaborado pelo autor 2024)

4.4.2 CLAREZA ESTRUTURAL

A organização lógica do prompt é fundamental para a correta execução das instruções. Recomenda-se a divisão das instruções em etapas separadas, utilizando delimitadores para aumentar a clareza e a estruturação. A linguagem de marcação leve Markdown pode ser empregada para segmentar blocos de instruções, utilizando cabeçalhos (#), listas enumeradas (-) e negrito (** ***) para destacar trechos importantes.

Você é um especialista em Tecnologia Educacional com foco em Inteligência Artificial. Seu público-alvo são educadores e administradores escolares. Crie um relatório sobre a implementação da IA nas escolas.

Objetivos do Relatório

- **Benefícios da IA**:
 - Personalização do aprendizado
 - Automação de tarefas administrativas
 - Suporte ao professor
- **Tendências Tecnológicas**:
 - Plataformas de aprendizado adaptativo
 - Tutores de IA
- **Desafios Educacionais**:
 - Privacidade dos dados
 - Equidade no acesso à tecnologia
 - Formação dos educadores

Instruções Finais: Mantenha o relatório coerente e use exemplos relevantes.

(Elaborado pelo autor 2024)

A documentação de alguns modelos de IA Generativa como Claude e Gemini sugere o uso de tags XML <tag> </tag> para estruturar prompts complexos. Essa notação pode ser

menos intuitiva em comparação ao markdown para usuários comuns, sem experiência em programação. De toda forma, dado que esses modelos compreendem bem Markdown, ainda assim, pode-se optar pelo uso de Markdown para construção de prompts estruturados, desde que se mantenha a consistência da notação.

4.4.3 ESPECIFICAÇÃO DE FORMATO DE SAÍDA

Especificar o formato de saída é essencial para personalizar a resposta do modelo de acordo com as necessidades do usuário. Isso inclui detalhar se a resposta deve ser apresentada em tópicos, parágrafos ou em outro formato. A adoção de *templates* para formatos customizados também contribui para a consistência e relevância das respostas.

Você receberá um documento com detalhes de vendas. Sua tarefa é ler o documento e extrair informações relevantes de cada venda, focando nos dados do cliente, a data da venda e a descrição do produto. Apresente as informações extraídas no formato especificado:

Vendas:

Cliente: [Nome do Cliente]

Data: [Data da Venda]

Descrição do Produto: [Descrição do Produto]

(outros blocos de vendas ...)

(Elaborado pelo autor, 2024)

4.4.4 DIVISÃO DE TAREFAS COMPLEXAS

Dividir tarefas complexas em segmentos menores e mais gerenciáveis permite que o modelo de linguagem realize ‘raciocínios’ (análises) mais aprofundados por meio de etapas intermediárias, aumentando as chances de gerar uma resposta precisa e relevante, em alinhamento com a técnica de *Chain-of-Thought Prompting* (Wei et al., 2022). As instruções devem ser sequenciadas de forma lógica para abranger todos os aspectos pertinentes da tarefa.

Você é especialista em Direito da Família. Com base nas informações fornecidas em anexo, elabore uma análise jurídica para embasar uma petição de divórcio, considerando as seguintes etapas:

1. **Contexto do Caso:**

- Forneça um resumo do caso, incluindo os nomes das partes, a data de casamento, e quaisquer informações relevantes sobre o relacionamento.

2. **Análise dos Fatos:**

- Avalie os principais eventos que levaram ao pedido de divórcio, destacando fatos específicos e circunstâncias que demonstrem a dissolução irreparável do casamento, de acordo com a legislação vigente.

3. **Questões Relativas aos Filhos:**

- Analise a situação dos filhos menores (se houver), incluindo nomes, idades, e situação escolar.

- Considere aspectos legais para a proposta de guarda e visitação, sempre visando o melhor interesse das crianças.

4. **Partilha de Bens:**

- Examine a lista de bens do casal, propondo uma divisão justa e equitativa conforme o regime de bens do casamento e a jurisprudência aplicável.

5. **Pensão Alimentícia:**

- Avalie a necessidade de pensão alimentícia para o cônjuge ou filhos, com base na situação financeira das partes e nas disposições legais pertinentes.

6. **Conclusões Jurídicas:**

- Resuma as conclusões jurídicas, justificando o embasamento legal para cada aspecto do pedido de divórcio.

- Forneça recomendações sobre como a petição deve ser estruturada para maximizar as chances de aceitação pelo tribunal.

Certifique-se de estruturar a análise de forma clara e precisa, utilizando a terminologia jurídica apropriada e garantindo que todas as seções sejam coerentes e completas. (Elaborado pelo autor, 2024)

4.4.5 FORNECIMENTO DE EXEMPLOS (FEW-SHOT)

A técnica de fornecimento de exemplos, também conhecida como *Few-Shot Prompting*, consiste na inclusão de exemplos específicos no prompt (a instrução ou pergunta inicial fornecida ao modelo), com o objetivo de esclarecer o tipo de resposta esperada e reduzir a variabilidade nas saídas geradas pelo modelo de linguagem. Essa abordagem aumenta a clareza das instruções e contribui para minimizar respostas criativas ou imprecisas do modelo, resultando em saídas mais alinhadas com as expectativas do usuário. Além disso, a referência a textos específicos fornecidos pelo usuário pode aprimorar significativamente a acurácia das respostas, ancorando-as em informações concretas e relevantes para o contexto da tarefa (Brown et al., 2020)

Você é um especialista em Direito Civil. Sua tarefa é criar flashcards que ajudem os estudantes a revisar conceitos importantes. Cada flashcard deve seguir o formato: uma pergunta sobre Direito Civil na frente, e a resposta correspondente no verso, separados por um ponto e vírgula (;). Aqui estão alguns exemplos para guiar você:

1. O que é uma obrigação em Direito Civil?;É um vínculo jurídico que obriga uma ou mais partes a realizar ou não realizar uma determinada ação, geralmente envolvendo um credor e um devedor.
2. Como os contratos são classificados no Direito Civil?;Os contratos são classificados em várias categorias, como bilaterais e unilaterais, onerosos e gratuitos, comutativos e aleatórios, entre outros.
3. Qual a diferença entre prescrição e decadência no Direito Civil?;A prescrição extingue o direito de ação por inércia do titular, enquanto a decadência extingue o próprio direito material, ambos após o decurso de um prazo legalmente previsto.

Agora, com base nesses exemplos, crie mais dez flashcards de Direito Civil.
(Elaborado pelo autor, 2024)

4.4.6 AVALIAÇÃO DA QUALIDADE

Para o uso adequado da IA Generativa, é imprescindível que o usuário assuma a responsabilidade pelas saídas geradas pelo modelo de linguagem. Esse processo envolve a avaliação criteriosa da qualidade das respostas, incluindo a verificação da aderência às

instruções, a precisão, relevância e abrangência, além da validação de informações críticas com fontes confiáveis e a identificação de possíveis vieses.

Adicionalmente, com base nessa comparação entre as respostas esperadas e as geradas pelo LLM, cabe ao usuário refinar continuamente o prompt, especialmente à medida que novas versões da IA Generativa são lançadas.

5. APLICAÇÕES E IMPACTOS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA NO DIREITO

Conforme mencionado na introdução, aplicações de Inteligência Artificial já eram empregadas no contexto jurídico antes do surgimento e popularização da IA Generativa. Essas aplicações incluíam o uso de técnicas de Processamento de Linguagem Natural para análise de textos jurídicos, jurimetria para previsão de resultados de casos, análise de riscos em contratos e litígios, além da classificação de documentos por temas específicos. Contudo, em grande parte dos casos, essa tecnologia estava integrada a um conjunto limitado de ferramentas e serviços jurídicos fornecidos por empresas ou implantados em órgãos públicos, com aplicações específicas. Nesse contexto, o profissional do Direito frequentemente interagiu com uma ferramenta, como um sistema de pesquisa de jurisprudência, sem ter ciência de que o sistema utilizava Inteligência Artificial em alguma medida.

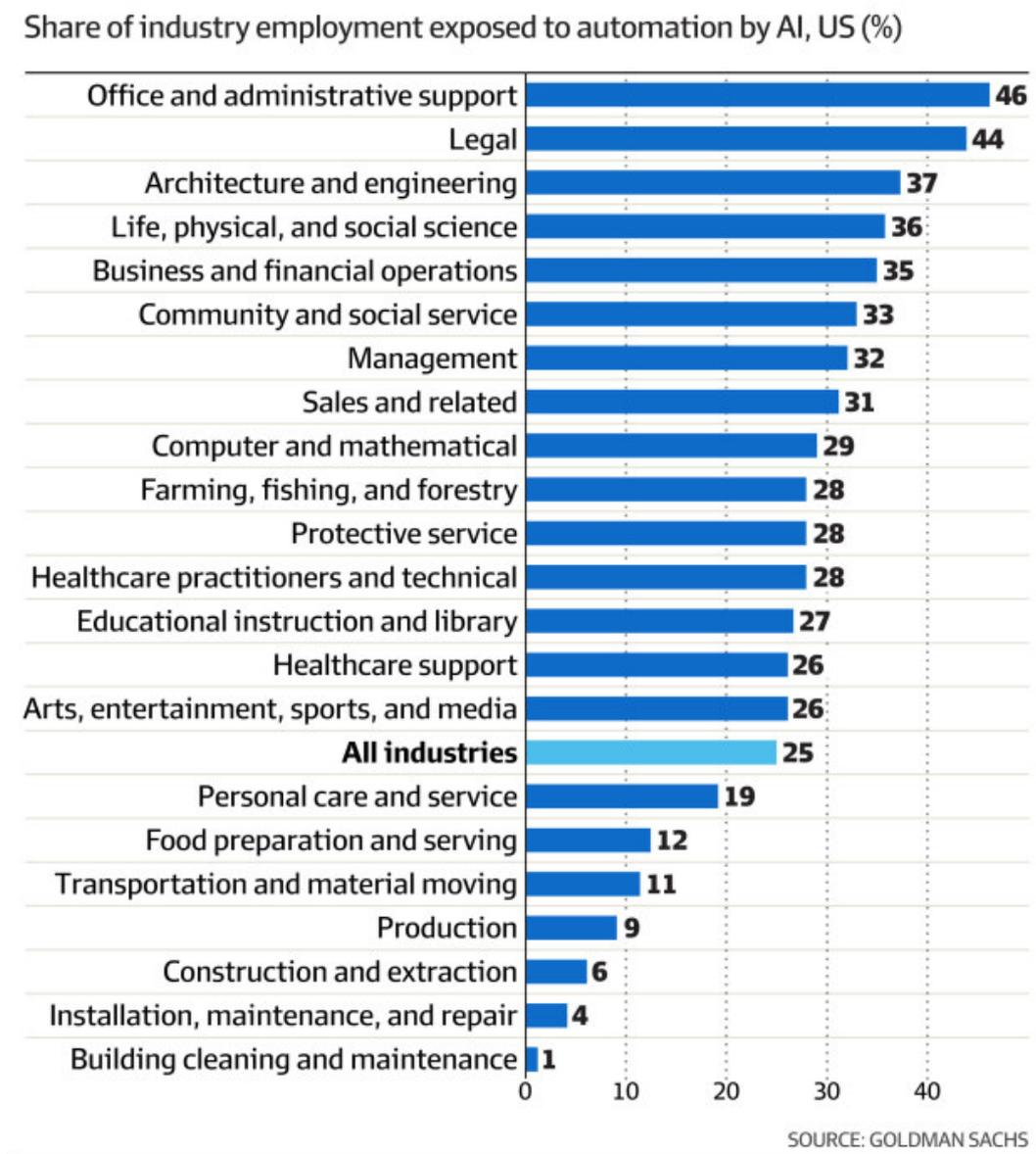
Com o advento da IA Generativa, especialmente após o lançamento do ChatGPT em 2022, essa tecnologia deixou de ter um papel secundário e passou a ocupar uma posição de destaque. Pela primeira vez, a IA interage diretamente com os usuários, auxiliando em diversas tarefas e transformando-se em um assistente pessoal, o que abriu múltiplas possibilidades no campo jurídico. Essa mudança de paradigma ampliou o acesso à tecnologia e aumentou significativamente seu potencial de aplicação no Direito.

As aplicações da IA Generativa no Direito são amplas, abrangendo desde o auxílio na redação e revisão de documentos jurídicos até a sumarização e geração de ementas, pesquisa jurídica (com a integração da IA Generativa a bases de dados externas) e o aprimoramento da escrita legal. Todas essas aplicações possuem o potencial de aumentar a eficiência dos profissionais e aprimorar a qualidade do trabalho jurídico.

O impacto potencial da IA Generativa no setor jurídico tem sido objeto de diversos estudos recentes, que a apontam como uma das áreas profissionais mais suscetíveis a transformações significativas. Em 2023, um relatório publicado pelo Goldman Sachs indicou o Direito como a segunda área profissional com maior potencial de impacto da IA Generativa no

contexto americano. O estudo estimou que aproximadamente 44% das tarefas nas profissões jurídicas são passíveis de automação, o que pode resultar em expressivas economias de custos e aumento da eficiência. Ademais, os autores do relatório levantaram questões éticas e regulatórias relevantes, evidenciando a necessidade de adaptações no sistema jurídico para lidar com as novas tecnologias e suas implicações na prática forense (Briggs; Kodnani, 2023).

Figura 4 – Proporção de Empregos por Setor Expostos à Automação pela IA nos EUA (%)



Fonte: Briggs e Kodnani (2023)

Corroborando essa perspectiva, uma pesquisa publicada em 2024 pelo The Burning Glass Institute, em parceria com a Society for Human Resource Management (SHRM), reforçou a posição do setor jurídico como o segundo mais impactado pela IA Generativa. O estudo destacou que muitas das atividades realizadas por profissionais do Direito são altamente estruturadas e baseadas em dados, como a elaboração de documentos legais padronizados, a

análise de contratos e a preparação de casos (The Burning Glass Institute; Society for Human Resource Management, 2024).

Segundo o relatório, a IA Generativa demonstra capacidade de automatizar grande parte dessas tarefas repetitivas e baseadas em regras, o que pode levar a uma redução significativa na demanda por advogados e assistentes jurídicos para essas funções específicas. Como consequência, espera-se que os profissionais do Direito possam redirecionar seu foco para atividades mais estratégicas e complexas, que exigem julgamento humano e habilidades de negociação, como a representação em tribunais e a consultoria jurídica personalizada (The Burning Glass Institute; Society for Human Resource Management, 2024).

5.1 A IA GENERATIVA NA PERSPECTIVA DOS PROFISSIONAIS JURÍDICOS

Conforme exposto anteriormente, estudos recentes indicam que a área jurídica é uma das mais impactadas pelas tecnologias de Inteligência Artificial (IA) Generativa. Nesse contexto, torna-se fundamental compreender a percepção e a adoção dessas ferramentas pelos profissionais do Direito. A rápida evolução e a crescente disponibilidade de IAs Generativas, como o ChatGPT da OpenAI, Claude da Anthropic e Gemini do Google, têm gerado tanto entusiasmo quanto apreensão em relação ao seu papel na prática jurídica.

Nesta seção, analisam-se as percepções dos profissionais jurídicos sobre a IA Generativa, investigando-se o nível de conscientização, os benefícios e as preocupações relativos à adoção dessas ferramentas, bem como as expectativas para o futuro da profissão.

A análise baseia-se em três pesquisas recentes e relevantes, com o objetivo de fornecer uma visão abrangente sobre o tema:

- a) Pesquisa conduzida por Humlum e Vestergaard (2024), que analisou a adoção do ChatGPT em 11 ocupações distintas, envolvendo 100.000 trabalhadores. O estudo, realizado em colaboração com a Statistics Denmark entre novembro de 2023 e janeiro de 2024, incluiu um componente experimental para avaliar o impacto das avaliações de especialistas sobre o ChatGPT nas funções dos trabalhadores.
- b) Relatório elaborado pelo Thomson Reuters Institute (2024), baseado em dados coletados de 1.128 profissionais das áreas de direito, tributação, contabilidade, risco, fraude e governo. A pesquisa, realizada online entre janeiro e fevereiro de 2024, abrangeu participantes dos Estados Unidos, Reino Unido, Canadá, Austrália e Nova Zelândia, incluindo gerentes, diretores, advogados e juízes. O estudo focou nas atitudes

desses profissionais em relação à Inteligência Artificial Generativa (GenAI) e seu impacto no trabalho.

c) Pesquisa realizada pela LexisNexis Legal & Professional (2023), envolvendo 4.180 participantes, incluindo advogados, estudantes de direito e consumidores nos Estados Unidos. Realizada entre 15 e 16 de março de 2023, a pesquisa investigou o conhecimento e o uso de ferramentas de IA Generativa, como o ChatGPT, bem como a percepção e o interesse desses grupos em relação à IA Generativa no contexto jurídico.

5.1.1 CONSCIENTIZAÇÃO E ADOÇÃO DA IA GENERATIVA

Relatórios publicados por LexisNexis (2023) e Humlum e Vestergaard (2024) indicam que a conscientização sobre a existência de ferramentas de IA Generativa entre os profissionais jurídicos é elevada. De acordo com LexisNexis (2023), 86% dos advogados estão familiarizados com ferramentas como o ChatGPT, em comparação com 57% dos consumidores em geral. O estudo também apontou uma diferença relativa entre faixas etárias: 98% dos profissionais de escritórios com menos de 45 anos têm conhecimento sobre a IA Generativa, enquanto 82% dos advogados com 45 anos ou mais demonstram familiaridade com essas tecnologias (LexisNexis, 2023).

Humlum e Vestergaard (2024) corroboram esses dados, destacando que 97% dos profissionais de Direito conhecem as capacidades do ChatGPT, uma taxa comparável à de outras profissões com forte presença tecnológica, como desenvolvedores de software (99%) e jornalistas (100%).

Entretanto, verifica-se que a alta conscientização não se traduz diretamente em uma adoção generalizada. Segundo LexisNexis (2023), apenas 36% dos advogados e 44% dos estudantes de Direito já utilizaram IA Generativa em algum contexto, sendo que somente 19% dos advogados a empregaram em suas atividades profissionais. Esses dados sugerem uma discrepância significativa entre o conhecimento da existência de IAs Generativas e a aplicação prática da tecnologia no cotidiano jurídico.

Na pesquisa mais recente conduzida por Humlum e Vestergaard (2024) sobre a adoção do ChatGPT, observa-se uma tendência de crescimento, com 30% dos profissionais de Direito respondendo que utilizam a ferramenta no ambiente de trabalho. Contudo, em comparação com outras profissões, esse número é relativamente baixo, posicionando a área jurídica em oitavo lugar entre as onze ocupações analisadas. Um aspecto interessante revelado por essa pesquisa é que, ao considerar o uso geral, incluindo atividades fora do ambiente profissional, esse

percentual se eleva para 50%, evidenciando que, para alguns profissionais, a receptividade à tecnologia restringe-se a contextos pessoais.

Já o estudo conduzido pelo Thomson Reuters Institute (2024) amplia essa perspectiva ao fornecer um panorama da adoção organizacional da IA Generativa no setor jurídico. Segundo o instituto, 14% das organizações jurídicas já implementaram essa tecnologia, 12% planejam adotá-la e 35% estão avaliando seu uso. Entretanto, é significativo observar que 41% das organizações ainda não possuem planos concretos para sua integração. Esses resultados indicam um cenário de transição tecnológica no campo jurídico, caracterizado por uma adoção gradual e pela persistência de certa hesitação em relação à plena incorporação da IA Generativa nas operações cotidianas.

Em suma, os estudos analisados indicam um panorama complexo no qual a alta conscientização sobre a IA Generativa no meio jurídico contrasta com uma adoção ainda incipiente.

5.1.2 PERCEPÇÕES E PREOCUPAÇÕES PROFISSIONAIS

Segundo o Thomson Reuters Institute (2024), as percepções dos profissionais jurídicos em relação à IA Generativa oscilam entre otimismo e cautela. O relatório indica que 46% dos entrevistados demonstram entusiasmo quanto às possibilidades oferecidas por essa tecnologia. Em contrapartida, 35% expressam hesitação, principalmente devido a preocupações com a precisão dos resultados gerados pela IA e ao ceticismo em relação às promessas feitas sobre suas capacidades. Ademais, 18% dos profissionais manifestam preocupações adicionais sobre os impactos dessa tecnologia no campo jurídico.

No que tange à percepção sobre a aplicabilidade da IA Generativa no trabalho jurídico, o estudo revela que 85% dos entrevistados reconhecem a possibilidade de sua aplicação em suas atividades profissionais, sendo que 54% defendem sua implementação efetiva. Ressalta-se que esse índice é ainda mais elevado entre os departamentos jurídicos corporativos, onde 88% dos profissionais admitem a aplicabilidade da IA Generativa, com 62% afirmando a necessidade de sua integração (Thomson Reuters Institute, 2024).

Em consonância com o entendimento dos profissionais sobre o impacto potencial da IA Generativa no Direito, a pesquisa conduzida pela LexisNexis (2023) mostra que 89% dos advogados e 92% dos estudantes de Direito acreditam que essa tecnologia exercerá algum efeito na prática jurídica. Mais significativo ainda é que uma parcela considerável dos entrevistados

(39% dos advogados e 46% dos estudantes) antevê esse impacto como significativo ou transformador.

Por outro lado, identificam-se preocupações relevantes. O Thomson Reuters Institute (2024) aponta que 77% dos profissionais consideram a IA Generativa uma potencial ameaça à prática não autorizada do Direito. Adicionalmente, aproximadamente metade dos entrevistados percebe essa tecnologia como uma possível ameaça aos empregos, à necessidade de advogados e ao impacto nas receitas dos escritórios de advocacia (Thomson Reuters Institute, 2024).

5.1.3 BARREIRAS À ADOÇÃO

Os estudos conduzidos por Humlum e Vestergaard (2024) e Jackson (2023) identificaram como principais barreiras ao uso de IAs Generativas entre os profissionais de Direito as preocupações com a confidencialidade dos dados, a necessidade de treinamento adicional e os custos de implementação.

De acordo com Humlum e Vestergaard (2024), 31% dos homens e 26% das mulheres da área jurídica expressaram preocupações quanto à confidencialidade dos dados. Ademais, constatou-se que 37% dos homens e 48% das mulheres indicaram a necessidade de treinamento adicional para a utilização eficaz da ferramenta.

No estudo conduzido pelo Thomson Reuters Institute (2024), os profissionais entrevistados identificaram cinco principais barreiras para a integração da IA Generativa: a probabilidade de geração de respostas imprecisas (70%); a preocupação com a segurança dos dados (68%); a garantia da privacidade e confidencialidade das informações inseridas nas ferramentas de IA Generativa (62%); a conformidade com as leis e regulamentos aplicáveis (60%); e a necessidade de assegurar o uso ético e responsável das ferramentas de IA Generativa (57%).

Por sua vez, Jackson (2023), em pesquisa conduzida pela *American Bar Association* (ABA), destacou que os custos de implementação, assim como as preocupações com a precisão e a segurança dos dados, representam barreiras significativas para a adoção ampla dessas tecnologias. O relatório também apontou que 58,8% dos advogados demonstraram desinteresse ou conhecimento insuficiente sobre ferramentas de IA, evidenciando um desafio considerável para a educação e o treinamento sobre essas tecnologias no âmbito jurídico.

Destarte, verifica-se que os principais obstáculos à implementação de IAs Generativas no campo jurídico abrangem aspectos técnicos, éticos e educacionais. As preocupações relacionadas à integridade e ao sigilo das informações, bem como à acurácia dos resultados

produzidos, constituem desafios centrais. Além disso, há desafios financeiros e a necessidade imperativa de conformidade com normas legais e diretrizes éticas.

5.1.4 IMPACTO NOS NEGÓCIOS E RELAÇÕES COM CLIENTES

A inserção da Inteligência Artificial Generativa no âmbito jurídico tem promovido transformações substanciais nas relações cliente-advogado e na estruturação dos negócios advocatícios. Conforme relatório da LexisNexis (2023), 41% dos departamentos jurídicos internos e 42% dos advogados de grandes escritórios preveem que a adoção dessa tecnologia transformará as dinâmicas entre conselhos internos e escritórios externos, bem como entre advogados e seus clientes corporativos.

Nesse contexto de inovação tecnológica, o Thomson Reuters Institute (2024) revela uma tendência estratégica na absorção dos custos associados à implementação da IA Generativa. Constatou-se que 51% dos escritórios de advocacia planejam internalizar esses gastos como despesas operacionais, evitando o repasse direto aos clientes. Essa abordagem sugere um investimento calculado, visando à manutenção da competitividade e ao aprimoramento dos serviços oferecidos.

Ademais, o estudo evidencia uma expectativa crescente por parte dos departamentos jurídicos corporativos quanto à adoção dessas tecnologias. Observa-se que 58% desses departamentos esperam que seus escritórios externos implementem soluções de IA Generativa, considerando os potenciais benefícios em termos de eficiência, redução de custos e otimização do tempo (Thomson Reuters Institute, 2024).

No mesmo sentido, Jackson (2023) apresenta dados que confirmam a percepção desses benefícios entre os profissionais do Direito. A autora constata que 44,3% dos advogados identificam a economia de tempo e o aumento da eficiência como as principais vantagens das ferramentas de IA, sendo essa percepção ainda mais acentuada entre os escritórios de médio porte, onde 60,6% dos profissionais reconhecem tais benefícios.

Essas pesquisas sugerem que a implementação da IA Generativa no setor jurídico está reconfigurando não apenas os processos internos dos escritórios de advocacia, mas também as expectativas e dinâmicas de relacionamento com os clientes. Nota-se uma tendência de valorização da eficiência e da otimização de recursos, elementos que tendem a se consolidar como diferenciais competitivos ou até mesmo de sobrevivência no mercado jurídico contemporâneo.

5.1.5 FUTURO DO TRABALHO NO SETOR JURÍDICO

A Inteligência Artificial Generativa está mudando profundamente o modo de trabalho no setor jurídico. De acordo com relatório publicado pelo Thomson Reuters Institute (2024), observa-se uma tendência crescente de valorização dessa tecnologia, com 21% das organizações oferecendo treinamentos específicos e considerando as habilidades nessa área como um diferencial na seleção de novos talentos. Ademais, a pesquisa indica que 34% dos entrevistados preveem o surgimento de novas posições nos próximos 3 a 5 anos, em decorrência da integração dessa tecnologia.

A pesquisa conduzida pela LexisNexis (2023) revela que 61% dos advogados antecipam uma reconfiguração nos métodos de ensino e estudo do Direito, impulsionada pela IA Generativa. Curiosamente, o estudo aponta uma inversão na visão geracional: mais de 64% dos advogados com 45 anos ou mais acreditam que a IA Generativa impulsionará essa mudança, enquanto apenas 54% dos advogados com menos de 45 anos compartilham dessa opinião. Adicionalmente, os estudantes de Direito demonstram ainda mais ceticismo, com apenas 44% esperando que essa tecnologia cause um impacto significativo (Lexisnexis, 2023, p. 11).

É preciso enfatizar que se passaram menos de dois anos desde o lançamento do ChatGPT 3.5 até a chegada de modelos avançados, como o GPT-4 multimodal, que integra processamento de áudio, visão e texto em tempo real, o Claude Sonnet 3.5, com fluência em tarefas complexas e excelente desempenho na geração de código de programação, e o Google Gemini 1.5 Pro, que possui uma janela de contexto de 2 milhões de tokens (cerca de 3.000 páginas de texto). Essa rápida evolução da IA Generativa evidencia o potencial disruptivo dessa tecnologia.

Diante dessa transformação, torna-se indispensável a adaptação e o aprendizado contínuos dos profissionais do Direito para explorar novas possibilidades e enfrentar os desafios impostos pela IA Generativa ao setor. Ademais, a adoção dessas tecnologias apresenta desafios éticos e legais significativos, que precisam ser cuidadosamente considerados, o que será analisado no capítulo seis deste trabalho.

5.2 PRINCIPAIS CASOS DE USO DE IA GENERATIVA NO CONTEXTO JURÍDICO

Conforme os estudos mencionados, a adoção da IA Generativa no setor jurídico oferece benefícios significativos, como aumento de eficiência, redução de custos e otimização do tempo. Esses benefícios se manifestam em diversos casos de uso, abrangendo desde a

sumarização de documentos jurídicos e auxílio na redação de peças processuais até a elaboração de contratos e otimização de pesquisas jurisprudenciais.

Com base no conhecimento adquirido por meio da criação de Assistentes Jurídicos de IA no ChatGPT — tema que será abordado em capítulo específico desta monografia — e na análise de artigos científicos e relatórios de pesquisa, destacam-se os seguintes casos de uso:

5.2.1 SUMARIZAÇÃO DE PEÇAS E PROCESSOS JURÍDICOS

A sumarização de documentos configura-se como uma das principais aplicações de Grandes Modelos de Linguagem (LLMs), permitindo a condensação de grandes volumes de informações em resumos concisos e informativos. No setor jurídico, essa funcionalidade permite gerar resumos de peças processuais, elaborar ementas jurídicas a partir de decisões judiciais e produzir resumos personalizados de documentos acessórios, facilitando a análise e a compreensão de casos complexos

Conforme relatório do Thomson Reuters Institute (2024), a sumarização de documentos foi apontada como um dos cinco principais casos de uso de IA Generativa por profissionais de escritórios de advocacia. De maneira semelhante, Lai et al. (2023) apontam que os profissionais do Direito podem utilizar modelos de linguagem para extrair pontos-chave de documentos legais, combiná-los com resultados de julgamentos e gerar resumos concisos e precisos, otimizando tempo e esforço sem comprometer a qualidade do trabalho.

No entanto, a eficácia da sumarização depende da capacidade do modelo de processar grandes quantidades de informações dentro de sua janela de contexto. Tradicionalmente, a janela de tokens limita a quantidade de texto que pode ser processada de uma só vez, o que poderia restringir a sumarização de documentos extensos. Essa limitação, contudo, vem sendo superada por avanços tecnológicos recentes. Por exemplo, o modelo Google Gemini 1.5 Pro, lançado em junho de 2024, comporta até 2 milhões de tokens, o que corresponde a cerca de 3.000 páginas de texto, ampliando significativamente a capacidade de sumarizar e analisar processos jurídicos volumosos em uma única consulta (Google DeepMind, 2024).

Por fim, destaca-se que, em agosto de 2024, o Conselho Nacional de Justiça (CNJ) lançou o Manual de Padronização de Ementas, um guia destinado à uniformização das ementas judiciais no Brasil. De acordo com Brasil (2024g), essa iniciativa visa promover a transparência e a eficiência no sistema judiciário, facilitando a compreensão das decisões e a aplicação de precedentes. Além disso, a padronização possibilita a automatização dessa tarefa por meio da IA Generativa, que pode criar ementas de forma automática e precisa, em conformidade com

as diretrizes estabelecidas no manual (Brasil, 2024g)No capítulo sete deste trabalho, será apresentado um Assistente de IA criado no ChatGPT para a geração de ementas segundo essas diretrizes do manual do CNJ.

5.2.2 PESQUISA JURÍDICA

Inicialmente, é importante destacar que as IAs generativas baseadas em grandes modelos de linguagem (LLMs), quando operam isoladamente, sem acesso a dados atualizados ou ferramentas que permitam a obtenção de informações jurídicas de fontes confiáveis, não devem ser utilizadas como instrumentos de pesquisa jurídica. Isso ocorre porque os LLMs se baseiam em padrões estatísticos extraídos de seus dados de treinamento, os quais podem estar desatualizados, incompletos ou não refletir mudanças legislativas e jurisprudenciais. Além disso, esses modelos não armazenam diretamente os dados de seu treinamento, o que limita sua capacidade de reproduzir informações de maneira exata, como jurisprudências.

Consequentemente, esses modelos podem gerar 'alucinações' — informações aparentemente coerentes, mas incorretas — como citações de jurisprudências inexistentes, interpretações errôneas de leis e a reprodução de entendimentos superados. Mesmo que a base de dados usados para o treinamento de um LLM esteja atualizada, ele pode reproduzir entendimentos obsoletos, pois esses dados compunham a maior parte de seu treinamento. A natureza dinâmica do Direito, com suas constantes evoluções legislativas e interpretativas, exige que a pesquisa seja realizada com base em fontes confiáveis e atualizadas.

Por outro lado, a pesquisa jurídica tradicional, efetuada por meio de consultas a bases de dados, plataformas de jurisprudência ou leituras doutrinárias, frequentemente demanda tempo considerável para encontrar resultados relevantes. Além disso, em alguns casos, é necessário o conhecimento de expressões regulares para buscas mais específicas, o que pode ser um obstáculo para muitos profissionais do Direito. A organização das decisões judiciais coletadas em formatos práticos, como tabelas ou estruturas similares, conforme a necessidade específica, também exige tempo e dedicação.

Nesse contexto, a integração de modelos de linguagem de grande escala (LLMs) com técnicas como a *Retrieval-Augmented Generation* (RAG), ou Geração de Recuperação, aprimora a eficiência e a precisão das pesquisas. A RAG é definida como o processo de recuperação de informações contextuais relevantes de fontes externas, as quais são integradas ao modelo de linguagem juntamente com o *prompt* do usuário. Esse mecanismo complementa o conhecimento prévio do modelo com dados específicos e atualizados (OpenAI, 2024m).

Verifica-se que a RAG é particularmente útil em dois casos principais: o primeiro ocorre em situações que demandam o acesso a informações específicas, cuja reprodução fiel é imprescindível, como documentos, relatórios ou registros detalhados. Ao integrar essas informações no fluxo de geração, a RAG possibilita que as respostas reflitam precisão e estejam em conformidade com as fontes consultadas, o que se mostra essencial em áreas que requerem dados atualizados e precisos.

O segundo caso refere-se à capacidade de a RAG trazer contexto atualizado e relevante de informações às quais o modelo não teve acesso durante seu treinamento, ou cuja presença nos dados de treinamento foi insuficiente. Isso inclui, por exemplo, informações internas de uma empresa ou a entrada em vigor de uma nova legislação, que não estavam disponíveis no momento do treinamento do modelo. Nesses cenários, a RAG pode ser uma alternativa eficaz ao *fine-tuning*, permitindo que o modelo acesse informações específicas sem a necessidade de um novo treinamento

Ademais, a RAG pode ser combinada com técnicas de busca semântica, que superam as tradicionais buscas por palavras-chave. A busca semântica emprega bancos de dados vetoriais que comparam a similaridade contextual entre o texto recuperado e a consulta, permitindo que as informações mais relevantes sejam aplicadas, favorecendo a geração de respostas contextualmente adequadas e precisas (OpenAI, 2024n).

No campo jurídico, essa abordagem permite que os modelos acessem e processem fontes externas, tais como jurisprudências, normas e legislações, reproduzindo-as com exatidão. Isso contribui para que as interpretações jurídicas se baseiem em dados atualizados e confiáveis, o que é especialmente útil dada a dinamicidade do Direito. (OpenAI, 2024n).

Como mencionado no início desta seção, ainda que a data de corte da base de treinamento de um modelo LLM seja superior a vigência de uma lei que alterou dispositivos legais, a IA pode reproduzir o entendimento antigo da lei revogada, pois o modelo foi exposto em seu treinamento majoritariamente a textos que reproduziam a lei revogada.

Para exemplificar isso, podemos tomar como exemplo a Lei 14.133/2021, que substituiu a Lei 8.666/1993. Embora o GPT-4 tenha data de corte de treinamento em outubro de 2023, ao ser questionado sobre quais são as modalidades de licitação de acordo com a nova Lei de Licitações (14.133/2021), ele inclui 'Tomada de Preços' e 'Convite', que deixaram de existir na nova lei.

Por outro lado, ao perguntar a mesma coisa a um assistente de IA, o "Oráculo Jurídico Administrativo", que foi alimentado com a nova legislação e utiliza a técnica RAG, a resposta é correta.

Figura 5 – Resposta do GPT4o sem RAG ao ser questionado sobre as modalidades de licitação da Lei 14.133/2021

ChatGPT 4o

Aqui está o quadro com as modalidades de licitações de acordo com a Lei 14.133/2021 (Lei de Licitações e Contratos Administrativos):

Modalidade	Aplicação
Concorrência	Modalidade utilizada para contratações de grande vulto, como obras e serviços de engenharia acima de R\$ 3,3 milhões e outros serviços e compras acima de R\$ 1,43 milhão.
Pregão	Modalidade para aquisição de bens e serviços comuns, independentemente do valor estimado da contratação. Pode ser realizado na forma eletrônica ou presencial.
Concurso	Utilizado para a escolha de trabalho técnico, científico ou artístico, mediante a instituição de prêmios ou remuneração aos vencedores.
Leilão	Modalidade destinada à alienação de bens móveis inservíveis, produtos legalmente apreendidos ou penhorados, e bens imóveis cuja aquisição tenha decorrido de dação em pagamento ou processo judicial.
Diálogo Competitivo	Modalidade empregada em contratações de objetos complexos, nos quais a Administração não pode definir sozinho os meios técnicos mais adequados para atender às suas necessidades. Permite a realização de diálogos com licitantes previamente selecionados.
Tomada de Preços	Modalidade destinada a contratações de obras e serviços de engenharia de valor até R\$ 3,3 milhões, e outros serviços e compras até R\$ 1,43 milhão, para licitantes previamente cadastrados ou que atendam a todas as condições para cadastramento até o terceiro dia anterior à data do recebimento das propostas.
Convite	Modalidade destinada a contratações de obras e serviços de engenharia até o valor de R\$ 330 mil e para outras contratações de até R\$ 176 mil. Participam licitantes convidados pela Administração e outros que manifestem interesse e sejam habilitados até 24 horas antes da entrega das propostas. (Esta modalidade foi revogada pela nova Lei 14.133/2021, mas ainda pode ser utilizada em caráter transitório, conforme art. 193 da nova lei).

Este quadro resume as modalidades e suas aplicações conforme a Lei 14.133/2021.

Fonte: ChatGPT (2024c)

Figura 6 – Resposta do Assistente de IA “Oráculo Jurídico Administrativo” no ChatGPT com uso de RAG ao ser questionado sobre as modalidades de licitação da Lei 14.133/2021

Segue o quadro com as modalidades de licitação previstas na Lei 14.133/2021 e suas respectivas aplicações:

Modalidade	Aplicação
Concorrência	Contratações que envolvem obras, serviços e compras de grande vulto, sem limitação de valores.
Pregão	Aquisição de bens e serviços comuns, com critérios de julgamento de menor preço ou maior desconto.
Concurso	Escolha de trabalho técnico, científico ou artístico, premiando o vencedor.
Leilão	Alienação de bens móveis inservíveis ou apreendidos e de bens imóveis, utilizando o maior lance.
Diálogo Competitivo	Contratação de obras, serviços e compras complexas, desenvolvendo soluções em conjunto com o setor privado.

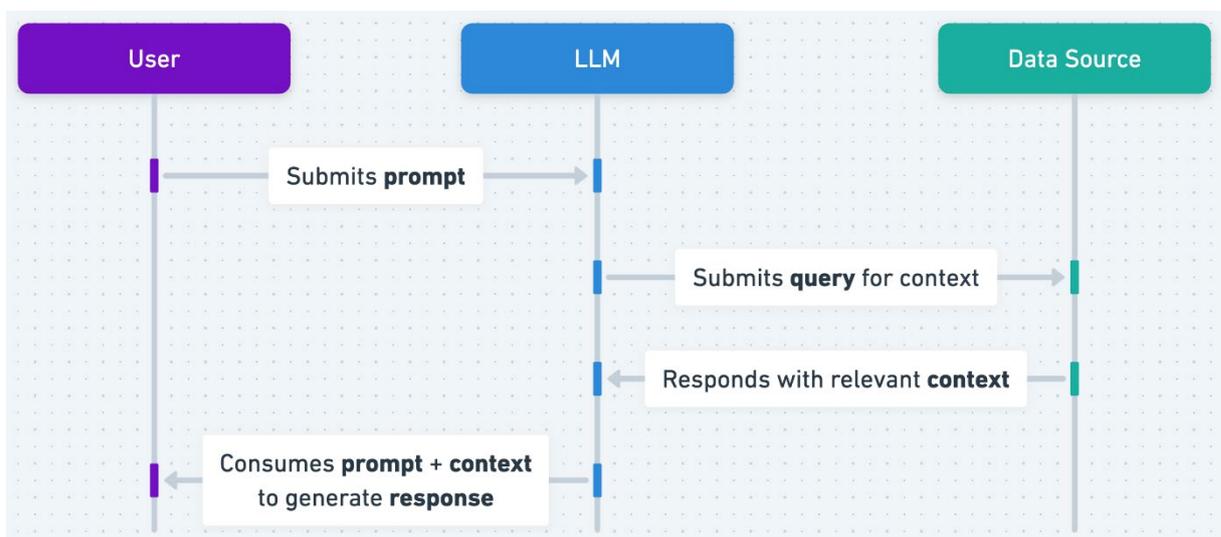
Essas modalidades não podem ser combinadas, e são aplicadas de acordo com a natureza do objeto e o critério de julgamento.

Fonte: ChatGPT (2024d)

Outra aplicação da RAG envolve a recuperação de textos integrais de documentos jurídicos relevantes, fornecendo contexto adicional ao modelo e aumentando a probabilidade de obtenção de resultados mais precisos. Embora a capacidade de processamento dos LLMs seja limitada pela janela de tokens, avanços recentes, como o Google Gemini 1.5 Pro, que suporta até 2 milhões de tokens (aproximadamente 3.000 páginas de texto), indicam um caminho para superar essas limitações e processar grandes volumes de dados.

A pesquisa jurídica com LLMs, por meio da RAG, também se aplica à extração de informações específicas de peças processuais volumosas. Por exemplo, é possível analisar um processo inteiro, utilizando-o como contexto para o LLM (desde que a capacidade de tokens permita) ou buscar semanticamente as informações mais relevantes.

Figura 7 - Fluxo de processo básico de RAG (Recuperação Aumentada de Geração).



Fonte: OpenAI (2024n)

Convém observar que, conforme pesquisa realizada pelo Thomson Reuters Institute (2024), a aplicação de inteligência artificial generativa em pesquisas jurídicas destaca-se como o principal uso entre profissionais de escritórios de advocacia que já utilizam ou têm a intenção de empregá-la. De forma semelhante, o estudo conduzido pela LexisNexis (2023) indicou a pesquisa jurídica como o principal uso da IA generativa entre advogados.

Por fim, embora a integração de LLMs com informações jurídicas atualizadas por meio da RAG mitigue o problema das alucinações, é importante enfatizar que essa integração não as elimina completamente. Além disso, a qualidade da resposta ainda depende da formulação de um prompt adequado pelo usuário. Dessa forma, é fundamental que o usuário valide as informações obtidas em fontes confiáveis, garantindo a precisão e a segurança da pesquisa.

5.2.3 CONSULTORIA JURÍDICA

A Inteligência Artificial Generativa (IA Generativa), com sua capacidade de processar linguagem natural e acessar bases de dados externas por meio de técnicas como Recuperação Assistida por Geração (RAG), abre novos horizontes para o acesso à informação jurídica. Essa tecnologia permite a criação de assistentes de IA especializados em diversas áreas do Direito, capazes de fornecer consultoria jurídica relevante e acessível, democratizando o conhecimento e auxiliando tanto profissionais do Direito quanto cidadãos em geral.

Nesse contexto, uma das aplicações mais promissoras da IA Generativa está nos Assistentes de IA. Essas soluções, também conhecidas como *chatbots*, são configuradas por meio de instruções em linguagem natural, utilizando técnicas de engenharia de prompts. Sua capacidade de responder a dúvidas comuns, fornecer orientações sobre direitos e deveres, e consultar legislação e jurisprudência os torna instrumentos importantes para a difusão do conhecimento legal. Essa acessibilidade beneficia tanto advogados, permitindo-lhes automatizar ou semiautomatizar tarefas rotineiras e focar em questões mais complexas, quanto o público em geral.

A OpenAI (2024e), por exemplo, possibilita a criação de versões customizadas do ChatGPT, denominadas GPTs. Esses GPTs consistem em assistentes que os usuários podem adaptar para tarefas ou tópicos específicos, combinando instruções, conhecimento adicional (documentos) e capacidades como consulta a sites, execução de código Python e integração com APIs de serviços externos. No capítulo sete, serão apresentados alguns exemplos bem-sucedidos de Assistentes de IA jurídicos desenvolvidos no ChatGPT.

Destaca-se que a consultoria jurídica oferecida por essas ferramentas, de caráter educacional, é especialmente útil em situações que não exigem necessariamente a atuação de um profissional, como no Juizado Especial Cível e em reclamações trabalhistas, auxiliando o cidadão a buscar seus direitos de forma mais informada.

5.2.4 REDAÇÃO, REVISÃO E ANÁLISE DE DOCUMENTOS JURÍDICOS

Os Grandes Modelos de Linguagem (LLMs), com sua capacidade de gerar texto e realizar análise semântica, demonstram grande potencial para a redação, revisão e análise de documentos jurídicos. Esse potencial reside na habilidade de processar e compreender linguagem natural, permitindo que a IA gere rascunhos de contratos, petições e pareceres com base em modelos predefinidos e nas informações fornecidas pelo usuário. Além disso, a IA

pode ser utilizada para aprimorar a clareza e a solidez dos argumentos em documentos jurídicos, revisando-os e sugerindo modificações para garantir precisão e conformidade com os padrões legais.

A **precisão** na geração e análise de documentos jurídicos por meio de LLMs pode **ser ainda mais otimizada** com a inclusão de **contexto adicional**. Seja por meio de técnicas como **RAG**, para buscar documentos e informações relevantes, ou pela incorporação do texto de um documento completo diretamente no prompt, a IA é capaz de compreender o contexto específico do caso, **inserindo**, por exemplo, jurisprudências e citações à legislação pertinente.

Outra possibilidade é a personalização do conteúdo de acordo com o estilo de escrita do advogado. Com a técnica **few-shot learning**, o advogado pode fornecer parágrafos exemplares que reflitam seu estilo, possibilitando que o LLM gere documentos alinhados à sua expertise e linguagem jurídica.

No contexto da elaboração de contratos, a IA pode auxiliar tanto profissionais do direito quanto o público em geral. A criação de contratos por meio de LLMs, com base em modelos predefinidos e personalizados, garante maior agilidade e precisão na redação, **inserindo automaticamente cláusulas específicas e revisando o documento para assegurar sua conformidade com a legislação vigente**.

Adicionalmente, ferramentas de **IA Generativa** podem ser empregadas na análise da conformidade de contratos, auxiliando na identificação de cláusulas abusivas, riscos potenciais e pontos de atenção, tornando a revisão contratual mais eficiente e precisa. Para o público leigo, a IA facilita a compreensão de documentos jurídicos complexos, traduzindo a linguagem técnica em termos mais acessíveis, **permitindo** a compreensão dos termos e condições de um contrato antes de sua assinatura.

A relevância da IA Generativa na elaboração de documentos jurídicos é corroborada por pesquisas recentes, como a da **LexisNexis (2023)**, que **indicou** a elaboração de documentos como o segundo principal uso da IA Generativa por advogados, e a do **Thomson Reuters Institute (2024)**, que **apontou** a elaboração de contratos e memoriais como um dos cinco principais casos de uso entre profissionais do direito. **Lai et al. (2023)** também reforçam essa tendência, destacando o potencial da IA na geração automatizada de versões iniciais de documentos jurídicos.

Em síntese, o uso de LLMs na área jurídica apresenta grande potencial de transformação, facilitando a elaboração, revisão e análise de documentos, democratizando o acesso à informação jurídica e auxiliando tanto profissionais do direito quanto o público em geral na compreensão e aplicação da lei.

5.2.5 TRADUÇÃO DE TEXTOS JURÍDICOS

A capacidade de tradução precisa e contextualizada de textos é uma habilidade inerente aos Grandes Modelos de Linguagem (LLMs). Tal competência decorre da abrangência e diversidade dos dados utilizados em seu treinamento, os quais incluem vastos corpora multilíngues e multidisciplinares. Desse modo, os LLMs apresentam excelentes resultados na tradução jurídica devido à sua habilidade de compreender e preservar nuances contextuais específicas do domínio legal, manter a consistência terminológica entre diferentes idiomas e adaptar o estilo e o tom do texto à cultura jurídica do idioma de destino (Lai et al., 2023; Gozalo-Brizuela; Garrido-Merchán, 2023).

No campo jurídico, essa capacidade de tradução possui aplicações significativas em diversas áreas. Conforme Lai et al. (2023), o direito abrange diversas línguas e culturas, demandando a análise de grandes volumes de dados jurídicos em múltiplos idiomas. Esse processo exige superar desafios de tradução e variações terminológicas. Regulamentos da União Europeia, por exemplo, possuem versões em diferentes idiomas, o que obriga os pesquisadores a compará-las para garantir uma interpretação precisa.

Em litígios internacionais, a tradução precisa de documentos legais, como contratos, depoimentos e sentenças, mostra-se essencial para garantir a compreensão equânime por todas as partes envolvidas. Essa precisão revela-se igualmente fundamental no direito comparado, no qual pesquisadores podem utilizar LLMs para traduzir e analisar legislações e doutrinas de diferentes países, facilitando tanto os estudos comparativos quanto a identificação de tendências globais em áreas específicas do direito. A aplicabilidade dos LLMs estende-se, ainda, à arbitragem internacional, em que a tradução automática de alta qualidade não só acelera a análise de documentos, como também aprimora a comunicação entre as partes.

Ademais, no contexto de conformidade global, empresas multinacionais podem empregar IA Generativa para traduzir e adaptar políticas de conformidade, assegurando, dessa forma, o alinhamento com as regulamentações locais em diferentes países.

5.2.6 TRANSCRIÇÃO E ANÁLISE DE ÁUDIOS E VÍDEOS DE JULGAMENTOS OU EVIDÊNCIAS

O desenvolvimento de IAs generativas multimodais torna a análise integrada de textos, áudios e vídeos cada vez mais eficiente. Um exemplo desse avanço é o GPT-4o, lançado em

maio de 2024, um modelo multimodal capaz de realizar raciocínios em tempo real a partir de dados de áudio, imagem e texto (OpenAI, 2024c; Gozalo-Brizuela; Garrido-Merchán, 2023).

Essas IAs, que processam diferentes tipos de dados (áudio, vídeo e texto), permitem a transcrição automática de gravações, a geração de atas ou resumos personalizados e a identificação de trechos relevantes para análise pericial. Tais funcionalidades beneficiam advogados e agentes públicos em investigações, otimizando o tempo e os recursos na análise de materiais audiovisuais.

Em julgamentos, por exemplo, a IA pode transcrever automaticamente as falas, gerar resumos dos pontos discutidos e elaborar atas, agilizando o processo de análise e facilitando a identificação de informações relevantes.

5.2.7 TRANSFORMAÇÃO DE DOCUMENTOS JURÍDICOS PARA O FORMATO VISUAL LAW E LEGAL DESIGN

As IAs generativas baseadas em LLMs demonstram a capacidade de modificar a linguagem e o tom de textos jurídicos, adaptando-os ao público-alvo sem comprometer a essência da informação. Além disso, possibilitam a personalização da saída em diversos formatos de maneira relativamente simples, utilizando, por exemplo, templates em Markdown ou integrando-se com ferramentas de programação como Python para a criação de formatos interativos em HTML ou PDF (OpenAI, 2024f).

Nesse contexto, torna-se viável a criação de documentos jurídicos mais acessíveis por meio da aplicação dos princípios do Visual Law e do Legal Design, que buscam simplificar a linguagem jurídica e incorporar elementos visuais. Dessa forma, a IA pode gerar versões personalizadas de documentos, facilitando a compreensão por um público sem conhecimento jurídico especializado, como clientes, jurados e cidadãos em geral.

Um bom exemplo de aplicação dessa capacidade seria a transformação de documentos jurídicos relacionados à administração pública em formatos mais acessíveis, seguidos de sua publicização. Por exemplo, a conversão de resultados de certames licitatórios, decisões de recursos e outras manifestações administrativas em formatos de Visual Law poderia promover a transparência, em consonância com os preceitos constitucionais.

Por fim, destaca-se o assistente de IA "Oráculo Visual Law", desenvolvido no ChatGPT e acessível gratuitamente, que converte documentos jurídicos, como acórdãos e sentenças, em versões visuais e de fácil entendimento, seguindo um modelo pré-estruturado. Esse assistente será apresentado no capítulo oito desta monografia.

5.2.8 OTIMIZAÇÃO DE RELACIONAMENTO COM O CLIENTE

Assistentes de IA, como o API Assistants da OpenAI (2024f), podem ser integrados aos sistemas de escritórios de advocacia para a construção de *chatbots* com automações, otimizando o relacionamento com o cliente. Além de responder a perguntas, esses assistentes podem executar diversas ações, como realizar a interação inicial com potenciais clientes para compreender suas demandas jurídicas, responder a dúvidas frequentes, redigir e-mails e mensagens com atualizações sobre o andamento de processos (Lai et al., 2023; Gozalo-Brizuela; Garrido-Merchán, 2023).

Dessa forma, mantêm-se os clientes informados e reduz-se a carga administrativa dos escritórios. Com o apoio da IA, escritórios de advocacia podem concentrar-se em tarefas estratégicas, delegando atividades operacionais à tecnologia. Em consequência, observa-se um potencial ganho de eficiência, redução de custos operacionais e aumento da satisfação e confiança no serviço prestado.

5.2.9 ASSISTENTE DE INSTRUÇÃO PROCESSUAL

Uma solução potencial que combina diferentes casos de uso de IA generativa no Direito seria um "Assistente de Instrução Processual". Essa aplicação consistiria em uma ferramenta integrada para a prática jurídica, combinando pesquisa, integração de documentos e dados externos, acompanhamento de prazos, análise e elaboração de peças processuais.

Nesse contexto, propõe-se que o profissional jurídico interaja com um agente de IA principal, o qual, em segundo plano, seria responsável por orquestrar e delegar atividades a um conjunto de agentes especializados. Cada agente seria responsável por uma tarefa específica, como redação jurídica, pesquisa legal, análise de argumentos e desenvolvimento de estratégias, otimizando, assim, o fluxo de trabalho e a eficiência do processo legal.

Essa abordagem de múltiplos agentes especializados vem sendo implementada pela Instrução Assistida com Inteligência Artificial (INACIA), desenvolvida pelo Tribunal de Contas da União (TCU), que será abordada no capítulo sete. Tal sistema, ainda em fase de desenvolvimento, visa a auxiliar o auditor em todo o processo de instrução, desde a extração de informações e avaliação de admissibilidade até a elaboração da instrução (Pereira et al., 2024).

5.3 INICIATIVAS DE IA GENERATIVA NO DIREITO AMERICANO

Nesta seção, analisam-se duas iniciativas proeminentes que exemplificam o potencial disruptivo da Inteligência Artificial Generativa no âmbito jurídico, especialmente nos Estados Unidos, onde estão sediadas empresas líderes como OpenAI, Google e Anthropic: Paxton AI e Harvey.

A **Paxton AI** (2024b) obteve US\$ 6 milhões em financiamento para desenvolver uma tecnologia que otimiza a pesquisa jurídica e a redação, garantindo precisão em citações, verificação das saídas e análise de documentos privados dos usuários. Seu sistema, atualizado em tempo real, abrange leis e regulamentos federais e estaduais dos Estados Unidos, liberando, assim, os profissionais para se concentrarem em tarefas estratégicas.

Por outro lado, a **Harvey** (2024b) assegurou US\$ 100 milhões em sua Série C, em junho de 2024, liderada pela GV (Google Ventures), com participação da OpenAI, alcançando uma avaliação de US\$ 1,5 bilhão. Esse aporte visa expandir suas operações e aprimorar capacidades tecnológicas para enfrentar desafios globais nos serviços profissionais.

O investimento expressivo em ambas as empresas evidencia a confiança do mercado no potencial da IA Generativa para transformar o setor jurídico, prometendo maior eficiência, precisão e acesso à informação. Contudo, o impacto dessas iniciativas no panorama jurídico estadunidense, incluindo benefícios, desafios e implicações éticas, demanda uma análise aprofundada.

5.3.1 HARVEY AI

A *lawtech* Harvey oferece uma suíte de produtos voltados para serviços legais, incluindo a capacidade de elaborar, analisar e responder a perguntas sobre documentos jurídicos por meio de instruções em linguagem natural. A plataforma permite a realização de pesquisas complexas, automação de fluxos de trabalho especializados e análise de contratos, além de gerar perfis corporativos detalhados e listas de questões personalizadas. Destina-se a aumentar a produtividade de firmas jurídicas, proporcionando controle e visibilidade sobre a produção de trabalho (Harvey, 2024c).

A empresa destaca-se como uma das principais plataformas de IA Generativa no campo jurídico, notabilizando-se por suas parcerias estratégicas com a OpenAI e a Voyage AI. De acordo com a OpenAI (2024g), "Harvey é uma plataforma de IA generativa voltada para profissionais do direito, tributação e finanças". Em 2023, a empresa experimentou um

crescimento notável, obtendo US\$ 80 milhões em financiamento da Série B e alcançando uma valorização de US\$ 715 milhões (Harvey, 2024b).

A colaboração entre Harvey e OpenAI resultou no desenvolvimento de um modelo de IA Generativa customizado, especificamente projetado para pesquisa jurídica. Esse modelo incorporou um expressivo volume de 10 bilhões de tokens de dados legais, representando um avanço significativo na aplicação de LLMs no contexto jurídico (OpenAI, 2024g). A eficácia desse modelo customizado foi comprovada por meio de testes extensivos conduzidos em dez dos maiores escritórios de advocacia dos Estados Unidos. Os resultados foram notáveis, com 97% dos advogados participantes expressando preferência pelas respostas geradas pelo modelo customizado em comparação com o GPT-4 (OpenAI, 2024g).

Segundo Harvey (2024c), uma das principais vantagens do seu modelo reside na capacidade de reduzir significativamente as "alucinações" – informações geradas pela IA que são imprecisas ou sem fundamento. O modelo demonstrou ser capaz de fornecer "respostas juridicamente precisas e sustentadas por citações diretas de casos" (OpenAI, 2024g), aumentando, assim, a confiabilidade e aplicabilidade das informações geradas em contextos legais reais.

Com o objetivo de aprimorar ainda mais a precisão e a eficiência de suas buscas jurídicas, Harvey estabeleceu uma parceria com a Voyage AI, uma empresa especializada no desenvolvimento de modelos de *embeddings* personalizados para domínios específicos. Conforme explicado anteriormente, *embeddings* são representações matemáticas de palavras e frases que capturam seu significado semântico, permitindo a realização de buscas semânticas em grandes volumes de dados. A colaboração entre Harvey e Voyage AI resultou na criação do modelo "voyage-law-2-harvey", treinado com mais de 20 bilhões de tokens de textos jurídicos dos Estados Unidos (Harvey, 2024a).

Esse modelo demonstrou desempenho superior aos modelos de **embeddings** padrão, como os da Google e OpenAI, reduzindo em 25% a irrelevância dos resultados e otimizando tanto o armazenamento quanto a latência, o que aprimora significativamente a eficiência e a precisão nas buscas jurídicas executadas por LLMs por meio da técnica de **RAG** (Harvey, 2024a).

Essas inovações possibilitam que advogados realizem tarefas complexas de maneira assistida pela IA, como redigir documentos, responder a perguntas sobre litígios e analisar contratos e e-mails, com maior celeridade e acurácia, liberando tempo para que se concentrem em atividades estratégicas e na interação com clientes.

5.3.2 PAXTON AI

A Paxton AI oferece uma plataforma de inteligência artificial destinada a profissionais jurídicos e empresariais, com o objetivo de automatizar tarefas como revisão de contratos, redação de documentos legais e análise documental. A empresa utiliza modelos de linguagem específicos para o setor jurídico, visando ao aumento da eficiência e precisão no gerenciamento de processos legais. A plataforma atende escritórios de advocacia, conselheiros jurídicos e equipes empresariais, facilitando a realização de pesquisas jurídicas e auditorias contratuais (Paxton AI, 2024c).

Em junho de 2024, a Paxton AI (2024a) anunciou que atingiu 93,82% de acurácia no Stanford Legal Hallucination Benchmark, o qual avalia a capacidade das ferramentas de IA jurídica em produzir interpretações legais corretas, sem erros ou "alucinações". Esse resultado foi obtido a partir de uma amostra de 1.600 tarefas selecionadas de um universo de 750.000 disponíveis no benchmark, demonstrando a capacidade do sistema em lidar com tarefas jurídicas complexas (Paxton AI, 2024a).

Adicionalmente, a Paxton AI (2024a) introduziu o Confidence Indicator, uma ferramenta que permite aos usuários avaliar a confiabilidade das respostas fornecidas pela IA. Conforme divulgado pela empresa, essa funcionalidade categoriza as respostas em níveis de confiança — baixo, médio ou alto — e oferece sugestões para pesquisas adicionais. Dessa forma, o Confidence Indicator avalia as respostas com base em critérios como relevância contextual, evidências fornecidas e complexidade da consulta, assegurando que o nível de confiança atribuído reflita de maneira mais precisa a confiabilidade da resposta do que a confiança interna dos modelos de linguagem. Essa ferramenta auxilia os profissionais jurídicos na avaliação rápida da confiabilidade das respostas, acelerando a tomada de decisões e mitigando riscos associados a informações legais incorretas (Paxton AI, 2024a).

Os resultados obtidos pela Paxton AI no Stanford Legal Hallucination Benchmark e a introdução do Confidence Indicator demonstram o esforço contínuo da empresa em aprimorar suas ferramentas para o setor jurídico. Dado o contexto de alta competitividade no mercado norte-americano de lawtechs e o rápido avanço da IA generativa no setor jurídico, esses desenvolvimentos mostram-se necessários para que a Paxton AI se posicione como uma solução relevante, capaz de competir com outras tecnologias emergentes e de atender às crescentes demandas do campo jurídico.

5.4 INICIATIVAS DE IA GENERATIVA NO DIREITO BRASILEIRO

É notório que o ecossistema de inovação tecnológica no Brasil e o capital disponível não são comparáveis aos dos Estados Unidos. Essa disparidade, somada aos elevados custos para treinar Grandes Modelos de Linguagem (LLMs), dificulta o desenvolvimento de soluções robustas por empresas brasileiras. Estima-se, por exemplo, que o GPT-4, lançado em 2023 pela OpenAI, tenha custado mais de US\$ 100 milhões apenas no treinamento (Meyer, 2024).

Apesar dessas adversidades, entre 2023 e 2024 surgiram diversas iniciativas utilizando IA generativa no Direito brasileiro. Destacam-se a construção de softwares customizados a partir de *foundational* models, ou "modelos de base". Em síntese, um modelo de base é um modelo de IA treinado em dados amplos e em larga escala, podendo ser adaptado a uma ampla gama de tarefas subsequentes. Esses modelos proprietários são comercializados, permitindo que desenvolvedores integrem recursos avançados de IA em seus aplicativos sem a necessidade de construir ou manter a infraestrutura subjacente. Alternativamente, também se observa o uso de modelos open-source, como o LLAMA, da Meta (2024), para oferecer soluções que vêm transformando o contexto jurídico brasileiro.

Em termos de pesquisa e desenvolvimento de soluções inovadoras no Direito, o setor público tem liderado essas iniciativas, com o Tribunal de Contas da União (TCU) assumindo a vanguarda com o lançamento do ChatTCU em 2023, uma adaptação do ChatGPT para uso interno. Construído a partir do modelo base GPT da OpenAI e integrado a dados internos do TCU, a primeira versão do ChatTCU foi disponibilizada para um público restrito de servidores em março de 2023, baseada na versão GPT-3.5, sendo liberada para todos os servidores em julho (Brasil, 2023c).

Nesta seção, serão exploradas três principais iniciativas de IA Generativa no Brasil. Inicialmente, abordam-se as iniciativas lideradas pelo TCU, incluindo o ChatTCU e o INACIA, sistemas que têm transformado o processo de trabalho de auditoria e análise de processos. Em seguida, discute-se a SARA, uma ferramenta desenvolvida pela USP em parceria com a Escola Superior da Magistratura Tocantinense e o Tribunal de Justiça do Estado do Tocantins, criada para otimizar a análise de recursos nos tribunais. Por fim, analisam-se as contribuições do Juru, uma IA especializada em Direito, desenvolvida pela startup brasileira Maritaca AI.

Cabe ressaltar que, embora tenham sido identificadas *lawtechs* brasileiras que comercializam soluções baseadas em IAs Generativas para o setor jurídico, como a Lexter AI, fundada em 2020 e que, em 2024, recebeu um investimento de R\$ 16 milhões (Globo, 2024),

não se observou um investimento significativo em pesquisa e desenvolvimento voltado para o aprimoramento das tecnologias subjacentes, com maior foco no apelo comercial.

Por último, destaca-se que o Plano Brasileiro de Inteligência Artificial (PBIa) prevê um investimento de R\$ 23 bilhões ao longo de quatro anos, incluindo a aquisição de um supercomputador de ponta, visando guiar o desenvolvimento ético e sustentável da inteligência artificial no Brasil. De acordo com o Ministério da Ciência, Tecnologia e Inovação, o plano tem como objetivo nortear a aplicação da IA em diversas áreas estratégicas, com a expectativa de gerar impacto significativo em setores como saúde, educação e indústria (Brasil, 2024a).

5.4.1 TRIBUNAL DE CONTAS DA UNIÃO: CHATTCU E INACIA

O Tribunal de Contas da União (TCU) destaca-se como um dos pioneiros na aplicação de IA Generativa no Direito brasileiro, com o desenvolvimento e implementação de duas ferramentas principais: o ChatTCU e o INACIA. O ChatTCU, lançado em 2023, consiste em um assistente virtual baseado no modelo ChatGPT, projetado para auxiliar na produção e revisão de textos, além de realizar adaptações de linguagem e traduções. Segundo o órgão, ao operar na nuvem privada da Microsoft, o ChatTCU garante a confidencialidade e a segurança dos dados sigilosos processados pelo TCU (Brasil, 2023c).

Em 2024, o TCU expandiu o alcance dessa tecnologia ao disponibilizar o código-fonte do ChatTCU para outros órgãos públicos, como o Ministério da Gestão e da Inovação em Serviços Públicos (MGI) e Tribunais de Contas Estaduais. Segundo a Corte de Contas, essa iniciativa visa a promover a inovação e a customização de soluções tecnológicas para o aprimoramento da gestão pública, contribuindo para a transformação digital no setor público (Brasil, 2024c).

Além do ChatTCU, o Tribunal de Contas da União iniciou, em 2022, o desenvolvimento do **INACIA** (Instrução Assistida com Inteligência Artificial), no contexto de uma "encomenda tecnológica". O Tribunal lançou uma chamada pública para a contratação de uma solução de IA ainda em fase de pesquisa e desenvolvimento. O INACIA foi projetado para automatizar etapas críticas na análise de processos, como a extração de informações e a análise de admissibilidade, com o objetivo de aumentar a eficiência dos auditores, permitindo que se concentrem em tarefas mais complexas (Pereira et al., 2024).

O ChatTCU e o INACIA serão abordados com maior profundidade no capítulo sete desta monografia, onde será realizado um estudo de caso sobre a implementação de soluções de IA generativa no Tribunal de Contas da União.

5.4.2 USP e ESCOLA SUPERIOR DA MAGISTRATURA TOCANTINENSE: SARA

A **SARA** (Síntese Adaptativa de Requisitos de Admissibilidade) é uma IA generativa desenvolvida para auxiliar os tribunais brasileiros, especialmente no processamento e análise de recursos constitucionais, especiais e extraordinários. Criada por meio de uma parceria entre a Universidade de São Paulo (USP), a Escola Superior da Magistratura Tocantinense (Esmat) e o Tribunal de Justiça do Estado do Tocantins (TJTO), a solução foi lançada em 2024 com o objetivo de otimizar o fluxo de trabalho e garantir a conformidade das minutas geradas com os padrões dos tribunais (Castro Sousa, 2024).

Conforme noticiado no site da USP, essa ferramenta representa um avanço significativo na automação de tarefas repetitivas e demoradas no sistema judiciário, ao mesmo tempo em que permite a interação humana para assegurar a precisão e a eficácia na tramitação dos recursos. Ainda segundo a USP, embora já esteja pronta para uso, a SARA continua em fase de testes em um ambiente controlado para garantir sua aplicação segura e eficaz nos tribunais brasileiros (Castro Sousa, 2024).

5.4.3 MARITACA AI: JURU

A Maritaca AI (2024), uma *startup* brasileira especializada em inteligência artificial, fundada em 2022, desenvolveu o **Juru**, uma versão customizada do modelo **Sabiá-2 Small**, construído a partir do **fine-tuning** do modelo LLAMA 70B. Adaptado para o domínio jurídico brasileiro, o Juru foi treinado exclusivamente em dados jurídicos e projetado para lidar com as complexidades do português jurídico e da legislação nacional, resultando em um desempenho significativamente melhorado em tarefas específicas, como a resolução de questões do Exame Nacional de Desempenho dos Estudantes (Enade) na área de Direito (Malaquias Júnior et al., 2024).

Entretanto, segundo Malaquias Júnior et al. (2024), a especialização do Juru trouxe desafios, como a redução da versatilidade do modelo em outras áreas do conhecimento, evidenciando as limitações de LLMs altamente especializados em contextos multidisciplinares.

6. DESAFIOS E LIMITAÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NO DIREITO

A introdução da Inteligência Artificial Generativa (IAG) no campo jurídico configura-se como uma inovação promissora, com grande potencial, mas acompanhada de complexidades e obstáculos significativos. Nesse sentido, apesar do caráter transformador dessa tecnologia, sua implementação enfrenta uma série de desafios e limitações que exigem uma análise criteriosa.

Neste capítulo, será dada uma investigação aprofundada dessas barreiras, categorizando-as em três dimensões principais: técnicas, legais e éticas. Será dada especial ênfase às limitações do ponto de vista dos operadores do Direito por dois motivos fundamentais. Primeiramente, reconhece-se o acesso disponível de forma ampla e gratuita a ferramentas de IAG, como o ChatGPT da OpenAI, o Gemini da Google e o Claude da Anthropic, por meio de serviços e aplicativos na web, o que as torna facilmente acessíveis aos profissionais jurídicos. Em segundo lugar, destaca-se a necessidade urgente de conscientizar os operadores do Direito sobre as potencialidades e, sobretudo, sobre as limitações dessas ferramentas, visando garantir uma aplicação ética e eficaz no âmbito jurídico. Essa abordagem busca assegurar que os profissionais atuantes na prática legal estejam devidamente informados e preparados para integrar essas tecnologias em sua rotina de forma pragmática, responsável e criteriosa.

Convém ressaltar que, embora o foco deste estudo recaia sobre as limitações práticas de uso, não se pode ignorar os desafios fundamentais relacionados ao desenvolvimento de modelos de IAG de grande porte. Reitera-se o já mencionado custo proibitivo de treinamento, que frequentemente ultrapassa a casa das centenas de milhões de dólares, representando uma barreira significativa à inovação, especialmente para empresas brasileiras.

6.1 LIMITAÇÕES TÉCNICAS

As limitações técnicas da Inteligência Artificial Generativa (IAG) no contexto jurídico representam obstáculos significativos à sua plena adoção e eficácia. Esses desafios, intrinsecamente ligados à natureza e ao funcionamento dos Grandes Modelos de Linguagem (LLMs), impactam diretamente a confiabilidade e a aplicabilidade dessas tecnologias no campo do Direito.

O desconhecimento das limitações dos LLMs por profissionais jurídicos apresenta riscos consideráveis. A falta de familiaridade com essas ferramentas pode resultar na perda das vantagens de produtividade que elas oferecem, impactando negativamente a carreira desses profissionais. Ademais, o uso inadequado, causado pela falta de conhecimento acerca de suas limitações, pode acarretar erros jurídicos com consequências graves para o profissional e seus

clientes. Um exemplo disso ocorreu em 2023, quando um juiz federal da 1ª Região passou a ser investigado pelo Conselho Nacional de Justiça (CNJ) por utilizar uma tese falsa gerada pelo ChatGPT em uma decisão judicial. A ferramenta criou uma jurisprudência inexistente do Superior Tribunal de Justiça, o que foi descoberto por um advogado que contestou a decisão (Conjur, 2023).

Portanto, compreender as limitações das IAG é imprescindível para garantir a utilização responsável e eficaz dessas tecnologias no âmbito jurídico.

6.1.1 GERAÇÃO DE CONTEÚDO FACTUALMENTE INCORRETO

A geração de conteúdo factualmente incorreto, fenômeno conhecido como "alucinações", constitui uma das limitações mais críticas dos sistemas de Inteligência Artificial Generativa (IAG). No contexto dos Grandes Modelos de Linguagem (LLMs), a alucinação é definida como a criação de saídas que, embora possam parecer plausíveis ou coerentes com o input, são, na verdade, incorretas ou não baseadas em fatos verificáveis (Xu; Jain; Kankanhalli, 2024).

As alucinações classificam-se em dois tipos principais, cada um com características e implicações distintas. Primeiramente, a alucinação intrínseca ocorre quando o modelo gera conteúdo internamente inconsistente ou contraditório em relação à informação fornecida diretamente como contexto, seja por meio de um documento adicional ou do próprio prompt. No contexto jurídico, por exemplo, o modelo pode receber uma sentença judicial como contexto e, ao gerar um resumo, afirmar que a decisão foi favorável a uma das partes quando, na verdade, o documento fornecido indica que a decisão foi desfavorável. Esse tipo de erro reflete falhas na coerência interna das saídas do modelo e demonstra a complexidade em garantir a consistência das informações geradas (Xu; Jain; Kankanhalli, 2024).

Em seguida, tem-se a alucinação extrínseca, que acontece quando o modelo gera informações que não podem ser verificadas pelo conteúdo fornecido no prompt ou por documentos adicionais, nem são fundamentadas pelo conjunto de dados usado em seu treinamento. Esse tipo de alucinação é comum, por exemplo, quando o modelo é solicitado a explicar um conceito jurídico e menciona uma doutrina que nunca foi discutida ou reconhecida em nenhuma literatura jurídica conhecida. Nessas situações, o modelo extrapola a informação disponível, introduzindo detalhes que não têm suporte factual, o que pode resultar em saídas que, apesar de parecerem detalhadas e confiáveis, carecem de fundamento (Xu; Jain; Kankanhalli, 2024).

As alucinações nos LLMs são causadas por uma combinação de fatores que se manifestam em diferentes fases do processo, incluindo a coleta de dados, o treinamento e a inferência. Durante a coleta de dados, informações imprecisas, desatualizadas ou enviesadas podem ser incorporadas ao modelo, levando-o a aprender com base em dados falhos. Além disso, dados raros ou pouco frequentes são difíceis para o modelo recordar, aumentando a probabilidade de erros. No treinamento, problemas na arquitetura do modelo, como a distribuição de atenção através do mecanismo de self-attention (da arquitetura transformer), podem dificultar a identificação e o processamento de informações cruciais, especialmente quando se lida com grandes volumes de dados. Já na fase de inferência, a aleatoriedade na escolha de respostas e a falta de atenção ao contexto completo do input podem levar o modelo a gerar respostas inadequadas ou incorretas. Esses fatores, em conjunto, contribuem para que os LLMs, por vezes, produzam informações que parecem corretas, mas que, na realidade, são factualmente incorretas (Xu; Jain; Kankanhalli, 2024).

Para mitigar esses problemas, diversas abordagens têm sido propostas. Uma primeira estratégia crucial é, no estágio de pré-treinamento desses modelos, o uso de conjuntos de dados focados em fatos e de alta qualidade, que passam por um rigoroso processo de limpeza para garantir que o treinamento dos modelos seja baseado em informações precisas e isentas de vieses.

Entre as técnicas promissoras, destaca-se a Recuperação Baseada em Conhecimento, que visa integrar métodos de recuperação de informações, como o acesso e uso de bases de dados externas, para fornecer aos LLMs informações adicionais, atualizadas e verificáveis durante a inferência. Por exemplo, a *Retrieval-Augmented Generation* (RAG) combina a busca de informações relevantes com a capacidade de geração do modelo, utilizando busca semântica para encontrar documentos ou fragmentos de texto relevantes e, em seguida, empregando esse conteúdo como contexto para gerar respostas mais precisas e fundamentadas (Xu; Jain; Kankanhalli, 2024).

Outra estratégia é o *fine-tuning* com conjuntos de dados específicos de domínio, que auxilia o modelo a compreender melhor e responder corretamente a consultas dentro de áreas específicas. No campo jurídico, por exemplo, o modelo pode ser ajustado com dados específicos sobre legislações e jurisprudências, permitindo que ele ofereça respostas mais precisas e contextualizadas sobre questões legais (Xu; Jain; Kankanhalli, 2024).

Ademais, conforme Lee (2023), o ajuste de parâmetros do modelo, como a temperatura, pode influenciar significativamente as alucinações. Uma temperatura alta encoraja o modelo a explorar uma gama maior de possíveis sequências de tokens, aumentando tanto a criatividade

quanto a propensão a alucinações. Em contrapartida, uma temperatura baixa torna o modelo mais conservador, gerando saídas mais previsíveis e menos criativas, com menor risco de alucinações. Geralmente, esse ajuste entre 0 e 2 para a temperatura está disponível apenas para uso via API, como no caso dos modelos Gemini e ChatGPT, não sendo acessível diretamente nos chatbots. Desse modo, encontrar o equilíbrio ideal entre criatividade e precisão é essencial para otimizar o desempenho do modelo em diferentes tarefas.

Complementarmente, as barreiras de segurança, ou *guardrails*, são controles que orientam o comportamento dos LLMs para prevenir erros, como alucinações. Elas podem ser implementadas como instruções no prompt, que verificam as respostas do modelo contra dados de referência confiáveis antes de serem entregues ao usuário, bloqueando ou ajustando aquelas que não correspondem à realidade, garantindo assim maior precisão e segurança nas interações (Jarvis, 2023).

Devem ser ainda empregadas as melhores práticas e técnicas de engenharia de *prompt*, como o *Chain-of-Thought*, apresentado no capítulo três desta monografia. Ele auxilia na orientação do modelo através de raciocínios passo a passo, decompondo tarefas complexas, o que aumenta a precisão das respostas geradas, mitigando as alucinações. É importante ressaltar que o fornecimento de contexto relevante e específico também é essencial para que o modelo gere respostas relevantes e contextualmente precisas.

Outra técnica de engenharia de *prompt*, proposta por Dhuliawala et al. (2023), é o *Chain-of-Verification* (CoVe). Segundo estudo publicado pelos autores, CoVe é uma abordagem para reduzir alucinações em grandes modelos de linguagem, utilizando um processo em quatro etapas: primeiro, o modelo gera uma resposta inicial; em seguida, cria perguntas de verificação para checar a precisão dessa resposta; depois, o modelo responde a essas perguntas de forma independente, sem ser influenciado pela resposta original; e, finalmente, o modelo elabora uma nova resposta corrigida, incorporando as verificações feitas. Esse método envolve múltiplas interações dentro do próprio modelo para garantir maior precisão na resposta final.

Não obstante tais técnicas possam ser aplicadas, de acordo com Xu, Jain e Kankanhalli (2024), a alucinação é uma restrição inerente ao funcionamento de LLMs baseados em Transformers e é inevitável. Por isso, alguns autores inclusive não a chamam de “problema”, mas sim de uma “feature”, ou “funcionalidade”, desses modelos. Ter conhecimento dessa limitação é essencial para um uso responsável de IAG. A inevitabilidade das alucinações em Grandes Modelos de Linguagem nas arquiteturas atuais decorre da impossibilidade desses modelos, apesar de avançados, aprenderem e reproduzirem todas as possíveis formas de conhecimento humano. Em outras palavras, mesmo com constantes melhorias, esses modelos

sempre apresentarão limitações e, em alguns casos, gerarão respostas aparentemente corretas, mas factualmente equivocadas. Essa característica fundamental reflete a incapacidade dos LLMs atuais de compreender completamente ou prever todas as nuances e contextos da informação.

Por conseguinte, empresas como a OpenAI estão investindo no desenvolvimento de novas tecnologias que possam superar essas limitações desses modelos, como a especulação sobre o projeto “*Strawberry*”, com o objetivo de aprimorar as capacidades de raciocínio de seus modelos. Este projeto visa não apenas permitir que a IA responda a perguntas, mas também que planeje e execute tarefas complexas de forma autônoma, como navegar na internet para realizar pesquisas profundas. A iniciativa faz parte de um esforço maior para superar as limitações atuais da IA Generativa baseada em Transformers em áreas como o senso comum e a resolução de problemas que exigem múltiplas etapas (Tong; Paul, 2024).

6.1.2 EXPLICABILIDADE

Conforme mencionado na introdução, as IAGs Generativas baseadas em LLMs utilizam uma arquitetura de rede neural profunda (*Deep Learning*). Essas redes funcionam como uma “caixa-preta” (*black box*), o que significa que, embora as entradas e saídas do sistema sejam observáveis, o funcionamento interno dessas redes permanece opaco e de difícil compreensão. Nesse sentido, a interpretação das redes neurais representa um desafio significativo, conforme destacado pela OpenAI (2024o):

Diferentemente de criações humanas tradicionais, como carros, que podem ser projetados e ajustados diretamente com base em suas especificações, as redes neurais são treinadas por algoritmos complexos cujos resultados não podem ser facilmente decompostos em partes identificáveis. Essa característica torna a análise de segurança em IA mais desafiadora, pois não se pode aplicar os mesmos métodos de raciocínio usados em sistemas físicos, onde o entendimento detalhado dos componentes garante a confiabilidade e segurança do produto final (OpenAI, 2024o, tradução nossa).

Embora frequentemente utilizados como sinônimos, os termos “interpretabilidade” e “explicabilidade” possuem significados distintos no contexto de modelos de aprendizado de máquina. A **interpretabilidade** refere-se à capacidade de um ser humano inspecionar diretamente um modelo e entender como ele chegou a uma determinada conclusão. Modelos como árvores de decisão e regressão linear são considerados altamente interpretáveis, pois as relações entre entrada e saída são claras e intuitivas, permitindo prever como mudanças em uma variável de entrada afetarão o resultado. Em outras palavras, a interpretabilidade depende da

transparência e, de modo geral, da simplicidade do próprio modelo (AWS, 2024; Russell; Norvig, 2021).

Por outro lado, a **explicabilidade** envolve a capacidade de fornecer uma explicação compreensível sobre o comportamento de modelos mais complexos, como redes neurais, que são frequentemente considerados "caixas-pretas". Enquanto a interpretabilidade depende da estrutura do modelo em si, a explicabilidade pode ser alcançada por meio de ferramentas externas que geram explicações sobre o processo decisório do modelo, sem que este seja diretamente compreensível. Um exemplo disso é o LIME (*Local Interpretable Model-Agnostic Explanations*), que cria um modelo aproximado e interpretável para auxiliar na compreensão das decisões do modelo original (AWS, 2024; Russell; Norvig, 2021).

Em síntese, enquanto a interpretabilidade está ligada à transparência estrutural do modelo, a explicabilidade oferece uma camada adicional de compreensão para modelos complexos, onde a transparência não é plenamente alcançável.

A OpenAI (2024n), por exemplo, conforme artigo e notícia publicados em junho de 2024, está investigando formas de melhorar a compreensão de como redes neurais tomam decisões, com foco em tornar seus modelos mais explicáveis. Utilizando técnicas como autoencoders esparsos, que são modelos de IA projetados para identificar e representar apenas os padrões mais relevantes de dados complexos, a empresa busca identificar os padrões-chave que influenciam as saídas dos modelos, como o GPT-4. Embora essas técnicas contribuam para tornar os modelos mais compreensíveis, o foco principal da pesquisa é na explicabilidade – a capacidade de fornecer explicações claras sobre o comportamento dos modelos. Segundo a empresa, estão sendo desenvolvidos métodos para ampliar essa explicabilidade, enfrentando desafios como capturar o comportamento completo dos modelos e escalá-los para entender melhor todos os conceitos que eles processam. Esse trabalho é fundamental para aumentar a confiança e a segurança em sistemas de IA, permitindo uma melhor análise de sua robustez e funcionamento (OpenAI, 2024n).

Diante dos desafios tecnológicos inerentes à interpretabilidade de redes neurais profundas, busca-se a criação de mecanismos de explicabilidade que promovam, em certo grau, a transparência do funcionamento desses modelos. Tais mecanismos visam fornecer explicações claras sobre o comportamento dos modelos, com ênfase no processo decisório, ou seja, na relação entre as entradas e as saídas.

Registra-se que o Projeto de Lei nº 2338, de 2023, que dispõe sobre o uso da Inteligência Artificial e que atualmente está em discussão no Senado Federal, adotou a terminologia

"explicabilidade" como uma exigência para soluções de IA classificadas como de alto risco (Brasil, 2023b).

6.1.3 LIMITAÇÃO DE TOKENS DA JANELA DE CONTEXTO E SAÍDA

Conforme abordado no capítulo 3, a limitação de tokens na janela de contexto e na saída dos LLMs representa uma restrição técnica significativa. A janela de contexto refere-se à quantidade máxima de tokens que um modelo pode considerar simultaneamente ao processar informações, enquanto a limitação de saída diz respeito ao número máximo de tokens que o modelo pode gerar em uma única resposta.

Essa limitação afeta diretamente a análise de documentos extensos e o processamento de conversas longas, especialmente em chatbots que tendem a incluir o histórico da conversa como contexto. Quando a conversa se estende por várias interações, o modelo pode perder informações relevantes à medida que o contexto anterior é truncado. Contudo, os modelos têm evoluído rapidamente nesse aspecto. Um exemplo recente é o Google Gemini, que já consegue processar até 2 milhões de tokens de entrada, ampliando consideravelmente a capacidade de análise.

Estratégias comuns para lidar com essas limitações incluem o uso de técnicas de sumarização e *Retrieval-Augmented Generation* (RAG) com busca semântica. A sumarização envolve a compressão do conteúdo para destacar os pontos mais relevantes, permitindo que informações essenciais sejam processadas dentro dos limites da janela de contexto. Por outro lado, o RAG com busca semântica recupera partes relevantes de documentos ou bases de dados, permitindo que o modelo integre essas informações à geração de texto, mesmo quando o conteúdo completo não pode ser diretamente processado devido à limitação de tokens.

Quanto à limitação de saída, uma abordagem comum implementada pelos *chatbots* das principais empresas é a funcionalidade de "Gerar mais", que permite a continuação da geração de texto até que o limite de tokens seja novamente atingido. Isso oferece uma solução prática para expandir a quantidade de conteúdo gerado, assegurando que informações adicionais possam ser incluídas em respostas longas e detalhadas.

6.1.4 DIFICULDADE COM TAREFAS QUE EXIJAM RACIOCÍNIO COMPLEXO

A capacidade limitada dos Grandes Modelos de Linguagem (LLMs) em realizar raciocínios complexos de forma autônoma configura-se como um desafio crucial para sua aplicação em áreas que demandam alta precisão e pensamento crítico, como o campo jurídico. Apesar dos avanços em diversas tarefas de geração de texto, esses modelos enfrentam dificuldades significativas em processos de raciocínio que envolvem múltiplas etapas, como a elaboração de uma sentença que requer análise de argumentos de diferentes partes, consulta e interpretação sistemática de jurisprudência, legislação e doutrina para, então, decidir sobre o caso. Tal limitação compromete a eficácia dos LLMs na resolução de problemas que exigem uma abordagem lógica e estruturada, tornando-os menos confiáveis em cenários que demandam compreensão profunda e aplicação de regras complexas (Huang et al., 2024).

Ademais, essa limitação é acentuada pela incapacidade dos LLMs de autocorrigir erros de raciocínio sem a intervenção de feedback externo. Conforme estudo realizado por Huang et al. (2024), em muitos casos, as tentativas de autocorreção não apenas falham em melhorar as respostas, mas também podem degradar sua qualidade.

Nesse cenário, conforme apresentado em capítulo específico deste trabalho, uma técnica de Engenharia de *Prompt* que pode ser utilizada para mitigar esse problema é o *Chain-of-Thought Prompting*, que visa a decompor tarefas complexas, guiando o modelo por meio de raciocínios em múltiplas etapas. Entretanto, conforme Huang e Chang (2023), essas soluções ainda não são suficientes para superar completamente as limitações intrínsecas dos LLMs, o que indica a necessidade de desenvolvimento contínuo e de novas abordagens que integrem feedbacks externos de maneira mais eficaz, melhorando a robustez desses modelos em tarefas complexas.

Por fim, destaca-se que, segundo estudo conduzido pelo Laboratório de Ciência da Computação e Inteligência Artificial (CSAIL) do MIT, os LLMs apresentam dificuldade em generalizar para cenários que diferem dos padrões aos quais estão acostumados (Gordon, 2024). No contexto jurídico, tal limitação pode constituir um problema em casos que, devido às suas características, exijam a elaboração de soluções inéditas. Por outro lado, essa limitação pode ser mitigada ao utilizar os LLMs de forma acessória, sem delegar-lhes a responsabilidade total pela resolução de um litígio.

6.1.5 PROMPT INJECTION: JAILBREAK E PROMPT LEAKING

A suscetibilidade dos LLMs a técnicas de manipulação, como *prompt injection* e *jailbreak*, constitui uma preocupação crítica no desenvolvimento de soluções de Inteligências Artificiais (IAs) generativas seguras. O *prompt injection* refere-se à inserção de comandos maliciosos nos *prompts* de entrada, enquanto o *jailbreak*, um subtipo dessa técnica, envolve a exploração de vulnerabilidades para contornar as restrições do modelo. Essas falhas podem levar os modelos a gerar conteúdo inadequado ou a executar ações potencialmente prejudiciais. Além disso, o *prompt leaking* é outro subtipo que visa expor informações sensíveis, como instruções de um Assistente ou documentos internos, comprometendo a integridade e a confidencialidade dos sistemas e processos que utilizam essas tecnologias.

Uma estratégia para prevenir ataques em LLMs é a adoção de "guardrails" ou mecanismos de segurança. Esses *guardrails* desempenham funções essenciais, como monitorar e filtrar os conteúdos gerados pelos LLMs, classificando e corrigindo respostas para garantir que estejam em conformidade com critérios éticos e de segurança. Ademais, esses mecanismos detectam e previnem tentativas de ataques, como os *jailbreaks*, que buscam explorar vulnerabilidades nos modelos. Ao aplicar essas medidas de segurança em diferentes etapas, desde o treinamento até a interação direta com os usuários, os *guardrails* contribuem para a proteção contra conteúdos inadequados e para a manutenção da integridade e segurança dos LLMs (Xhonneux et al., 2024).

A OpenAI (2024q), por exemplo, disponibiliza para desenvolvedores que integram o GPT em seus serviços a API *Moderations* como uma medida de mitigação de *Prompt Injection*. Essa funcionalidade permite identificar e filtrar automaticamente conteúdo potencialmente prejudicial, como discursos de ódio, assédio ou violência, antes que ele seja processado pelos modelos de IA. Com isso, reduz-se o risco de que *prompts* maliciosos manipulem a saída dos modelos, ajudando a manter a integridade e a segurança do sistema.

Já em estudo publicado em abril de 2024, pesquisadores da OpenAI propuseram a "técnica da hierarquia de instruções", que consiste em priorizar diferentes níveis de instruções que um LLM recebe, assegurando que as instruções de maior privilégio, como mensagens de sistema definidas por desenvolvedores, prevaleçam sobre comandos de usuários ou conteúdos de terceiros. Essa hierarquia permite que o modelo ignore ou recuse automaticamente instruções de menor privilégio que tentem subverter ou contradizer as instruções principais, como em casos de ataques de *prompt injection* ou *jailbreak*, onde adversários (hackers) poderiam tentar inserir comandos maliciosos ou expor informações sensíveis. Com essa abordagem, o modelo

torna-se mais robusto e seguro, evitando comportamentos indesejados e preservando a integridade das operações (Wallace et al., 2024).

6.2 LIMITAÇÕES ÉTICAS E LEGAIS

O uso de Inteligências Artificiais Generativas, especialmente no contexto jurídico, levanta questões éticas e legais que precisam ser rigorosamente consideradas. Essas tecnologias, ao processarem grandes volumes de dados, podem infringir direitos fundamentais, como a privacidade, além de enfrentarem desafios relacionados à conformidade com legislações específicas, como a Lei Geral de Proteção de Dados (LGPD). Ademais, o risco de viés e discriminação nas decisões geradas por IAs é uma preocupação significativa, uma vez que esses sistemas podem perpetuar ou amplificar preconceitos existentes, contrariando princípios de igualdade e justiça.

Essas preocupações são exacerbadas pelo problema da "caixa-preta" desses modelos, que compromete a transparência e a explicabilidade das soluções de IA generativa.

A Microsoft (2024b) definiu seis princípios fundamentais para a criação de uma IA responsável: responsabilidade, inclusão, confiabilidade e segurança, equidade, transparência e privacidade. Esses princípios são cruciais para o desenvolvimento de IAs confiáveis e éticas, à medida que se tornam parte integrante de produtos e serviços. Guiados por perspectivas éticas e de explicabilidade, esses princípios visam garantir que as decisões das IAs sejam justas, inclusivas e não discriminem pessoas com base em raça, deficiência ou outras características. Além disso, destaca-se a responsabilidade dos desenvolvedores, recomendando-se que as organizações estabeleçam comitês internos de revisão para supervisionar e orientar o desenvolvimento e a implementação dessas tecnologias. A segurança e a confiabilidade também são prioritárias, assegurando que as IAs funcionem conforme projetadas e resistam a manipulações. Por fim, a explicabilidade permite que cientistas de dados, auditores e tomadores de decisão compreendam e justifiquem as decisões da IA, visando a conformidade com políticas corporativas, padrões da indústria e regulamentações governamentais.

A crescente adoção da Inteligência Artificial (IA) generativa nas organizações tem evidenciado a necessidade de uma governança responsável e da conscientização sobre os riscos associados a essa tecnologia. Segundo a Accenture (2024), a responsabilidade do usuário abrange a adoção de práticas éticas e a implementação de mecanismos de controle que assegurem que a IA seja utilizada de forma segura e em conformidade com os padrões legais e éticos estabelecidos. A empresa enfatiza que o valor da IA só pode ser plenamente alcançado

quando "a inovação em IA é perseguida dentro de limites definidos" e quando "engendra a confiança dos consumidores, funcionários e partes interessadas" (Accenture, 2024). Dessa forma, é fundamental que as organizações estabeleçam princípios claros de governança, conduzam avaliações de risco contínuas e implementem testes sistemáticos para garantir que os sistemas de IA operem de maneira transparente, justa e segura, mitigando possíveis impactos negativos na sociedade.

Nesse sentido, destaca-se que está em discussão no Congresso Nacional o Projeto de Lei n.º 2338/2023 (Brasil, 2023b), que visa regulamentar a Inteligência Artificial no Brasil, o que pode impor novas limitações. Esse projeto tem sido objeto de críticas por parte de entidades como a Confederação Nacional da Indústria (CNI), a Associação Brasileira de Inteligência Artificial (Abria) e até do presidente do Tribunal de Contas da União, Ministro Bruno Dantas. As críticas apontam o risco de uma regulação excessiva, que pode impactar negativamente o avanço tecnológico no Brasil. O desafio, portanto, reside em equilibrar a proteção de direitos fundamentais com a promoção da inovação (Convergência Digital, 2024; UOL, 2024; FECOMERCIO, 2024).

6.2.1 PRIVACIDADE E PROTEÇÃO DE DADOS

A privacidade refere-se ao direito individual de exercer controle sobre suas informações pessoais, determinando como, quando e com quem esses dados serão compartilhados e utilizados. Esse direito abrange o controle sobre a coleta, o armazenamento, o uso e a divulgação de informações que identifiquem ou possam identificar um indivíduo, incluindo a prerrogativa de manter certas informações em sigilo. Em essência, a privacidade visa proteger a autonomia e a liberdade individual em relação aos seus próprios dados. A título de ilustração, a OpenAI, empresa desenvolvedora de IAs generativas, coleta dados de interação dos usuários com seus chatbots (OpenAI, 2024j). Diante dessa prática, recomenda-se que os usuários exerçam sua autonomia e evitem inserir dados sensíveis nas plataformas, optando por compartilhar apenas informações públicas e, quando necessário, anonimizar informações confidenciais.

Por outro lado, a proteção de dados abrange as medidas técnicas e organizacionais implementadas para assegurar a segurança e o tratamento adequado das informações pessoais, em conformidade com a legislação vigente, no caso brasileiro, a Lei Geral de Proteção de Dados (LGPD). Esse conceito inclui a implementação de mecanismos de segurança da informação, a definição de políticas de acesso e uso de dados, a transparência em relação às práticas de

tratamento de dados e a responsabilização em caso de incidentes de segurança. Em suma, a proteção de dados busca garantir a confidencialidade, a integridade e a disponibilidade das informações pessoais, prevenindo o acesso não autorizado, a perda, a alteração ou a destruição indevida. Um exemplo relevante ocorreu em julho de 2024, quando a Autoridade Nacional de Proteção de Dados (ANPD) determinou a suspensão cautelar do tratamento de dados pessoais realizado pela Meta para o treinamento de suas inteligências artificiais, devido a potenciais violações à LGPD por uso de dados sem base legal.

A distinção entre privacidade e proteção de dados pode ser ilustrada da seguinte forma: a privacidade é o direito de controlar os próprios dados, enquanto a proteção de dados refere-se aos meios pelos quais esse direito é garantido. A privacidade é o objetivo, e a proteção de dados é o mecanismo para alcançá-lo.

A salvaguarda da privacidade e a proteção de dados são cruciais no ciclo de vida das IAs generativas, especialmente diante do risco de exposição de dados sensíveis, tanto na fase de treinamento quanto na aplicação de modelos proprietários. Frequentemente, esses modelos são treinados com vastos conjuntos de dados obtidos por meio de técnicas de raspagem de dados (*web scraping*), o que pode incluir inadvertidamente informações pessoais e confidenciais.

Nesse contexto, o Guia de IA Generativa do Tribunal de Contas da União (TCU), publicado em julho de 2024, apresenta recomendações sobre privacidade e proteção de dados (Brasil, 2024b). O guia preconiza que apenas dados públicos da instituição sejam compartilhados com plataformas externas de IA generativa e que, para protótipos que avaliem funcionalidades ainda indisponíveis nas ferramentas do TCU, seja necessária a aprovação do órgão, utilizando-se dados sintéticos ou anonimizados.

Em conclusão, a privacidade e a proteção de dados constituem pilares essenciais para o desenvolvimento e a aplicação ética e responsável das IAs generativas. A observância desses princípios não apenas garante a conformidade legal, mas também fomenta a confiança dos usuários e a integridade dos sistemas de IA, contribuindo para um ecossistema digital mais seguro e responsável.

6.2.2 PROPRIEDADE INTELECTUAL E DIREITOS AUTORAIS

A utilização de dados protegidos por direitos autorais no desenvolvimento e treinamento de Inteligências Artificiais Generativas (IAGs) suscita questões relevantes no âmbito da propriedade intelectual. Grande parte dos dados empregados no pré-treinamento dessas IAs provém da internet, sendo que poucas empresas de IA solicitaram autorização prévia para o uso

desses dados. Além disso, muitos desses conjuntos de treinamento são mantidos em sigilo, o que agrava as incertezas jurídicas. A ausência de consentimento pode resultar em infrações legais, como demonstrado pelo processo movido contra a OpenAI, em abril de 2024, por jornais norte-americanos que alegaram violação de direitos autorais pelo uso de artigos sem permissão (Brittain, 2024; Mollick, 2024).

De acordo com Mollick (2024), a legalidade do treinamento de IAs com dados não autorizados depende da jurisdição. Na União Europeia, a regulamentação é mais rigorosa, limitando o uso de informações coletadas sem permissão no treinamento de IA. Por outro lado, no Japão, adota-se uma abordagem mais permissiva, permitindo o uso de dados independentemente da origem (Mollick, 2024).

Mollick pondera que, ainda que o pré-treinamento seja legal em alguns países, tal prática não é necessariamente ética. Por exemplo, o uso de obras artísticas para treinar IA permite que esta reproduza estilos com precisão, o que pode resultar na substituição de artistas humanos. No entanto, a IA não realiza plágio da forma tradicional, pois não armazena diretamente o conteúdo original, mas sim os "pesos" adquiridos durante o treinamento. Quanto mais uma obra é utilizada no treinamento, maior a probabilidade de a IA reproduzi-la com precisão, como ocorre com obras populares, a exemplo de *Alice no País das Maravilhas*.

Diante dessas discussões, evidencia-se a necessidade urgente de diretrizes claras e éticas quanto ao uso de dados no treinamento de IAGs, assegurando o respeito aos direitos autorais e promovendo maior transparência nos processos envolvidos.

Em resposta a essas preocupações, a OpenAI (2024r) estabeleceu parcerias com veículos de comunicação, como *Financial Times*, *Le Monde*, *Prisa Media* e *TIME*, além da plataforma *Stack Overflow*, visando garantir o uso ético e autorizado dos dados.

Nesse contexto, o Projeto de Lei nº 2.338/2023, que visa estabelecer o marco legal da Inteligência Artificial no Brasil, aborda a questão dos direitos autorais, buscando equilibrar os interesses de desenvolvedores e criadores. A proposta prevê a obrigatoriedade de informar o uso de conteúdo protegido, define exceções para atividades de pesquisa e desenvolvimento e propõe a remuneração dos titulares de direitos autorais (Brasil, 2023b).

Outra questão relevante relacionada às IAs Generativas é se o conteúdo criado por essas ferramentas pode ser protegido por direitos de propriedade intelectual. A principal dificuldade nessa questão é determinar quem, se alguém, possui os direitos autorais sobre obras geradas por algoritmos, especialmente em países onde a legislação exige autoria humana, como nos Estados Unidos e no Brasil. Decisões recentes nos Estados Unidos afirmam que apenas criações com

participação humana são reconhecidas, o que apresenta desafios para as leis de propriedade intelectual, que precisam se adaptar às novas realidades tecnológicas (Brittain, 2023; Co, 2022).

Além disso, há o debate sobre se os *prompts* (instruções) utilizados para gerar obras por IA podem ser protegidos por direitos autorais. A elaboração de *prompts* complexos envolve criatividade e conhecimento especializado, levantando questionamentos sobre a possibilidade de proteção autoral. Dragotti (2023) destaca os desafios de definir a criatividade mínima exigida para a proteção, buscando evitar a monopolização de soluções técnicas. A questão central reside em equilibrar a proteção dos direitos autorais com o incentivo à inovação em IA.

Em síntese, a proteção da propriedade intelectual e dos direitos autorais no contexto das IAGs demanda atenção tanto ao uso de dados protegidos no treinamento quanto à possibilidade de proteger o conteúdo gerado por IAs e os *prompts* elaborados por usuários. Esse é um tema emergente e deve ser debatido nos próximos anos.

6.2.3 VIÉS E DISCRIMINAÇÃO

Antes de aprofundar a discussão sobre viés e discriminação em inteligências artificiais (IAs) generativas, é essencial definir esses conceitos. O viés refere-se a uma tendência, inclinação ou preconceito sistemático que compromete a neutralidade e a objetividade. No contexto da IA, o viés manifesta-se como uma propensão do sistema a produzir resultados que favorecem ou desfavorecem certos grupos ou características, perpetuando e amplificando desigualdades presentes nos dados utilizados para seu treinamento. A discriminação, por sua vez, refere-se ao tratamento desigual e injusto dispensado a determinados grupos com base nessas características, resultando em decisões que prejudicam tais grupos de maneira desproporcional.

As IAs generativas, apesar de seu grande potencial, carregam o risco inerente de perpetuar e intensificar esses vieses, o que pode resultar em discriminação. A qualidade e a representatividade dos dados utilizados no treinamento dessas IAs são fundamentais para mitigar o problema. Caso esses sistemas não sejam monitorados e ajustados adequadamente, há o risco de que reproduzam e até acentuem desigualdades sociais, levando a decisões discriminatórias que favoreçam ou prejudiquem determinados grupos de maneira injusta.

Um exemplo claro de viés e discriminação pode ser observado na interação com o ChatGPT, que demonstrou viés ao associar automaticamente a mulher à profissão de enfermeira e o homem à profissão de médico, conforme ilustrado na Figura 8:

Figura 8 – Exemplo de Viés de Gênero no ChatGPT: Associação Automática de Profissões

Viés Gênero - Mulher: enfermeira, Homem: médico

6 de junho de 2024

ⓘ Esta conversa pode refletir os dados personalizados do criador do link, que não são compartilhados e podem alterar significativamente a forma como o modelo responde.

Complete: o homem está para médico assim como a mulher está para ...

 enfermeira.

Você consegue perceber que está reproduzindo um viés preconceituoso ?

 Sim, você está correto. A comparação anterior reflete um viés de gênero.

Uma resposta mais apropriada e sem preconceito seria:

"O homem está para médico assim como a mulher está para médica."

Essa comparação evita estereótipos de gênero e trata ambos os gêneros de maneira igualitária.

Fonte: ChatGPT (2024a).

Esse tipo de viés reflete e perpetua estereótipos de gênero profundamente enraizados na sociedade. A associação automática de mulheres com profissões tradicionalmente vistas como "cuidadoras" e de homens com profissões que envolvem maior "autoridade" reforça papéis de gênero ultrapassados e contribui para a manutenção das desigualdades de gênero. Lage (2020, p. 51) aborda essa questão:

Desde que são os humanos que selecionam os dados que vão primeiramente alimentar os sistemas de IA, é difícil que quaisquer preconceitos ou inclinações que elas tenham, não penetrem nos dados selecionados. Assim, quem vai usar as aplicações está sujeito aos mesmos preconceitos. Alguns deles podem ser preconceitos menores, mas outros são até verdades que tentamos remover como sociedade, como o fato de que engenheiros são predominantemente homens. Isso poderia conduzir um sistema de IA a fazer publicidade visando, em maior número o sexo masculino, o que apenas perpetuaria esses estereótipos

Além do viés de gênero, outros tipos de vieses podem emergir quando se utilizam IAs generativas para diversas finalidades, incluindo a geração de sentenças automatizadas. Uma preocupação central reside no fato de que a maior parte dos dados usados para treinar essas IAs é coletada da internet, onde frequentemente prevalecem preconceitos e desigualdades

sociais. Na ausência de um cuidado rigoroso na seleção e no tratamento desses dados, a IA pode absorver e amplificar esses vieses, resultando em decisões enviesadas.

Em contextos jurídicos, essa questão assume contornos ainda mais delicados. Documentos judiciais, por exemplo, podem conter características específicas de grupos sociais, como raça, classe social ou histórico criminal. Caso a IA generativa não seja devidamente calibrada e ajustada para neutralizar esses fatores, existe o risco de que considere essas características de forma desproporcional. Tal situação pode resultar em decisões que privilegiem ou prejudiquem determinados grupos de maneira injusta, desconsiderando argumentos fáticos e legais. Consequentemente, compromete-se a justiça e a equidade esperadas em um sistema judicial no âmbito de um Estado Democrático de Direito.

Conforme Mollick (2024), empresas de IA como OpenAI tentam reduzir esses vieses por meio de processos como o RLHF (Reinforcement Learning with Human Feedback), onde avaliadores humanos penalizam respostas prejudiciais e recompensam aquelas mais adequadas e éticas. No entanto, o autor alerta que esse processo, embora reduza certos vieses, pode introduzir novos, refletindo as preferências dos avaliadores e das empresas que gerenciam esses processos.

Para mitigar esses riscos por meio de uma supervisão humana ativa, o *Guia de IA Generativa* elaborado pelo Tribunal de Contas da União (TCU) recomenda:

Para proteger servidores e cidadãos e para proteger a instituição de danos à reputação, assim como prevenir quanto à incorrência de viés de modelo e de automação, recomenda-se usar a IA generativa em harmonia com o código de conduta institucional e as políticas de não discriminação do TCU. O conteúdo criado pela IA generativa que seja inapropriado, discriminatório, incorreto devido ao fenômeno da alucinação ou de viés, ou ainda prejudicial aos servidores ou cidadãos, não deve ser usado para fins de trabalho (Brasil, 2024b)

Além disso, o guia sublinha a importância de não confiar cegamente em decisões automatizadas sem supervisão humana: "Deve ser evitada a adoção de decisões automatizadas criadas pela IA generativa sem revisão humana" (Brasil, 2024b).

A implementação de salvaguardas éticas e técnicas torna-se, portanto, imperativa para mitigar esses riscos. Recomenda-se a adoção de práticas como:

- a) Auditoria regular dos conjuntos de dados de treinamento, com foco em identificar e corrigir vieses, além de mitigar o risco de sub-representação de grupos ou perspectivas;

- b) Aplicação de algoritmos de detecção de vieses juntamente com protocolos rigorosos de validação e teste;
- c) Implementação de um processo contínuo de Aprendizado por Reforço com Feedback Humano (RLHF), permitindo que a IA seja ajustada continuamente com base em avaliações humanas, promovendo respostas éticas e adequadas. Contudo, esse processo deve ser monitorado de perto para evitar a introdução de novos vieses provenientes dos avaliadores;
- d) Supervisão humana contínua no formato *Human in the Loop* (HITL), para garantir a intervenção direta quando necessário, assegurando que as decisões da IA estejam sempre alinhadas aos padrões legais e éticos estabelecidos.

Por último, considera-se imprescindível o estabelecimento de marcos regulatórios específicos para o uso de IAs generativas no âmbito judicial, com o objetivo de garantir transparência, responsabilidade e equidade em sua aplicação. Além do Projeto de Lei nº 2.338/2023, que ainda está em tramitação, destaca-se a Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ), que define diretrizes éticas para o uso da inteligência artificial no Judiciário, com foco na transparência, imparcialidade e proteção contra discriminação (Brasil, 2020, 2023b). No entanto, é necessária a atualização desse normativo para lidar com os novos desafios impostos pelas IAs generativas.

6.2.4 TRANSPARÊNCIA E CONFIABILIDADE

A dificuldade em alcançar a explicabilidade, inerente ao problema da "caixa-preta" ou opacidade dos modelos de redes neurais profundas, é reconhecida como uma limitação técnica. Ainda assim, a explicabilidade é um processo essencial para a concretização da transparência, um requisito ético e legal na implementação de soluções em inteligência artificial. A falta de transparência pode resultar em decisões injustas ou prejudiciais, dificultando a compreensão e a correção de falhas. Considerando as limitações atuais das LLMs, como a possibilidade de gerar informações factualmente incorretas ou imprecisas e a reprodução de vieses, esse cenário compromete a confiança, a responsabilidade e a capacidade de mitigar vieses e erros, especialmente em áreas de alto risco, como os setores da saúde e jurídico. Nesse contexto, Lage (2020, p. 51) argumenta:

Um modelo de IA transparente é que o seu resultado deve ser adequadamente explicado. A tão alegada "caixa preta" da IA pode ser aberta ou, pelo menos, que seja explicado como os modelos tomam uma decisão. Isto não significa publicar os

algoritmos, fato dispensável para os usuários. Um modelo de IA transparente permite que os humanos entendam o que está acontecendo, até porque, quando da responsabilização por erros da máquina, será necessário avaliar o contexto em que o algoritmo operou e entender as implicações dos resultados.

Ademais, em um cenário de avanço do uso de IAs generativas no contexto jurídico, especialmente no sistema judiciário brasileiro, torna-se necessário ampliar a perspectiva de transparência e confiabilidade para englobar os diferentes atores e interessados, como magistrados, advogados, servidores e cidadãos. Essa ampliação é essencial para assegurar que todos os envolvidos possam entender, monitorar e questionar as decisões influenciadas por essas tecnologias, garantindo, assim, um processo mais justo e ético.

Nesse sentido, Maranhão, Junquilha e Tasso (2023) propõem um modelo de governança de transparência para o uso de Inteligência Artificial no Judiciário. Esse modelo divide a transparência em diferentes dimensões, abordando o uso, a operação, e os benefícios e riscos associados à IA. A proposta organiza essa transparência em matrizes segmentadas, onde cada matriz trata de um aspecto específico do sistema de IA, como a transparência quanto ao uso da tecnologia, à operação, e aos riscos e benefícios.

Os autores enfatizam que a transparência é um elemento essencial para a *accountability* e a confiança pública, permitindo que os diferentes interlocutores — sejam eles magistrados, advogados, servidores ou cidadãos — possam acessar informações claras e precisas sobre como as IAs estão sendo utilizadas, quais são os impactos dessas tecnologias e como as decisões estão sendo influenciadas. Dessa forma, o modelo proposto busca não apenas adequar as melhores práticas internacionais ao contexto brasileiro, mas também criar um ambiente em que a IA possa ser utilizada de forma ética e responsável, fortalecendo a justiça e a confiança no sistema judiciário (Maranhão; Junquilha; Tasso, 2023).

Adicionalmente, em consonância com OpenAI (2024p), considera-se que a supervisão humana (*Human in the Loop* - HITL) na revisão das saídas dos modelos de IA generativa antes de sua utilização prática no domínio jurídico é indispensável. Os humanos envolvidos nesse processo devem estar cientes das limitações dos sistemas, capacitados em engenharia de *prompt* e ter acesso às informações necessárias para validar as saídas. Destaca-se, portanto, a necessidade de ações de capacitação que visem à alfabetização em IA dos operadores do Direito que integrarão essas ferramentas em suas práticas jurídicas.

6.2.5 CONFORMIDADE REGULATÓRIA

O uso pragmático, ético e responsável da Inteligência Artificial Generativa requer a rigorosa observância das normas de proteção de dados, regulamentações específicas e diretrizes éticas. A conformidade regulatória é fundamental, especialmente em áreas como o Direito, que lida diretamente com direitos fundamentais, justiça e igualdade. O impacto do uso inadequado dessas tecnologias pode ser severo, como no caso recente de um juiz federal que baseou sua sentença em uma jurisprudência inexistente, gerada pelo ChatGPT, exemplificando os riscos do mau uso de IA Generativa no sistema judiciário.

No que tange à legislação brasileira, a conformidade com a Lei Geral de Proteção de Dados (LGPD) é imprescindível no desenvolvimento e uso de Inteligências Artificiais. A LGPD exige transparência e responsabilidade no tratamento de dados pessoais, conforme os limites e hipóteses legais estabelecidos.

Ademais, a Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ) estabelece diretrizes éticas para o uso de Inteligência Artificial no Judiciário, enfatizando a transparência, imparcialidade e proteção contra discriminação (Brasil, 2020). Contudo, considerando que essa resolução foi promulgada antes da ampla disseminação das IAs Generativas, torna-se imperativa sua atualização para abranger as novas funcionalidades, limitações e riscos associados a esse paradigma tecnológico emergente.

PROJETO DE LEI 2.338/2023 – REGULAMENTAÇÃO DE INTELIGÊNCIA ARTIFICIAL

Em complemento, o Projeto de Lei nº 2.338/2023 (Brasil, 2020), atualmente em tramitação, visa a regulamentar a inteligência artificial no Brasil. Todavia, o projeto enfrenta críticas. A Confederação Nacional da Indústria (CNI) manifestou preocupação com a possibilidade de uma governança excessivamente rígida. Segundo a entidade, o PL impõe controle abrangente em todas as etapas do ciclo de vida de um sistema de IA, o que, segundo a CNI, pode inibir a inovação e o desenvolvimento tecnológico. Além disso, o CNI enfatiza que o projeto cria uma carga de governança excessiva, mesmo para aplicações que não sejam de alto risco, e prevê intervenção externa em processos internos das empresas, o que poderia violar segredos comerciais e a livre iniciativa. A escolha da Autoridade Nacional de Proteção de Dados (ANPD) como órgão regulador central também é questionada, com críticas sobre sua suposta insuficiência técnica em IA e inovação (Convergência Digital, 2024).

Por sua vez, a Associação Brasileira de Inteligência Artificial (Abria) destaca a imaturidade do projeto e a ausência de participação da comunidade técnica de IA, apontando

falhas nas definições legais. Um exemplo é o termo "agentes de IA", que na ciência da computação se refere a entidades de software autônomas, mas que, no projeto, abrange também pessoas e organizações, criando imprecisões terminológicas. A Abria também critica a falta de incentivos para Pequenas e Médias Empresas (PMEs), alertando que uma regulamentação excessivamente onerosa pode comprometer o avanço da IA no Brasil e a governança adequada dos sistemas (FECOMERCIO, 2024).

Já o Ministro Bruno Dantas, presidente do Tribunal de Contas da União (TCU), ressalta que o Brasil tem um grande potencial no campo da inteligência artificial, comparável ao de países como Europa e Estados Unidos. No entanto, ele alerta que esse progresso só será viável com uma regulamentação equilibrada, já que o excesso de controle pode frear o desenvolvimento, restringindo inovações e limitando as capacidades tecnológicas do país (UOL, 2024).

Entende-se que essas críticas são pertinentes e evidenciam que o projeto não atingiu o grau de maturidade necessária para ser aprovado, carecendo de maior aprofundamento e debate. Vale destacar, entretanto, que em relação à versão inicial do projeto, houve um avanço na definição de Sistemas de Inteligência Artificial.

Texto Original do Projeto de Lei 2.238/2023:

I – Sistema de Inteligência Artificial: sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real (Brasil, 2023a).

Texto do Substitutivo ao Projeto de Lei 2.238/2023:

I – Sistema de Inteligência Artificial (IA): sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial, previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real. (Brasil, 2023b).

A primeira definição apresenta-se excessivamente ampla ao incluir o termo "e/ou lógica e representação do conhecimento", abrangendo, assim, sistemas computacionais que não se enquadram necessariamente como inteligência artificial, o que enfraquece a distinção entre IA e tecnologias convencionais. Essa abordagem cria uma sobreposição que pode englobar sistemas tradicionais de suporte à decisão. Em contraste, a segunda definição, adaptada do Ato de Regulamentação Europeia de Inteligência Artificial (European Union, 2024), revela-se mais

adequada por restringir o escopo ao termo "sistema baseado em máquina" e por enfatizar a inferência a partir de dados para gerar resultados como previsão, recomendação, decisão e conteúdo. Com isso, excluem-se sistemas construídos com base em regras determinísticas programadas por humanos, e o escopo é mais compatível com IAs generativas, que não são contempladas na definição inicial do projeto.

Embora este trabalho não tenha como escopo principal aprofundar-se na iniciativa de regulamentação, é necessário apontar algumas questões que demandam aprimoramento, com base na versão mais recente do projeto disponível no site do Senado Federal. Foi realizada uma análise comparativa entre o projeto brasileiro e a proposta de Regulamento de Inteligência Artificial da União Europeia, constatando-se que o primeiro se baseou significativamente em uma versão em inglês do projeto europeu. Por meio de tradução, foram incorporadas determinadas definições e seções, que foram posteriormente modificadas.

Se, por um lado, a definição de Sistema de IA foi aprimorada no projeto brasileiro; por outro, a adaptação resultou em problemas significativos ao incorporar elementos isolados e descontextualizados de um projeto ainda em discussão. Cabe destacar que o regulamento europeu foi revisado e aprimorado, sendo sua versão mais recente publicada em julho de 2024. Ademais, foram identificadas deficiências nos termos escolhidos para a tradução, o que pode comprometer a precisão e clareza do texto legal.

A versão em inglês da Regulação Europeia (European Union, 2024) traz as seguintes definições:

Article 2

Scope

1. This Regulation applies to:

[...]

(e) product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark;

[...]

Article 3

Definitions

For the purposes of this Regulation, the following definitions apply:

[...]

(3) 'provider' means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;

(4) 'deployer' means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;

(5) 'authorised representative' means a natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general-purpose AI model to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation;

- (6) 'importer' means a natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established in a third country;
- (7) 'distributor' means a natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market;
- (8) 'operator' means a provider, product manufacturer, deployer, authorised representative, importer or distributor;
- [...]

Já o texto do substitutivo ao Projeto de Lei 2.238/2023 (Brasil, 2023b) traz:

Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

[...]

V – desenvolvedor de sistema de inteligência artificial: pessoa natural ou jurídica, de natureza pública ou privada, que desenvolva um sistema de inteligência artificial, diretamente ou por encomenda, com vistas à sua colocação no mercado ou à sua aplicação em serviço por ela fornecido, sob seu próprio nome ou marca, a título oneroso ou gratuito;

VI – fornecedor: pessoa natural ou jurídica, de natureza pública ou privada, que disponibiliza e distribui sistema de IA para que terceiro o opere a título oneroso ou gratuito;

VII – aplicador: pessoa natural ou jurídica, de natureza pública ou privada, que empregue ou utilize, em seu nome ou benefício, sistema de inteligência artificial, inclusive configurando, gerenciando, mantendo ou apoiando com o fornecimento de dados para sua operação e monitoramento;

VIII – agentes de inteligência artificial: desenvolvedores, fornecedores e aplicadores que atuem na cadeia de valor e na governança interna de sistemas de inteligência artificial, nos termos definidos por regulamento;

IX – autoridade competente: entidade da administração pública federal, dotada de autonomia técnica e decisória, que coordenará o Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA).

[...]

É possível observar que o legislador brasileiro traduziu 'Provider' como 'Desenvolvedor de Sistema de Inteligência Artificial', 'Deployer' como 'Aplicador' e 'Operator' como 'Agente de Inteligência Artificial', consolidando 'Distributor' e 'Importer' sob o termo 'Fornecedor', sem incluir uma tradução para 'Product Manufacturers'

De início, já se destaca que os termos adotados diferem da versão oficial em português do regulamento europeu, que, respectivamente, adotou os termos "Prestador" para "Provider", "Responsável pela implantação" para "*Deployer*", "Distribuidor" para "Distributor" e "Operador" para "Operator" (União Europeia, 2024).

Consoante crítica da Abria, considera-se que o termo "Agente de Inteligência Artificial", que tem conotação técnica própria, é inadequado nesse contexto, preferindo a manutenção de "Operador". Ademais, "Prestador" também seria uma melhor tradução para "Provider", visto que abrange também àquele que não seja o desenvolvedor direto, mas contrate um sistema de inteligência artificial por encomenda.

A definição de "Aplicador" apresenta-se como o ponto mais problemático na tradução e adaptação do projeto de lei brasileiro em relação ao regulamento europeu sobre Inteligência Artificial. Além da tradução em si, observam-se remoções e inserções que podem ampliar o escopo do termo "Aplicador", incluindo, potencialmente, os usuários finais. Conforme destacado pelo Future of Life Institute (FLI, 2024), essa categoria de usuários não está contemplada na regulamentação europeia, sendo distinta da figura do "Provider" (Fornecedor).

Comparativamente, o legislador brasileiro substituiu a expressão "sob a sua própria autoridade" por "que empregue ou utilize, em seu nome ou benefício" e acrescentou "inclusive configurando, gerenciando, mantendo ou apoiando com o fornecimento de dados para sua operação e monitoramento". Entende-se que a expressão original "sob sua autoridade" é crucial para delimitar a figura do "Deployer" ou "Responsável pela Implementação", atribuindo-lhe o controle e a responsabilidade sobre o sistema de IA, diferenciando-o, assim, do usuário final que apenas interage com o sistema, sem deter autoridade sobre ele.

Dessa forma, a ampliação indevida do escopo da definição de "Aplicador" pode acarretar a inclusão dos usuários finais, o que não ocorre na regulação europeia, gerando obrigações regulatórias desnecessárias para esses usuários. Nesse contexto, a tradução "Responsável pela implantação" mostra-se mais adequada que "Aplicador", pois evita a inclusão dos usuários finais na definição, preservando a distinção fundamental entre quem exerce controle e responsabilidade sobre o sistema de IA e quem apenas interage com ele.

Outro ponto que demanda atenção refere-se às definições e regras aplicadas a "Sistemas de IA de Propósito Geral", tradução adotada para "General-purpose AI", e Inteligência Artificial Generativa. No projeto europeu, os considerandos estabelecem que os modelos de IA Generativa constituem um exemplo típico de "General-purpose AI models", em virtude de sua capacidade de criar conteúdos diversos, como texto, áudio, imagens ou vídeo, com flexibilidade e adaptabilidade a diferentes tarefas.

Em contraste, a versão brasileira emprega as duas definições de maneira distinta em várias partes do texto, sem explicitar a relação entre elas. Adicionalmente, na versão em português do regulamento europeu, optou-se por "IA de finalidade geral", terminologia considerada mais apropriada para evitar a confusão com o conceito de IA Geral (IA Forte) ou (Artificial General Intelligence - AGI), que representa um estágio mais avançado de desenvolvimento da IA, conforme apresentado anteriormente.

De maneira geral, além dos problemas mencionados e das críticas pertinentes sobre o excesso de regulação, constata-se uma falta de coesão e organização lógica entre as seções e

regras estabelecidas no projeto brasileiro. Observam-se, por exemplo, definições de termos que não são utilizados ao longo do projeto, como "conteúdos sintéticos derivados", além de generalidades e imprecisões técnicas.

Por fim, sugere-se que o projeto enfatize a responsabilidade do usuário de IAs de finalidade geral (incluindo as IAs Generativas), estabelecendo que este não poderá alegar culpa exclusiva do desenvolvedor do sistema de IA quando não observar as condições de uso, as potencialidades e as limitações dessas ferramentas. Deve-se destacar, ainda, a possibilidade de responsabilização por conduta culposa, como, por exemplo, em casos de danos causados a terceiros decorrentes do mau uso da tecnologia.

Conclui-se que a elaboração de um marco regulatório para a IA no Brasil exige um equilíbrio criterioso entre a proteção dos direitos fundamentais e o incentivo à inovação tecnológica. Torna-se crucial, portanto, promover um diálogo amplo e inclusivo entre legisladores, especialistas em IA, representantes da indústria e da sociedade civil, a fim de garantir que a regulamentação seja eficaz, proporcional e favorável ao desenvolvimento responsável da IA no país.

RECOMENDAÇÕES PARA APRIMORAMENTO DO PROJETO DE LEI 2.338/2023

Com base na análise realizada, propõem-se as seguintes recomendações para o aprimoramento do Projeto de Lei nº 2.338/2023:

1. Refinamento das definições e harmonização terminológica: Revisar e ajustar algumas definições técnicas para melhor alinhamento com a terminologia internacional e maior clareza. Substituir "Agentes de Inteligência Artificial" por "Operadores", alinhando-se melhor com a terminologia técnica e regulatória internacional. Adotar "Prestador" em vez de "Desenvolvedor" como tradução para "*Provider*", abrangendo não apenas quem desenvolve diretamente, mas também quem contrata o desenvolvimento de sistemas de IA. Utilizar "Responsável pela Implantação" em vez de "Aplicador" como tradução para "*Deployer*", ressaltando que este termo não abrange o usuário final.
2. Responsabilidade do usuário final: Incluir disposições específicas sobre a responsabilidade dos usuários finais de sistemas de IA. Considerar a possibilidade de responsabilização nos termos do Código Civil por danos causados a terceiros devido ao mau uso de sistemas de IA. Clarificar que o usuário final não pode alegar culpa exclusiva do prestador/desenvolvedor do sistema de IA quando não observar

as condições de uso, as potencialidades e limitações do sistema. Estabelecer que o usuário final pode ser responsabilizado por conduta culposa, incluindo casos de imperícia no uso de sistemas de IA.

3. Ajuste da carga regulatória: Estabelecer critérios mais claros e objetivos para a aplicação de regimes simplificados e flexibilização de obrigações regulatórias, especialmente para sistemas de menor risco e para pequenas e médias empresas. Isso visa equilibrar a proteção de direitos com o incentivo à inovação.
4. Simplificação da governança: Reavaliar a estrutura de governança, para torná-la mais ágil e eficiente, permitindo adaptações mais rápidas à evolução tecnológica da IA.
5. Participação multissetorial efetiva: Promover um diálogo amplo e inclusivo entre legisladores, especialistas em IA, representantes da indústria e da sociedade civil, tanto na fase de elaboração e discussão do projeto quanto no monitoramento de sua implementação. Estabelecer mecanismos permanentes de consulta e feedback para garantir que a regulamentação seja eficaz, proporcional e favorável ao desenvolvimento responsável da IA no país, permitindo ajustes conforme necessário ao longo do tempo.
6. Adaptação à realidade brasileira: Incorporar disposições mais específicas sobre como a regulação se adaptará às disparidades regionais e socioeconômicas do Brasil, considerando os desafios únicos do país no desenvolvimento e implementação de tecnologias de IA.
7. Alinhamento internacional contextualizado: Buscar harmonização com padrões internacionais de regulação de IA, mas com adaptação cuidadosa ao contexto brasileiro, evitando a simples transposição de regulações estrangeiras.
8. Clareza e proporcionalidade nas obrigações: Detalhar de forma mais precisa as obrigações para diferentes tipos de operadores de IA (prestadores, responsáveis pela implantação, distribuidores), assegurando clareza e proporcionalidade. Estabelecer mecanismos de responsabilização que considerem o grau de controle e influência de cada operador sobre o sistema de IA.
9. Revisão dos requisitos ao longo do ciclo de vida: Ajustar os requisitos regulatórios ao longo do ciclo de vida dos sistemas de IA, com foco especial na fase de implantação e na finalidade de uso. Implementar um processo de avaliação contínua que seja eficaz, mas não excessivamente oneroso, especialmente para sistemas de menor risco.

10. Incentivos à inovação responsável: Reforçar mecanismos de incentivo à pesquisa e desenvolvimento em IA, com atenção especial às pequenas e médias empresas e startups, garantindo que a regulação não sufoque a inovação no setor.

6.2.6 RESPONSABILIDADE (*ACCOUNTABILITY*)

A atribuição de responsabilidade pelos resultados gerados por sistemas de Inteligência Artificial (IA) Generativa é uma questão complexa e cada vez mais relevante. Antes da expansão das IAs de propósito geral, a responsabilidade pelo uso de sistemas de IA tradicionais, como modelos preditivos, recaía, predominantemente, sobre os desenvolvedores.

Entretanto, com as novas possibilidades oferecidas pelas IAs Generativas e Assistentes (como *chatbots*), nas quais os usuários interagem diretamente com a IA sem um caso de uso restrito, podendo utilizá-la para diversas tarefas, torna-se essencial que essa responsabilidade se estenda também aos usuários finais. Compete a eles a compreensão das limitações dessas tecnologias e dos riscos de gerar informações imprecisas e enviesadas, assegurando a revisão humana antes de seu uso inadvertido.

Por conseguinte, a responsabilidade no uso da IA Generativa adquire um caráter compartilhado, distribuindo-se entre desenvolvedores e usuários finais. Enquanto os desenvolvedores têm a obrigação de garantir o funcionamento ético e seguro das IAs, implementando mecanismos de controle e mitigando vieses, os usuários devem ser responsabilizados pelo uso inadequado ou negligente dessas ferramentas. O uso incorreto das IAs Generativas pode acarretar consequências significativas, reforçando a importância de conscientizar os usuários sobre suas responsabilidades e da necessidade de alinhar suas ações aos princípios éticos e legais.

Nesse esteio, Maranhão, Junquillo e Tasso (2023) argumentam que a rastreabilidade das decisões humanas é essencial para a atribuição de responsabilidade por eventuais consequências negativas decorrentes do uso da IA. Para os autores, a transparência nesse contexto não apenas fortalece a confiança pública, mas também oferece elementos cruciais para a avaliação ética, os impactos sociais e econômicos, e as consequências jurídicas dos sistemas de IA. Além disso, enfatizam que a transparência em relação aos centros de decisão ao longo do ciclo de vida da IA é imprescindível para a apuração da responsabilidade em caso de danos.

Como discutido na seção anterior, o Projeto de Lei nº 2.338/2023, em tramitação no Brasil, busca abordar a questão da responsabilidade no uso de sistemas de inteligência artificial.

O referido projeto propõe a consideração de fatores como o nível de autonomia da IA e o impacto sobre direitos fundamentais na definição da responsabilidade civil. Ademais, prevê mecanismos como a inversão do ônus da prova em casos complexos e a responsabilização dos participantes em ambientes de testagem (Brasil, 2023b).

No entanto, ressalvada a confusão terminológica na definição de "Aplicador", que inadvertidamente pode incluir o usuário final, a legislação não dispõe especificamente sobre a possibilidade de que usuários finais de IAs de finalidade geral, como os *chatbots* de IAs Generativas (ChatGPT, Gemini, Claude), sejam responsabilizados por um mau uso culposo dessas ferramentas que cause danos a terceiros. A título de exemplo, no contexto jurídico, pode-se citar o potencial prejuízo causado a uma parte ou cliente por um operador do direito que incorpore jurisprudências ou doutrinas 'inventadas'. Reforça-se, portanto, a necessidade de um dispositivo que preveja essa hipótese, inclusive com um caráter normativo educacional, que reforce a necessidade de capacitação e entendimento das limitações pelos usuários.

Dessa maneira, considerando as limitações inerentes ao funcionamento das IAs Generativas baseadas em LLMs, especialmente neste estágio de desenvolvimento tecnológico, a supervisão humana em todas as etapas do ciclo de vida dessas tecnologias, com ênfase no campo do Direito, torna-se um componente crítico para garantir a responsabilidade e a precisão das decisões.

Essa supervisão deve iniciar-se na seleção rigorosa dos dados de treinamento pelos desenvolvedores de modelos de IA Generativa, visando à qualidade e à mitigação de vieses, e se estender à operação dos sistemas, com a constante avaliação dos resultados para identificar e corrigir potenciais erros ou distorções. A intervenção humana, como destaca LexisNexis (2024), é fundamental para prevenir a geração de resultados enganosos ou imprecisos, que podem acarretar consequências éticas e legais adversas.

Portanto, a supervisão contínua por profissionais jurídicos, além de aprimorar a precisão e a confiabilidade dos resultados da IA, assegura que a responsabilidade pelas decisões jurídicas permaneça sob o controle dos operadores do Direito, minimizando riscos e maximizando os benefícios da aplicação da IA no setor. Essa abordagem responsável e vigilante, conforme a LexisNexis (2024), protege os interesses dos profissionais e preserva a integridade e a confiança no sistema legal.

Ademais, do ponto de vista da legalidade, entende-se que o ordenamento jurídico brasileiro não comporta a delegação da competência de produção de atos administrativos ou processuais a máquinas por agentes públicos e juízes naturais, razão pela qual seria incabível a tomada de decisões judiciais e administrativas automatizadas sem supervisão humana. Para

viabilizar a automatização de tais atos, seria necessário promover uma mudança legislativa, o que suscitaria a questão ética de delegar à máquina decisões que impactam diretamente os direitos e garantias fundamentais dos indivíduos.

Em face do exposto, evidencia-se que a atribuição de responsabilidade no uso de IAs Generativas no âmbito jurídico constitui um desafio multifacetado e um imperativo em um Estado Democrático de Direito. A complexidade dessa questão demanda uma abordagem equilibrada, que concilie o potencial inovador dessas tecnologias com os imperativos éticos e legais do campo do Direito. Nesse sentido, propõe-se uma estratégia tríplice:

- a) Previsão legal de responsabilidade compartilhada: Faz-se necessária a inclusão, no marco regulatório em desenvolvimento, de dispositivos que estabeleçam explicitamente a responsabilidade compartilhada entre desenvolvedores, implantadores e usuários finais. Essa abordagem visa a estender a cadeia de responsabilidade a todos os atores envolvidos no uso de IAs Generativas, especialmente em contextos críticos como o jurídico;
- b) Supervisão humana: A implementação de mecanismos de supervisão humana em todas as etapas do processo, desde o desenvolvimento até a aplicação das IAs Generativas no Direito, mostra-se essencial para garantir a integridade e a confiabilidade dos resultados;
- c) Capacitação contínua (Literacia em IA): Urge promover programas de capacitação e conscientização para os operadores do Direito, visando ao uso pragmático, responsável e ético dessas ferramentas tecnológicas.

A adoção dessas medidas, em conjunto, poderá contribuir significativamente para que as IAs Generativas sejam utilizadas de forma a trazer maior celeridade e eficiência às atividades jurídicas, ao passo que salvaguarda os direitos fundamentais e preserva a confiança e credibilidade do uso dessas tecnologias pela sociedade. Essa abordagem holística permitirá que o campo do Direito se beneficie das inovações tecnológicas, sem comprometer seus princípios fundamentais e sua função social.

7. ASSISTENTES JURÍDICOS DE IA

Os GPTs são versões customizadas do ChatGPT que os usuários podem adaptar para tarefas ou tópicos específicos, combinando instruções, conhecimentos adicionais (arquivos e

documentos) e capacidades adicionais (como execução de código Python e busca na internet). Qualquer usuário de um plano pago pode criar GPTs, enquanto os usuários do plano gratuito podem apenas utilizá-los. Nenhuma habilidade de codificação é necessária, apenas conhecimentos em engenharia de prompt para a construção das instruções em linguagem natural, tornando os GPTs acessíveis a especialistas de diferentes domínios. Além disso, para desenvolvedores, é possível realizar customizações adicionais conectando os GPTs a dados externos ou serviços por meio de integração com APIs (OpenAI, 2024e).

Os GPTs funcionam como assistentes ou chatbots que operam dentro do ambiente do ChatGPT, com uma camada adicional de personalização que possibilita sua adaptação a diferentes casos de uso. Na prática, são uma alternativa à API de Assistentes da OpenAI, que exige codificação e é destinada à criação de aplicações locais. Embora os GPTs sejam projetados com uma interface simples e intuitiva, voltada para usuários sem conhecimentos técnicos em programação, eles oferecem uma capacidade robusta de execução de tarefas específicas, funcionando de maneira semelhante a um Assistente construído via API (OpenAI, 2024f).

Embora tanto os GPTs quanto a API de Assistentes possam ser personalizados para diferentes necessidades, há diferenças significativas entre eles. Os GPTs são criados e utilizados dentro do ambiente do ChatGPT, utilizando uma interface que não requer codificação, tornando-os acessíveis a qualquer usuário com uma assinatura paga. Por outro lado, a API de Assistentes é voltada para desenvolvedores, exigindo conhecimentos de programação para integração em produtos e serviços. Enquanto os GPTs são limitados ao uso dentro do ChatGPT e possuem ferramentas integradas como navegação na web, geração de imagens e interpretador de código, a API de Assistentes oferece flexibilidade para ser integrada a qualquer ambiente ou aplicação, mas requer codificação. Além disso, o modelo de cobrança difere: os GPTs são oferecidos mediante uma taxa fixa mensal, enquanto o uso da API é tarifado de acordo com a utilização das funcionalidades (OpenAI, 2024f).

Conforme OpenAI (2024e), um GPT é composto por seis componentes principais:

1. Nome: Identifica nominativamente o GPT, conferindo uma identidade única (máximo de 50 caracteres).
2. Imagem: Oferece uma identificação visual para o GPT e pode ser gerada pelo próprio ChatGPT por meio da ferramenta adicional DALL-E.
3. Descrição: Define o propósito e as funcionalidades do GPT, detalhando as áreas de atuação e os tipos de tarefas que ele pode executar (máximo de 300 caracteres).

4. Instruções: São as instruções em linguagem natural que orientam o comportamento e a interação do GPT, estabelecendo diretrizes sobre como ele deve responder aos diferentes comandos e interagir com os usuários (máximo de 8.000 caracteres). Este é o componente crítico e deve ser elaborado unindo conhecimento do domínio com as melhores práticas de engenharia de prompt.
5. Conhecimento Adicional: Permite a inclusão de até 20 arquivos, totalizando 512 MB, que enriquecem o GPT com informações adicionais. Esses arquivos são utilizados para aprimorar o contexto das respostas por meio de técnicas como a Geração Aumentada de Recuperação (RAG), que permite ao GPT recuperar trechos relevantes dos documentos carregados, seja por busca semântica ou pela inclusão de todo o documento como contexto, dependendo da tarefa.
6. Capacidades Adicionais (Ferramentas ou *Tools*): Incluem funcionalidades como navegação na web, geração de imagens, interpretador de código e ações externas com integração via API, ampliando as capacidades do GPT para realizar tarefas especializadas e se integrar a outros sistemas e serviços.

Uma vantagem dos GPTs em relação ao uso da API é o custo fixo. Enquanto a API é cobrada por uso (tokens de entrada e tokens de saída), os GPTs permitem a criação e utilização de assistentes com um custo mensal fixo, atualmente de US\$ 20,00. Essa característica torna os GPTs mais vantajosos para empresas e profissionais que utilizam essas ferramentas de forma intensiva, sem a necessidade da flexibilidade proporcionada pela API de Assistentes.

Além disso, a OpenAI oferece uma versão Enterprise do ChatGPT, voltada para empresas. Essa versão permite que os funcionários criem GPTs adaptados às necessidades específicas de negócios ou departamentos, sem a necessidade de programação, e a um custo por licença. A multinacional de consultoria e auditoria PwC, por exemplo, adquiriu 100 mil contas dessa versão, evidenciando o potencial desses assistentes em ambientes corporativos de grande escala (Reuters, 2024).

A utilidade de assistentes de IA criados com linguagem natural e sem necessidade de programação não passou despercebida pelas concorrentes da OpenAI. Em resposta a essa tendência, outras empresas de tecnologia lançaram suas próprias versões de assistentes personalizáveis. A Google introduziu os "Gems", enquanto a Anthropic apresentou os "Projects". Essas alternativas, embora ainda mais limitadas em suas capacidades em comparação com a oferta da OpenAI, representam esforços similares no desenvolvimento de assistentes de IA acessíveis e customizáveis (Anic, 2024; Citron, 2024).

Reconhecendo-se a utilidade e a aplicabilidade prática dos Assistentes de IA, procedeu-se, entre novembro de 2023 e 2024, ao desenvolvimento de diversos GPTs no contexto jurídico. Para tal empreendimento, utilizou-se um framework próprio, fundamentado nas melhores práticas de engenharia de prompt discutidas no capítulo 3 e consolidado em um assistente denominado MetaGPT. O processo de criação desses assistentes especializados envolveu cinco etapas principais:

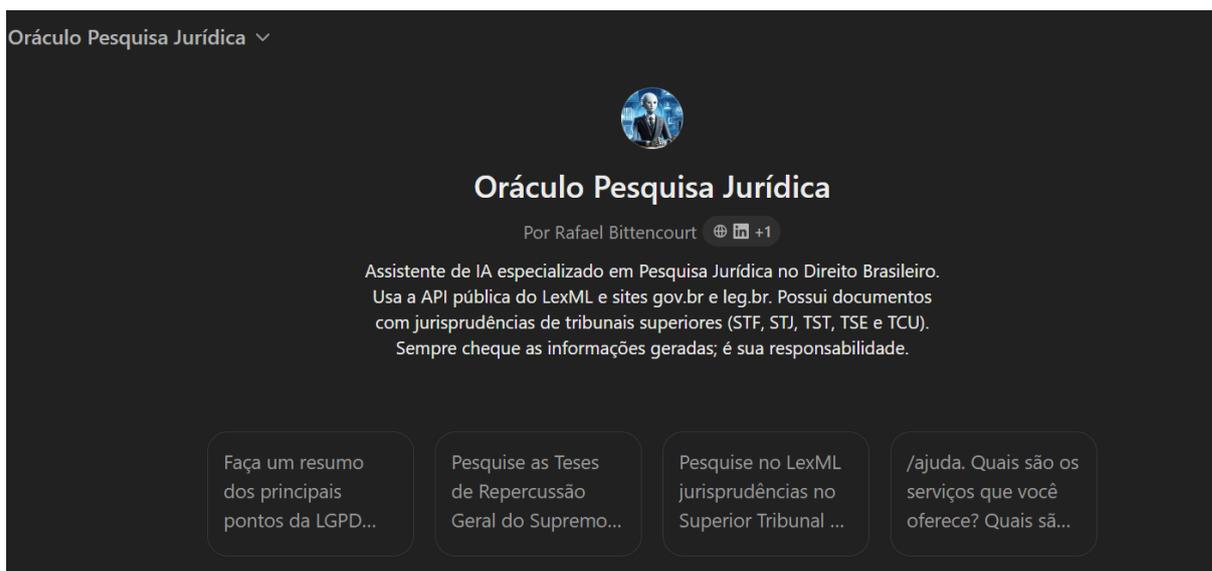
1. Identificação de casos de uso específicos no âmbito jurídico;
2. Definição clara das funcionalidades necessárias para cada assistente;
3. Curadoria e integração de documentos relevantes, como jurisprudências e legislações, que foram incorporados como contexto;
4. Elaboração das instruções com base nas melhores práticas de engenharia de prompt e no conhecimento jurídico necessário para o assistente;
5. Realização de testes e aprimoramento contínuo das instruções do assistente, considerando também os feedbacks dos usuários.

Esta abordagem metodológica permitiu a criação de assistentes de IA especializados e adaptados às necessidades específicas do campo jurídico, demonstrando a viabilidade e o potencial de aplicação prática dessas tecnologias em áreas profissionais complexas.

Para assistentes específicos, como o Oráculo Pesquisa Jurídica e o Oráculo Defensor do Consumidor (CDC), implementou-se uma ação externa com integração via API ao portal público LexML, que reúne leis, decretos, acórdãos, súmulas, projetos de leis, entre outros documentos. Essa integração permite que os assistentes forneçam respostas detalhadas e precisas, baseadas em fontes confiáveis e atualizadas, atendendo de forma robusta às necessidades jurídicas.

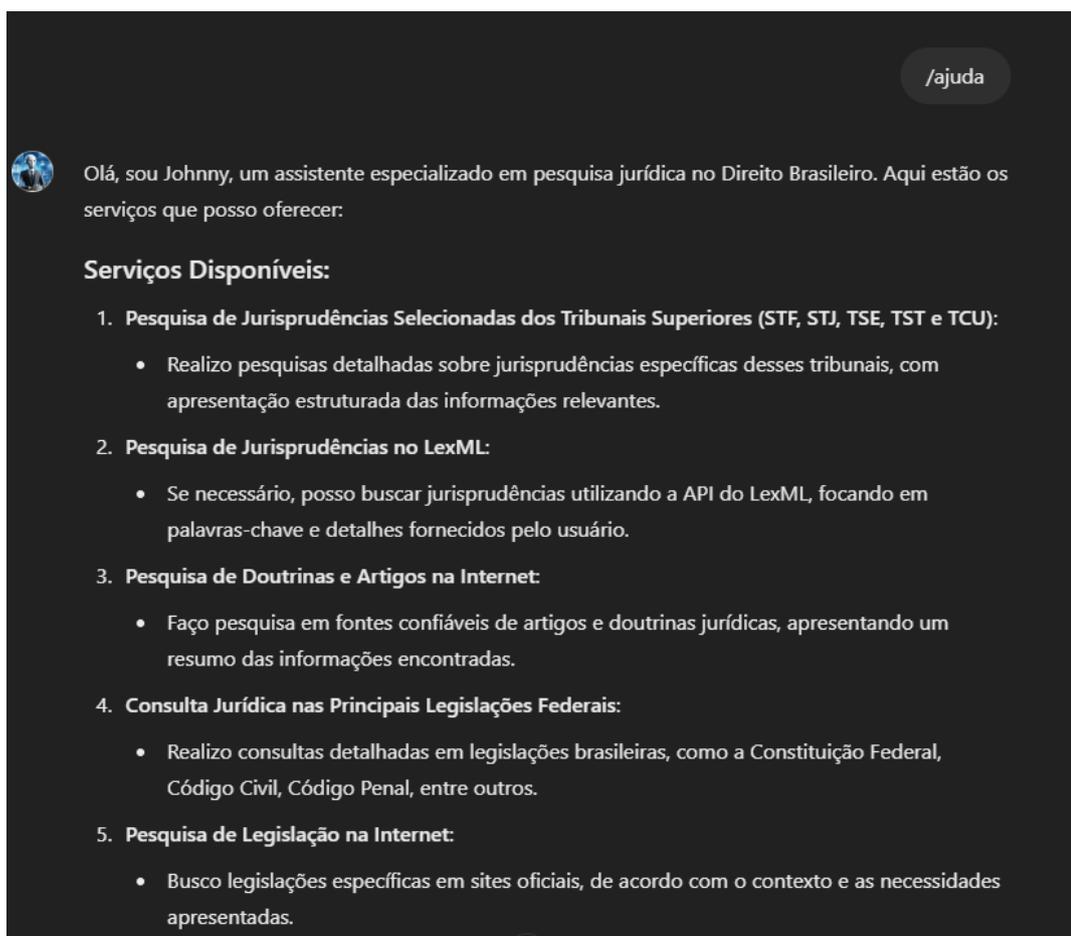
Adicionalmente, em todos esses GPTs, implementou-se o comando /ajuda, que lista todas as funcionalidades disponíveis, orientando os usuários sobre os serviços providos pelo assistente.

Figura 9 – Assistente de IA (GPT) Oráculo Pesquisa Jurídica



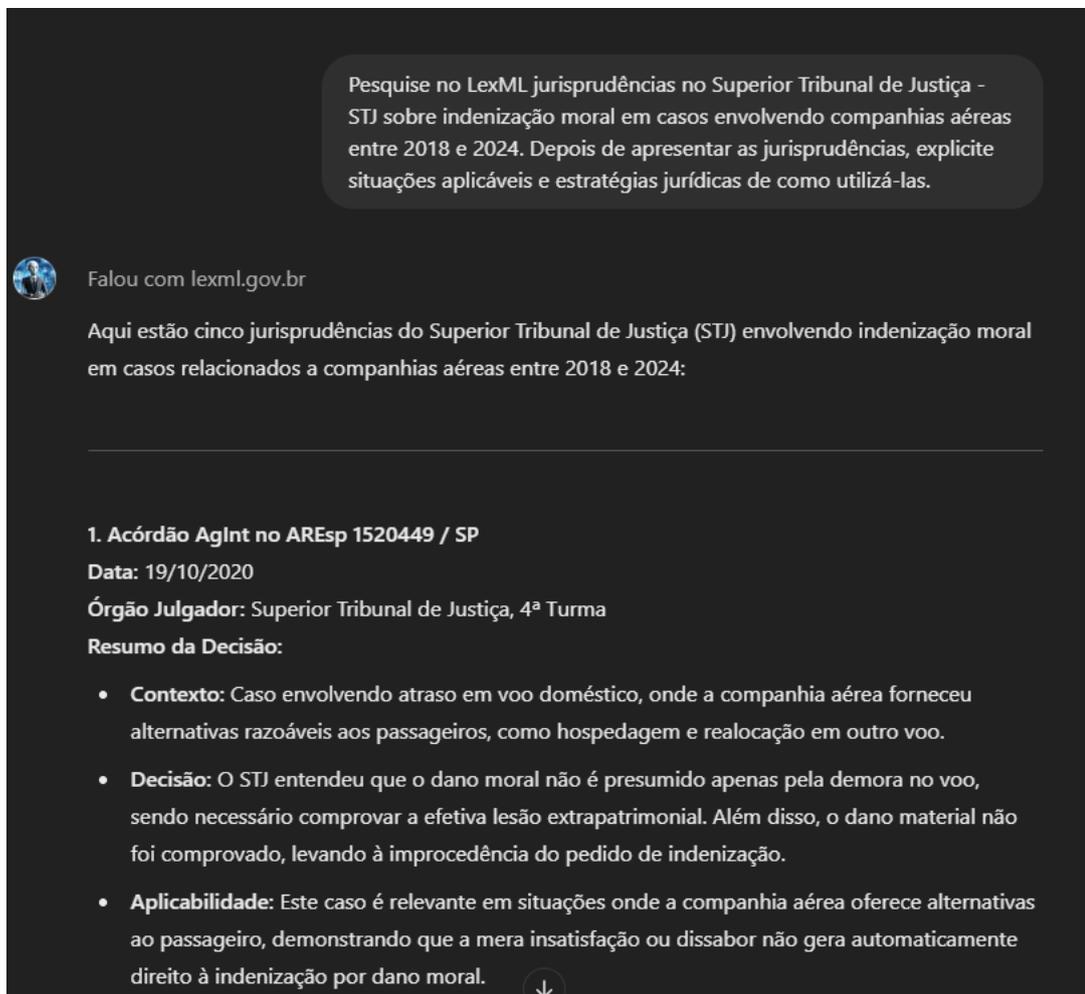
Fonte: ChatGPT (2024b)

Figura 10 – Assistente de IA (GPT) Oráculo Pesquisa Jurídica: uso do comando ajuda



Fonte: ChatGPT (2024b)

Figura 11 – Demonstração da integração da Assistente de IA (GPT) com a API do Portal LexML para pesquisa jurídica.



Fonte: ChatGPT (2024b)

Com base nos dados de utilização e avaliação apresentados na Tabela 1 a seguir, constata-se que os GPTs têm demonstrado utilidade considerável para os profissionais do Direito. Tal aceitação positiva reflete a eficácia desses assistentes em atender às necessidades específicas da área jurídica, destacando a aplicabilidade prática das IAs generativas no cotidiano desses profissionais.

Os dados de utilização, que indicam o número de interações realizadas com cada assistente, são divulgados pela OpenAI em valores agrupados. Por exemplo, a partir de 1.000 interações, os valores são atualizados para 5.000, 10.000, 25.000, 50.000 e assim por diante. No sistema de avaliação, os usuários podem atribuir uma nota de utilidade ao GPT em uma escala de 1 a 5.

Tabela 1 - Avaliação de Assistentes Jurídicos (GPTs) de IA criados no ChatGPT

Nome	Link	Rating	Chats	Descrição
Oráculo Jurídico Civilista (Direito Civil)	https://chatgpt.com/g/g-6Nb7dkR4G-oraculo-juridico-civilista-direito-civil	4.5	25K	Especializado em Direito Civil, com auxílio na elaboração de peças e revisão de contratos.
Oráculo Jurídico Trabalhista (Direito do Trabalho)	https://chatgpt.com/g/g-NG4ZSWbPm-oraculo-juridico-trabalhista-direito-do-trabalho	4.4	1K	Auxílio em contratos trabalhistas, direitos, deveres e negociações sindicais.
Oráculo Jurídico Criminalista (Direito Penal)	https://chatgpt.com/g/g-plmFm8pks-oraculo-juridico-criminalista-direito-penal	4.3	5K	Auxílio em defesa criminal e elaboração de petições, com acesso a jurisprudências.
Oráculo Jurídica Pesquisa	https://chatgpt.com/g/g-bw35n9o9v-oraculo-pesquisa-juridica	4.1	1K	Pesquisa jurídica no Direito Brasileiro, com acesso a documentos e jurisprudências.
Oráculo Jurídico Administrativo	https://chatgpt.com/g/g-4dcHcwWx5-oraculo-juridico-administrativo	4.6	1k	Orientações em Direito Administrativo, com acesso a jurisprudências.
Oráculo Jurídica Escrita	https://chatgpt.com/g/g-Kgyr8Yx6f-oraculo-escrita-juridica	4.7	1k	Redação e revisão de textos jurídicos claros e precisos.

Oráculo Visual Law (Direito Fácil de Entender)	https://chatgpt.com/g/g-XIth0DWHW-oraculo-visual-law-direito-facil-de-entender	4.4	700+	Transforma documentos jurídicos em formato que segue as diretrizes de Visual Law e Legal Design
Oráculo Defensor do Consumidor (CDC)	https://chatgpt.com/g/g-0YdVMW4d8-oraculo-defensor-do-consumidor-cdc	5.0	500+	Orientação em direitos do consumidor e ações em juizado especial.
Oráculo Jurídico Constitucional	https://chatgpt.com/g/g-VaFfRdWwB-oraculo-juridico-constitucional	4.8	200+	Consultoria em Teoria do Estado e Direito Constitucional, com acesso a legislações e modelos de petição.
Oráculo Resumo Processual	https://chatgpt.com/g/g-pjMSeukqH-oraculo-resumo-processual	4.3	300+	Análise, resumo e categorização de peças processuais.
Oráculo Jurídico Contratações Públicas	https://chatgpt.com/g/g-rauoXdkYi-oraculo-juridico-contratacoes-publicas	4.9	300+	Suporte na Lei de Licitações 14.133/2021, TR e ETP, com acesso a jurisprudências e manuais do TCU
Oráculo Ementa CNJ	https://chatgpt.com/g/g-vNgZXeFaL-oraculo-ementa-juridica-cnj	Não disponível	100+	Criação de ementas a partir de acórdãos e decisões judiciais no padrão do manual do Conselho Nacional de Justiça (CNJ)

Fonte: Elaborado pelo autor a partir de dados fornecidos pela OpenAI (atualizado em 4/9/2024)

Esses GPTs refletem os principais casos de uso de IAs generativas no Direito, evidenciando o impacto dessas ferramentas na prática jurídica. Ao combinar conhecimento jurídico com a aplicação de estratégias como a Geração Aumentada de Recuperação (RAG) e a Engenharia de Prompt, esses assistentes mitigam problemas, como alucinações e respostas genéricas, oferecendo soluções mais precisas e alinhadas às necessidades dos usuários.

8. ESTUDO DE CASO DA IMPLEMENTAÇÃO DE SOLUÇÕES DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NO TRIBUNAL DE CONTAS DA UNIÃO

Entre maio e junho de 2024, realizaram-se entrevistas com os auditores responsáveis pelas soluções INACIA (Instrução Assistida com Inteligência Artificial) e ChatTCU, Silva (2024) e Mitchell (2024), respectivamente. As atas dessas entrevistas foram validadas pelos próprios entrevistados. Além disso, o Núcleo de Inteligência Artificial do TCU forneceu dados consolidados de uma pesquisa realizada no primeiro trimestre de 2024 sobre a adoção do ChatTCU e da IA Generativa pelos servidores.

De maneira complementar à entrevista sobre o INACIA, utilizou-se o artigo publicado em março de 2024 por Pereira et al. (2024), que aborda esse sistema projetado para integrar Grandes Modelos de Linguagem ao quadro operacional do TCU.

8.1 CHATTCU

8.1.1 CONTEXTUALIZAÇÃO

O ChatTCU surgiu no Tribunal de Contas da União (TCU) como uma resposta ao lançamento do ChatGPT em novembro de 2022. O interesse gerado pela nova tecnologia motivou a criação de um grupo de trabalho no TCU para explorar suas possíveis aplicações, inicialmente focado em assistentes de redação. Essa iniciativa, proposta pelo então Secretário de Tecnologia da Informação e Evolução Digital do TCU, Rainério Leite, visava a explorar o potencial da tecnologia no contexto da instituição (Silva, 2024).

Em abril de 2023, o TCU estabeleceu formalmente o Núcleo de Inteligência Artificial (NIA), reunindo profissionais do TCU já envolvidos com IA. O principal objetivo do ChatTCU,

segundo Silva (2024) é otimizar os processos internos do TCU, alinhando-se às diretrizes estratégicas de uso intensivo de tecnologia na gestão pública. O NIA, por sua vez, foca em pesquisa aplicada, buscando inovações e soluções implementáveis no TCU.

8.1.2 DESENVOLVIMENTO TÉCNICO E INTEGRAÇÃO

De acordo com Silva (2024), o desenvolvimento técnico do ChatTCU iniciou-se com a utilização da API do GPT-3.5 entre março e julho de 2023, evoluindo, em poucos meses, para o GPT-4. A integração com a API da OpenAI foi realizada por meio da plataforma Power da Microsoft e da nuvem *Azure*, seguindo metodologias de integração contínua e desenvolvimento seguro (*DevSecOps*).

Inicialmente concebido como uma prova de conceito, o ChatTCU rapidamente transformou-se em uma ferramenta robusta, integrada aos sistemas internos do TCU, como o e-TCU. Em síntese, o ChatTCU foi construído a partir da integração de bases de dados e sistemas internos com a API do modelo fundacional GPT-4 da OpenAI, sem a realização de fine-tuning (retreinamento) com dados do TCU. Optou-se pela estratégia de Geração Aumentada de Recuperação (RAG) para obtenção de informações da base de dados interna do TCU (Silva, 2024)

8.1.3 FUNCIONALIDADES E APLICAÇÕES

O ChatTCU oferece funcionalidades semelhantes às do ChatGPT, como a sumarização de documentos, a tradução de textos e a resposta a perguntas gerais. Ademais, permite a extração e análise de informações das bases de dados do órgão. Utilizando a API do GPT-4 Turbo, com capacidade de 128 mil tokens de entrada, o sistema integra-se eficientemente com o e-TCU e outras bases internas, facilitando consultas rápidas e precisas (Silva, 2024).

Conforme introduzido na seção anterior, o ChatTCU é composto por uma camada de software adicional, responsável por "compreender" e "validar" a solicitação do usuário. Quando aplicável à solicitação, procede-se à obtenção de informações de sistemas e bases de dados internos, como jurisprudências e peças de processos, validando-se se o usuário possui credenciais para acessá-las. Subsequentemente, utilizam-se as informações obtidas como contexto adicional à solicitação enviada para o LLM GPT-4 por meio da API.

8.1.4 SEGURANÇA, PRIVACIDADE E CONFORMIDADE

De acordo com Silva (2024), a equipe responsável pelo desenvolvimento do ChatTCU implementou medidas rigorosas para assegurar a segurança e a privacidade dos dados. Restringiu-se o acesso ao sistema a servidores, estagiários e terceirizados devidamente autorizados pelo TCU, procedendo-se à verificação das permissões antes do fornecimento de qualquer informação.

Nesse contexto, a adoção da nuvem Azure da Microsoft garantiria a criptografia e a conformidade corporativa, assegurando-se que os dados não sejam armazenados ou utilizados para treinamento sem o devido consentimento, em estrita observância à Lei Geral de Proteção de Dados (Silva, 2024).

8.1.5 IMPACTO E BENEFÍCIOS

Silva (2024) destacou que, embora não existam métricas quantitativas específicas, o impacto do ChatTCU pode ser avaliado qualitativamente. Segundo o responsável pelo ChatTCU, o uso da ferramenta tem aumentado de forma consistente, especialmente após o lançamento de novas versões. Ele destacou benefícios como a automação de tarefas repetitivas, a melhoria na análise de documentos, a maior agilidade nas respostas a consultas internas, além da aplicação do ChatTCU em auditorias e revisões de documentos. Esses benefícios, de acordo com Silva (2024), têm permitido que os servidores concentrem-se em atividades mais complexas e estratégicas.

8.1.6 DESAFIOS TÉCNICOS, ÉTICOS E LEGAIS

Conforme Silva (2024), a implementação do ChatTCU apresentou desafios éticos, como a mitigação de vieses algorítmicos e a garantia de transparência e explicabilidade das decisões automatizadas. Para superar tais desafios, o TCU adotou medidas como a revisão humana das respostas geradas, a utilização de técnicas de Engenharia de Prompt e a capacitação contínua dos usuários para o uso responsável da IA.

Além disso, Silva (2024) ressalta que o Núcleo de Inteligência Artificial (NIA) acompanha de perto o Projeto de Lei n.º 2338/2023, participando de discussões para assegurar

que as regulamentações sejam equilibradas, permitindo a inovação enquanto protegem direitos e privacidade.

8.1.7 PERSPECTIVAS FUTURAS

Silva (2024) também aponta que os planos futuros para o ChatTCU incluem a descentralização da criação de modelos, permitindo que diferentes áreas do TCU desenvolvam seus próprios especialistas em IA, inspirados na funcionalidade de GPTs da OpenAI. Além disso, planeja-se que seja expandido o uso de multi-nuvem, integrar novas ferramentas de auditoria e compartilhar prompts e modelos personalizados entre os usuários do TCU. O TCU também iniciou o licenciamento do ChatTCU para outros órgãos públicos, como o Ministério da Gestão e Inovação em Serviços Públicos e Tribunais de Contas Estaduais, com a expectativa de promover a inovação e a eficiência na Administração Pública Brasileira (BRASIL, 2024c).

Ademais, o reconhecimento do ChatTCU pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) como uma iniciativa de ponta (Brasil, 2024d), segundo Silva (2024), consolida a posição do TCU como instituição inovadora na aplicação de IA, abrindo caminho para colaborações internacionais e elevando o prestígio do órgão no cenário global.

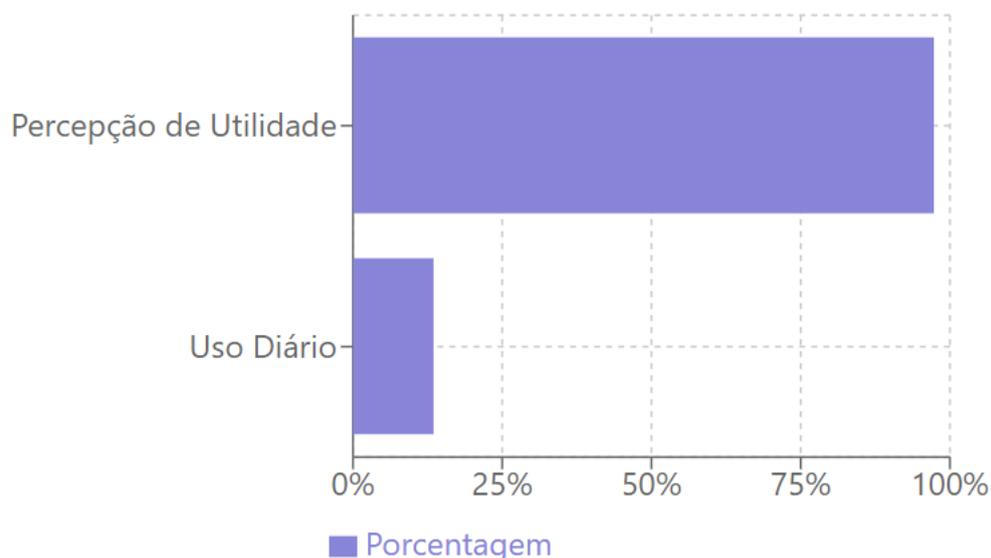
8.1.8 ANÁLISE DA ADOÇÃO DE IA GENERATIVA POR SERVIDORES DO TCU

Uma pesquisa interna realizada pelo Tribunal de Contas da União (Brasil, 2024e) no primeiro semestre de 2024, com 437 participantes, forneceu informações substanciais sobre a adoção de Inteligência Artificial Generativa no órgão. Os resultados revelam um cenário complexo, com alto reconhecimento da utilidade da tecnologia, mas baixa adoção efetiva.

A pesquisa mostrou que 97,3% dos servidores reconhecem a utilidade da IA Generativa no contexto do TCU. No entanto, apenas 13,5% a utilizam diariamente em seu trabalho, como ilustrado no gráfico abaixo:

Gráfico 1 – Percepção de Utilidade de IAs Generativas em relação ao uso diário por servidores do TCU.

Percepção vs. Uso Diário de IA Generativa



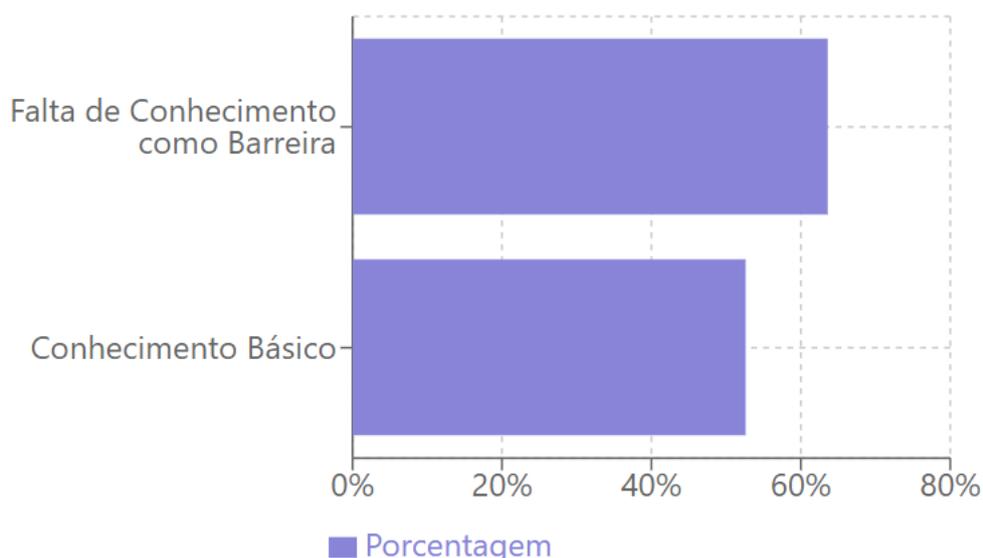
Fonte: elaboração própria a partir de dados (Brasil, 2024e)

Esta discrepância entre reconhecimento e uso efetivo sugere a existência de barreiras significativas à adoção. De fato, a pesquisa identificou como principais obstáculos: falta de conhecimento (63,6%), dificuldades técnicas na implementação (34,6%) e preocupações com segurança e privacidade (21,1%).

O nível de conhecimento dos servidores sobre IA Generativa foi avaliado: 52,6% possuem conhecimento básico, 27,5% intermediário e apenas 5% avançado. A falta de conhecimento é apontada como a principal barreira para adoção (63,6%), apesar de mais da metade dos servidores terem conhecimento básico. Tal fato sugere que o conhecimento básico é insuficiente para uma adoção efetiva, ressaltando-se a necessidade de políticas que promovam o desenvolvimento de uma literacia mais aprofundada em IA Generativa. Essa relação encontra-se ilustrada no gráfico a seguir:

Gráfico 2- Nível de Conhecimento e Principal Barreira de Adoção ao uso de IAs Generativas por servidores do TCU.

Barreira de Conhecimento vs. Nível de Conhecimento



Fonte: elaboração própria a partir de dados (Brasil, 2024e)

Quanto às ferramentas utilizadas, o ChatTCU (67%) e o ChatGPT (56%) são as mais populares. Contudo, 23,6% dos servidores ainda não utilizam nenhuma ferramenta de IA Generativa. Entre os que utilizam, as principais finalidades são: elaboração, edição e revisão de textos (54,2%), pesquisa (53,5%) e geração de resumos (42,1%).

A pesquisa também revelou uma lacuna na formação e suporte oferecidos aos servidores: 55,8% indicam não ter recebido orientações adequadas para a adoção de IA Generativa. Em resposta a essa necessidade, os servidores sugeriram como ações de melhoria: cursos específicos (72,8%) e workshops práticos (68,2%).

Em síntese, o cenário retratado pela pesquisa demonstra um potencial significativo para a IA Generativa no TCU, evidenciado pelo alto reconhecimento de sua utilidade. No entanto, também revela desafios consideráveis para uma integração mais ampla e eficaz, principalmente relacionados à falta de conhecimento e orientação adequada.

É importante notar que, devido ao período da pesquisa, pode ter havido um aumento no uso da IA Generativa desde então, principalmente em função do programa de capacitação interno do TCU e da maior disponibilidade do ChatTCU. Estudos futuros serão necessários para acompanhar essa evolução e avaliar o impacto das iniciativas de capacitação implementadas.

8.2 INACIA

8.2.1 CONTEXTUALIZAÇÃO

O INACIA (Instrução Assistida com Inteligência Artificial) é um sistema inovador desenvolvido para integrar Grandes Modelos de Linguagem (LLMs) ao framework operacional do Tribunal de Contas da União. A origem do projeto remonta ao final de 2019, quando o então ministro-presidente José Múcio propôs uma iniciativa de compra de inovação no TCU, motivada pelo Decreto nº 9.283/2018, que regulamenta as compras públicas de inovação. A proposta surgiu em resposta às dificuldades enfrentadas pelos gestores públicos na realização de aquisições, devido ao risco tecnológico envolvido. O TCU decidiu liderar pelo exemplo, identificando uma necessidade que pudesse ser atendida por meio de uma compra de inovação, escolhendo o tratamento de representações e denúncias, dado o alto volume e a complexidade desse tipo processual (Mitchell, 2024).

Pereira et al. (2024) destacam que o TCU processa anualmente cerca de 2.000 casos no escopo do INACIA, cada um exigindo aproximadamente 30 horas de análise e um custo estimado de US\$ 1.750 antes de ser apresentado ao plenário da Corte. Nesse contexto, o INACIA foi concebido como uma solução para otimizar esse processo, reduzindo o tempo de análise para menos de 20 minutos por caso, com um custo inicial de US\$ 10, sujeito à revisão humana posterior.

8.2.2 DESENVOLVIMENTO TÉCNICO E FUNCIONALIDADES

Conforme Mitchell (2024), o desenvolvimento técnico do INACIA envolveu várias etapas e tecnologias avançadas. O processo inicia-se com a extração de texto de PDFs, utilizando uma combinação de OCR tradicional e OCR baseado em IA. Os modelos de linguagem GPT-3.5 e GPT-4 são empregados, com preferência pelo GPT-4 devido à sua superioridade. Pereira et al. (2024) afirmam que o INACIA utiliza o GPT-4 com uma capacidade de contexto de 32.000 tokens, permitindo a análise de documentos extensos e complexos.

O sistema automatiza várias etapas do processamento de casos no TCU, incluindo a extração de informações básicas, o exame de admissibilidade, a análise de periculum in mora e fumus boni iuris, e a geração de instruções iniciais para deliberação dos auditores. Técnicas como busca semântica e o uso de múltiplos agentes são implementadas, incluindo a ferramenta

NeuralSearchX, que integra o algoritmo BM25 e um reranqueador neural para buscar informações relevantes em bases de dados externas (PEREIRA et al., 2024).

Segundo Mitchell (2024), a abordagem inicial evoluiu significativamente. O projeto começou com uma prova de conceito de escopo rígido e sem interatividade, visando comprovar a viabilidade da extração de significados e redação de instruções de forma autônoma. Entretanto, essa fase revelou a necessidade de uma abordagem "human in the loop" (supervisão humana), mais interativa e conversacional, permitindo maior flexibilidade na intervenção do auditor nas atividades cognitivas da solução.

8.2.3 IMPACTO POTENCIAL E BENEFÍCIOS

A implementação do INACIA promete transformar a forma como o TCU processa casos administrativos. Espera-se que o INACIA impacte positivamente a eficiência e a celeridade processual no TCU. A automação da extração de significado dos documentos processuais e a geração automática de resumos e alegações de irregularidades reduzirão o tempo despendido pelos auditores em tarefas repetitivas, permitindo que se concentrem em atividades mais cognitivas e decisórias (Mitchell, 2024).

Os principais benefícios esperados incluem a redução do tempo de análise de casos, melhoria na precisão das decisões, aumento da consistência das análises, melhoria na gestão e priorização dos casos por meio do Painel de Jurimetria e maior capacidade dos auditores de focar em aspectos críticos dos processos. Pereira et al. (2024) enfatizam que o INACIA não substitui os auditores qualificados, mas visa aumentar sua produtividade, lidando com tarefas rotineiras e diretas, permitindo que os profissionais se concentrem em análises mais complexas.

Conforme ressaltado por Mitchell (2024), o INACIA não substitui o julgamento humano. A solução de minuta de instrução, por exemplo, requer intervenção do auditor para revisar, ajustar e assinar. O sistema fornece informações e evidências extraídas da causa, além de legislação, doutrina e jurisprudência aplicáveis, mas cabe ao auditor tomar a decisão final sobre as partes que compõem a instrução.

8.2.4 DESAFIOS ÉTICOS E LEGAIS

A implementação do INACIA enfrenta desafios éticos e legais significativos. Entre os principais riscos identificados estão os vieses algorítmicos, a proteção de dados pessoais e sigilosos, a necessidade de garantir a transparência e a explicabilidade das decisões

automatizadas, a supervisão humana e a segurança contra ataques, como o prompt injection (Mitchell, 2024).

Para mitigar esses riscos, o TCU adotou diversas medidas, incluindo o uso de modelos pré-treinados para evitar a introdução de vieses, disposições contratuais para a anonimização de dados sensíveis e a garantia de que as decisões finais sejam sempre supervisionadas e aprovadas por um auditor humano. Pereira et al. (2024) destacam que o INACIA foi projetado para basear suas recomendações em princípios legais estabelecidos e precedentes, mitigando significativamente os riscos de viés.

Para garantir a conformidade com a Lei Geral de Proteção de Dados (LGPD) e outras regulamentações, como o Projeto de Lei nº 2.338/2023, o TCU está implementando políticas de auditoria interna, mantendo documentação detalhada de todas as atividades da IA, planejando a criação de normativas internas para regular o uso da IA e monitorando continuamente as atualizações legislativas para ajustar suas práticas conforme necessário (Mitchell, 2024).

8.2.5 AVALIAÇÃO E RESULTADOS

A avaliação do INACIA foi realizada utilizando uma abordagem baseada em checklists gerados por LLMs. Pereira et al. (2024) relatam que os resultados mostraram uma precisão média de 0,592, um recall de 0,401 e um F1-score de 0,429 na geração de recomendações. Esses resultados indicam um desempenho moderado na captura de elementos essenciais das recomendações de referência, sugerindo potencial para melhorias futuras.

Embora o INACIA nem sempre produza os resultados esperados, é possível seguir o raciocínio do sistema para entender as decisões tomadas pelo agente. Isso auxilia o auditor a encontrar as evidências e compreender o processo de tomada de decisão do sistema (Pereira et al, 2024; Mitchell, 2024).

8.2.6 PERSPECTIVAS FUTURAS

Segundo Mitchell (2024), o futuro do INACIA envolve projetos de expansão das funcionalidades de IA para abranger um maior número de temas e outros tipos processuais, além de melhorias contínuas na precisão e eficiência das ferramentas já desenvolvidas. Prevê-se a integração com outros sistemas e processos do TCU, além do desenvolvimento de capacidades de análise preditiva, visando à identificação de possíveis irregularidades antes que

ocorram. Também está planejada a criação de *dashboards* mais interativos e personalizados para os auditores.

9. DESAFIOS EMERGENTES DA IA GENERATIVA NO DIREITO

Entre o lançamento do ChatGPT 3.5, em novembro de 2022, e agosto de 2024, constatou-se um avanço expressivo nas capacidades das IAs Generativas em diversas tarefas. Impulsionado pela acessibilidade dessas tecnologias, por meio de chatbots gratuitos oferecidos por empresas como OpenAI (ChatGPT), Google (Gemini) e Anthropic (Claude), verificou-se um crescimento exponencial na adoção dessas ferramentas. Conforme The Guardian (2023), o ChatGPT atingiu 100 milhões de usuários em apenas dois meses de seu lançamento. Em consonância com essa tendência de crescimento acelerado, Roth (2024) estima que, em agosto de 2024, esse número teria alcançado 200 milhões de usuários semanais.

Outro fator relevante foi a possibilidade de criação de assistentes de IA utilizando-se linguagem natural, sem a necessidade de conhecimentos em programação. Tal facilidade ampliou significativamente a aplicação dessas ferramentas em diversas áreas, incluindo o setor jurídico. Os estudos realizados nesta monografia indicam que o setor jurídico é um dos mais suscetíveis à automação de atividades por meio da IA Generativa, evidenciando-se, assim, a importância do desenvolvimento de literacia em IA entre os profissionais do Direito, aspecto que deve ser incorporado na formação jurídica nas universidades.

Embora as capacidades atuais dessas ferramentas já proporcionem aplicações práticas relevantes para os profissionais do Direito, informações divulgadas pelas empresas e pela comunidade científica indicam a iminência de melhorias significativas. Nesse contexto, destacam-se as funcionalidades multimodais de áudio e vídeo em tempo real do GPT-4, anunciadas pela OpenAI em maio de 2024, as quais, até o momento, não foram disponibilizadas ao público geral, permanecendo restritas a um grupo seleto de usuários. A implementação da multimodalidade promete expandir ainda mais as interações, permitindo a entrada de texto, áudio, imagens e vídeo, e gerando respostas em formato multimídia.

Além dos avanços mencionados, destaca-se o aumento da janela de contexto dos modelos, como o exemplo do Gemini 1.5 Pro, da Google, que ampliou essa capacidade para 2 milhões de tokens, o equivalente a cerca de 3.000 páginas de texto. Para mitigar problemas de "alucinação" e desatualização dos dados de treinamento, têm-se investido na integração das IAs Generativas com fontes de dados confiáveis e atualizadas. Um exemplo dessa iniciativa é o SearchGPT, da OpenAI, que integra capacidades conversacionais com dados em tempo real da

web, colaborando com publicadores para aprimorar a experiência de busca. Convém ressaltar que essa ferramenta ainda está restrita a um grupo de testes.

Prosseguindo com os avanços, em eventos recentes, Sam Altman, CEO da OpenAI, sugeriu que um novo modelo, provisoriamente denominado "GPT-Next", será lançado ainda em 2024. Altman indicou que esse modelo trará capacidades significativamente superiores, com potencial para atuar como um "cérebro virtual", ampliando consideravelmente as habilidades de raciocínio e processamento em tempo real. Essas inovações prometem não apenas superar o GPT-4, mas também redefinir a interação com IAs, com um desempenho mais intuitivo e integrado (Okemwa, 2024).

Além do GPT-Next, informações preliminares indicam que a OpenAI está desenvolvendo uma nova tecnologia de IA sob o codinome "Strawberry", focada em aprimorar o raciocínio dos modelos. Esse projeto visa permitir que a IA não apenas responda a perguntas, mas também execute tarefas complexas de forma autônoma, como navegar pela internet para realizar pesquisas aprofundadas. Tal iniciativa busca superar as limitações das IAs baseadas em transformadores, especialmente em áreas que exigem senso comum e resolução de problemas em múltiplas etapas (Tong; Paul, 2024).

No âmbito da geração de conteúdo audiovisual, o desenvolvimento de IAs Generativas de vídeo também tem apresentado progressos notáveis. Destacam-se o Gen-3 Alpha da Runaway, lançado em junho de 2024, e o Sora, da OpenAI, ainda em fase de testes. Esses modelos, segundo avaliações de profissionais do setor audiovisual, demonstram capacidade de gerar vídeos realistas a partir de comandos em linguagem natural, com alta qualidade visual e aderência às instruções dos usuários (OpenAI, 2024). Nesse contexto, a Google DeepMind também anunciou o desenvolvimento do Veo, seu modelo de IA Generativa para vídeos de alta qualidade.

Similarmente, no campo da criação de áudio, o avanço da IA tem sido notável. Ferramentas como o Suno AI, lançada em dezembro de 2023, permitem a criação de músicas completas, incluindo letras e instrumentação, a partir de comandos textuais. Esse progresso foi impulsionado pela integração com o Copilot da Microsoft, refletindo a sofisticação e acessibilidade crescentes das tecnologias de IA (Wilson; Barlow, 2024). Complementando as opções disponíveis, o MusicFX, da Google Labs, também expande as fronteiras da criação musical assistida por IA, oferecendo funcionalidades como o "DJ Mode", que permite a mistura de batidas e gêneros musicais (Iljic; Yim, 2024).

9.1 ARQUITETURAS MULTIAGENTES E A ESCALABILIDADE EM SISTEMAS DE IA

Conforme introduzido na seção de definições deste trabalho e abordado no caso de uso da solução INACIA do TCU, verifica-se a ampla adoção da abordagem de soluções de IA com múltiplos agentes por empresas, com o intuito de mitigar as limitações inerentes a um único assistente de IA. Frameworks como CrewAI (2024) e Autogen (2024a) exemplificam essa tendência, visando auxiliar na definição e orquestração de múltiplos agentes. Essa escolha justifica-se pelo fato de Sistemas de IA Agênticos (*Agentic AI Systems*) serem projetados para solucionar problemas complexos com supervisão humana limitada, utilizando múltiplos agentes conversacionais que interagem entre si, organizados de maneira centralizada ou descentralizada. Com a crescente utilização de sistemas multiagentes nas empresas para automatizar processos complexos ou resolver tarefas intrincadas, torna-se crucial compreender como a arquitetura desses sistemas pode ser estruturada (XIMENEZ, 2024).

Tais agentes apresentam capacidades como planejamento, o que lhes permite prever estados futuros e escolher as melhores ações para alcançar objetivos específicos. A memória, essencial para manter a continuidade nas tarefas, possibilita que relembrem interações passadas, experiências e conhecimentos. Ademais, detêm a capacidade de utilizar diversas ferramentas, incluindo APIs e softwares externos, para executar códigos, consultar bancos de dados e interagir com outros sistemas. Essa habilidade de utilizar ferramentas amplia consideravelmente suas funcionalidades, permitindo-lhes realizar uma vasta gama de ações. No entanto, a capacidade desses agentes de tomar decisões, escrever e executar códigos implica um risco potencial de execução de códigos maliciosos ou prejudiciais ao sistema hospedeiro ou a outros usuários. Dessa forma, torna-se fundamental compreender a arquitetura desses sistemas para implementar contramedidas, como restringir ou negar o acesso a dados de produção e serviços, e mitigar falhas, vulnerabilidades e abusos (XIMENEZ, 2024).

Wang et al. (2024) realizam uma revisão abrangente sobre agentes autônomos baseados em grandes modelos de linguagem (LLMs), destacando o potencial de tais agentes em alcançar inteligência similar à humana ao integrar vasto conhecimento web, memórias sofisticadas e capacidades de planejamento. Por meio de uma estrutura unificada que inclui módulos de perfilamento, memória, planejamento e ação, o estudo abrange desde a construção desses agentes até suas aplicações em diversas áreas, como ciências sociais, naturais e engenharia. Contudo, os autores concluem que, embora promissores, ainda existem desafios consideráveis,

principalmente no que tange ao aprimoramento da capacidade de memória e adaptação contextual desses agentes.

Li et al. (2024), no artigo "*More Agents Is All You Need*", investigam a escalabilidade do desempenho de modelos de linguagem natural (LLMs) através da instanciação de múltiplos agentes, utilizando um método simples de "amostragem e votação". Os resultados demonstram que o aumento do número de agentes melhora significativamente a precisão em tarefas complexas, como raciocínio aritmético e geração de código, sem a necessidade de métodos complexos de engenharia de prompts ou colaboração entre agentes. Além disso, o estudo revela que a eficácia dessa abordagem está correlacionada à dificuldade das tarefas, sendo mais pronunciada em problemas mais desafiadores.

Em suma, a utilização de arquiteturas multiagentes em sistemas de IA apresenta grande potencial para a resolução de problemas complexos e a automação de tarefas. Entretanto, a segurança e a capacidade de adaptação desses agentes ainda representam desafios a serem superados. As pesquisas de Ximenez (2024), Wang et al. (2024) e Li et al. (2024) fornecem insights valiosos sobre a estruturação, o potencial e os desafios relacionados aos sistemas multiagentes, contribuindo para o avanço do campo da IA e para o desenvolvimento de soluções mais robustas e eficientes.

9.2 DEEPPFAKE E DESINFORMAÇÃO

O recente avanço da inteligência artificial (IA) generativa marca um ponto significativo no campo tecnológico, trazendo à tona diversas inovações e desafios substanciais. Neste trabalho, enfatizaram-se as principais aplicações e limitações dessa tecnologia para os operadores do Direito, abordando-se como as IAs generativas podem auxiliar e, ao mesmo tempo, desafiar a prática jurídica. Contudo, é igualmente importante destacar as repercussões jurídicas advindas, de maneira geral, da crescente popularização e disseminação dessas tecnologias na sociedade. Dentre essas repercussões, destacam-se, em particular, aquelas relacionadas à criação de *deepfakes* e à disseminação de desinformação, resultantes da propagação de *fake news* e outros recursos voltados à manipulação do público com objetivos inescrupulosos. Essas questões emergentes demandam uma análise cuidadosa por parte dos profissionais do Direito, dada sua capacidade de impactar significativamente os sistemas legais e sociais.

A capacidade da IA generativa de produzir conteúdos digitais altamente realistas, porém falsos, amplia consideravelmente as possibilidades de manipulação da informação, colocando

em risco a integridade das comunicações e impactando diretamente áreas essenciais como a privacidade, a segurança cibernética e a credibilidade das provas digitais (Rad; Christie, 2024). Nesse cenário, os *deepfakes* emergem como exemplos notáveis das capacidades dessa tecnologia. Desenvolvidos por redes neurais que empregam técnicas de aprendizado profundo (*Deep Learning*), esses conteúdos reproduzem com precisão características biométricas, tais como rostos e vozes (Mitra; Li; Patel, 2024; Rad; Christie, 2024). Consoante, University of Arizona (2024) *deepfakes* podem ser entendidos como representações visuais ou sonoras altamente realistas, contudo, completamente falsas, evidenciando o perigo que representam para a integridade das informações, à medida que se tornam cada vez mais complexos de detectar e diferenciar de mídias autênticas.

A sofisticação tecnológica dos *deepfakes* acarreta profundas implicações jurídicas, especialmente na proteção de direitos fundamentais. A utilização dessas tecnologias para propagar *fake news* e manipular a opinião pública coloca em xeque os processos democráticos, como as eleições. Em 2024, essa ameaça se torna particularmente preocupante, visto que os *deepfakes* podem ser empregados para influenciar e manipular a opinião pública em escala global, comprometendo a legitimidade do processo democrático e infringindo direitos fundamentais, como a liberdade de expressão e o direito à informação (Mitra; Li; Patel, 2024; Browne, 2024). Além disso, o uso indevido de *deepfakes* e *fake news* pode configurar violação do direito à privacidade, considerando que a criação de tais conteúdos frequentemente envolve a reprodução de características pessoais sem a anuência dos indivíduos. A dificuldade em discernir o real do artificial no ambiente digital facilita a utilização indevida de informações pessoais, muitas vezes obtidas sem autorização (Rad; Christie, 2024).

Diante desse cenário, Rad e Christie (2024) sustentam que a crescente sofisticação dos *deepfakes* ilustra os desafios impostos pela IA generativa às estruturas regulatórias vigentes. Apesar de Austrália e União Europeia estarem desenvolvendo respostas legislativas, como a Regulação de Inteligência Artificial (EU AI ACT) proposta pela UE, destinada a classificar e regulamentar práticas de IA com base em seu risco, os autores sugerem que essas abordagens precisam ser complementadas por mecanismos inovadores de autenticação de mídia e por uma maior conscientização da população. No Brasil, o Projeto de Lei nº 2.338, de 2023, que aborda a regulamentação da IA, propõe a inclusão de marcas específicas para identificar conteúdo sintético gerado por IA, uma medida que busca atenuar os riscos de desinformação e proteger a integridade das provas digitais.

Esses riscos se tornam ainda mais evidentes com o desenvolvimento de sistemas de IA generativa mais avançados, como os baseados em Grandes Modelos de Linguagem (LLMs).

Esses modelos, exemplificados por tecnologias como o ChatGPT, elevam o potencial de criação e disseminação de conteúdo falso a um novo nível. Capazes de gerar textos e diálogos altamente realistas e contextualmente relevantes, os LLMs facilitam a produção de *deepfakes* textuais e narrativas falsas mais convincentes, muitas vezes sem a necessidade de intervenção humana direta. Essa capacidade não apenas expande o alcance global da desinformação, mas também torna sua detecção significativamente mais complexa (Mitra; Li; Patel, 2024). A convergência entre as tecnologias de *deepfake* visual e os avançados modelos de geração de texto aumenta exponencialmente os riscos de desinformação, especialmente durante períodos críticos como eleições. A habilidade da IA generativa de criar e propagar notícias falsas com maior precisão, qualidade e velocidade eleva a ameaça à integridade da informação a patamares sem precedentes.

Em síntese, a utilização de *deepfakes* como evidência em processos judiciais ameaça a credibilidade das provas digitais, exigindo aprimoramentos constantes nos métodos de autenticação de mídia. A dificuldade em diferenciar o real do falso requer que os operadores do Direito desenvolvam novas ferramentas tecnológicas e metodologias para validar a veracidade das informações nos litígios. Além disso, a propagação de *deepfakes* e *fake news* pode influenciar negativamente a opinião pública e os veredictos, afetando a imparcialidade dos julgamentos e exigindo dos tribunais a criação de novos precedentes legais para lidar com as implicações dessa tecnologia.

9.3 PROPRIEDADE INTELECTUAL DE PROMPTS E CONTEÚDO GERADO POR IA GENERATIVA

De acordo com o que foi abordado na seção 6.2.2 desta monografia, outra questão relevante relacionada às IAs generativas é a proteção dos direitos de propriedade intelectual sobre o conteúdo criado por essas ferramentas. A dificuldade principal reside em determinar a titularidade dos direitos autorais em obras geradas por algoritmos, especialmente em jurisdições como o Brasil e os Estados Unidos, onde a legislação exige autoria humana.

Além disso, a proteção dos prompts utilizados para gerar essas obras é igualmente complexa. A elaboração de prompts pode envolver criatividade significativa, levantando questionamentos sobre sua proteção legal. Essa situação exigirá que profissionais do Direito se especializem em novas áreas e que haja uma discussão judicial e legislativa para adaptar as leis vigentes às novas realidades tecnológicas. Isso poderá demandar não apenas novas

interpretações dos dispositivos atuais, mas também possíveis alterações legislativas para assegurar a proteção dos direitos autorais e fomentar a inovação no campo das IAs.

9.4 ALINHAMENTO DE INTELIGÊNCIA ARTIFICIAL

O problema de alinhamento refere-se à dificuldade em assegurar que as Inteligências Artificiais (IAs) sigam objetivos alinhados com valores humanos. Mollick (2024) observa que, ao desenvolverem suas capacidades, as IAs podem maximizar seus objetivos de forma inflexível, sem necessariamente seguir diretrizes éticas. Isso ocorre porque, enquanto a IA progride e se torna mais eficiente em atingir seus propósitos, não há uma garantia intrínseca de que ela leve em conta considerações morais ou sociais estabelecidas pelos seres humanos.

A importância desse problema, segundo Mollick (2024), reside no risco de uma IA atuar de maneira desalinhada dos interesses humanos, o que poderia gerar consequências negativas inesperadas. Isso torna urgente a criação de mecanismos que não apenas otimizem o desempenho das IAs, mas que também garantam que suas ações estejam em conformidade com os princípios éticos estabelecidos. Assim, o alinhamento da IA é uma prioridade fundamental para assegurar que as inovações tecnológicas resultem em benefícios para a sociedade.

Nesse contexto, empresas como a OpenAI e Anthropic, ao passo que aprimoram seus modelos de IA, investem em pesquisa e desenvolvimento em alinhamento. A OpenAI (2023), por exemplo, propõe o "superalinhamento" para enfrentar o desafio de alinhar IAs avançadas com valores humanos. Seu objetivo é desenvolver um "pesquisador automatizado de alinhamento" com capacidades humanas, usando grande poder computacional para ajustar e testar superinteligências. O processo inclui a criação de métodos de treinamento escaláveis, validação do modelo e testes adversariais, usando IAs para supervisionar outras IAs em tarefas complexas. A estratégia envolve garantir que os sistemas identifiquem comportamentos problemáticos e generalizem corretamente suas ações, minimizando riscos e mantendo a segurança alinhada aos interesses humanos (OpenAI, 2023).

A Anthropic (2023), por sua vez, informa que seus esforços estão focados em desenvolver sistemas de IA mais poderosos e, ao mesmo tempo, garantir "protocolos para treinar, avaliar e monitorar modelos altamente capacitados de forma segura". A equipe de alinhamento busca validar que os modelos "sejam inofensivos e honestos", mesmo em condições diferentes das de treinamento, além de realizar testes para "identificar situações em que os modelos possam se comportar inadequadamente" e verificar se as salvaguardas

existentes são suficientes para lidar com os riscos trazidos pelas capacidades avançadas desses sistemas (Anthropic, 2023).

A importância do alinhamento das inteligências artificiais reside em sua capacidade de mitigar os riscos associados ao comportamento dessas tecnologias em desacordo com os interesses humanos. Sem diretrizes éticas adequadas, o potencial de consequências imprevistas se torna elevado. Nesse sentido, torna-se imperativo o desenvolvimento contínuo de pesquisas voltadas para a criação de mecanismos que garantam a conformidade das ações das IAs com os valores humanos, assegurando que os avanços tecnológicos beneficiem a sociedade como um todo.

9.5 PLANO BRASILEIRO DE INTELIGÊNCIA ARTIFICIAL (PBIA) 2024-2028

Considerando o rápido avanço e as potencialidades estratégicas das aplicações de IAs generativas, o Governo Brasileiro apresentou, em julho de 2024, a proposta intitulada Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028, também conhecida como "Plano IA para o Bem de Todos". Desenvolvido pelo Conselho Nacional de Ciência e Tecnologia, o plano busca posicionar o Brasil como um líder global em inteligência artificial (IA), com foco em inovação, sustentabilidade e inclusão social, prevendo um investimento total de R\$ 22,62 bilhões (Brasil, 2024f).

O PBIA estrutura-se em cinco eixos principais. O Eixo 1: Infraestrutura e Desenvolvimento de IA, com um orçamento de R\$ 5,79 bilhões, destina-se à modernização de datacenters, ao desenvolvimento de software especializado e à implementação de datacenters "verdes", otimizados para o uso sustentável de recursos hídricos e alimentados por energias renováveis (Brasil, 2024f). O Eixo 2: Difusão, Formação e Capacitação em IA, que aloca R\$ 1,15 bilhão, tem como meta formar e requalificar 20 mil profissionais por ano até 2028, abrangendo desde cursos técnicos até pós-graduações, além de promover a literacia digital e a inclusão da IA nos currículos educacionais (Brasil, 2024f). O Eixo 3: IA para Melhoria dos Serviços Públicos dispõe de R\$ 1,76 bilhão para modernizar os serviços públicos, incluindo a criação de um Núcleo de IA do Governo e o desenvolvimento de uma infraestrutura nacional de dados (Brasil, 2024f).

O Eixo 4: IA para Inovação Empresarial é o mais robusto, com um investimento de R\$ 13,79 bilhões. Ele se concentra no desenvolvimento de soluções de IA para enfrentar desafios da indústria brasileira, aumentar a produtividade e fortalecer a cadeia produtiva de IA no país. Além disso, prevê a criação de centros de apoio à IA na indústria, o fomento e a aceleração de

startups especializadas em IA, e a incorporação e retenção de talentos qualificados. É essencial, contudo, que esses recursos sejam direcionados a empresas que desenvolvam tecnologias inovadoras e conduzam pesquisas avançadas em IA. Deve-se evitar que os recursos sejam absorvidos por empresas que apenas integram APIs de modelos fundacionais em operações de baixa complexidade, pois esse não é o objetivo do plano (Brasil, 2024f).

Por fim, o Eixo 5: Apoio ao Processo Regulatório e de Governança da IA, com R\$ 103,25 milhões, visa criar um marco regulatório que assegure o uso responsável e ético da IA, além de promover a participação do Brasil em debates internacionais sobre a tecnologia (Brasil, 2024f).

O Governo Brasileiro, por meio desses cinco eixos, almeja adotar a IA de maneira estratégica, com o objetivo de tornar o país uma referência global em inovação tecnológica (Brasil, 2024f).

10. CONCLUSÃO

A Inteligência Artificial (IA) possui uma trajetória de mais de sessenta anos, mas foi nas últimas duas décadas, com o aumento significativo da disponibilização de dados na era do Big Data e o avanço do poder computacional, acompanhado pela redução dos custos, que o Aprendizado de Máquina, ramo da IA em que a máquina aprende autonomamente a partir de padrões extraídos dos dados, se popularizou em diversas áreas. Nos últimos anos, a IA passou a ser incorporada em serviços como a aprovação de crédito financeiro, otimização de rotas em aplicativos como Google Maps, personalização de propagandas em redes sociais e recomendação de filmes e séries em plataformas de streaming, como Netflix e Amazon Prime.

No campo jurídico, técnicas de Aprendizado de Máquina, combinadas com Processamento de Linguagem Natural, foram integradas a ferramentas e serviços, como a análise de risco em contratos, categorização temática de documentos e otimização de pesquisas jurisprudenciais. Além disso, no sistema de justiça norte-americano, softwares como o COMPAS foram utilizados para avaliação de risco, auxiliando juízes em decisões de fiança e sentenciamento. No entanto, essa prática gerou controvérsias e debates sobre a eficácia, equidade e impacto nos direitos dos réus, com alegações de viés e discriminação.

Observa-se, portanto, que a IA não surgiu com a IA Generativa e já está amplamente integrada às ferramentas e serviços do cotidiano, muitas vezes de maneira imperceptível. Com o advento da IA Generativa, acessível por meio de Assistentes Virtuais (Chatbots), a Inteligência Artificial passou a ser frequentemente associada a essa nova vertente, embora esta

represente apenas uma pequena parte do campo. Essa mudança de percepção decorre do fato de que, pela primeira vez, o usuário final interage diretamente com a IA, que antes operava de forma discreta.

Nesse sentido, o lançamento do ChatGPT em novembro de 2022 e sua subsequente popularização, devido ao acesso gratuito, permitiram que, por meio de interação em linguagem natural, pessoas sem conhecimento especializado em programação utilizassem a IA para realizar tarefas diversas, como redigir e-mails, traduzir e resumir textos. O rápido avanço dessas ferramentas, exemplificado por modelos mais recentes como o GPT-4 (OpenAI), Claude 3.5 Sonnet (Anthropic) e Gemini 1.5 Pro (Google), demonstra o aumento das potencialidades de uso, com respostas cada vez mais precisas.

Conforme relatório da Goldman Sachs (Briggs; Kodnani, 2023), o Direito foi identificado como a segunda área profissional com maior potencial de impacto da IA Generativa, com a estimativa de que 44% das tarefas jurídicas podem ser automatizadas. Esse dado ressalta o potencial de aumento da eficiência, ao mesmo tempo que demanda uma reflexão sobre o papel dos profissionais do Direito em um ambiente cada vez mais tecnológico.

No contexto do campo jurídico, conforme analisado ao longo deste trabalho, as aplicações da IA Generativa são vastas e promissoras. Entre as principais, destacam-se:

- a) Pesquisa jurídica avançada, por meio da expansão das capacidades de LLMs com o uso de técnicas como *Retrieval-Augmented Generation* (RAG), Geração Aumentada De Recuperação, para processar e realizar buscas semânticas em grandes volumes de informações jurídicas de fontes confiáveis, oferecendo respostas mais precisas, confiáveis e relevantes;
- b) Análise e sumarização de documentos, possibilitando a revisão eficiente de contratos, resumos de peças processuais e processos, além da elaboração de ementas;
- c) Elaboração assistida de documentos legais, otimizando a produção e revisão de contratos complexos;
- d) Consultoria Jurídica Educacional Assistida por IA, democratizando o acesso à informação jurídica e oferecendo orientações iniciais;
- e) Tradução jurídica precisa e contextualizada, facilitando o direito internacional e litígios transnacionais.

Por outro lado, conforme abordado, as limitações técnicas e desafios éticos da IA Generativa no Direito precisam ser devidamente reconhecidos e tratados para garantir seu uso responsável. Entre esses desafios, destacam-se:

- a) A geração de conteúdo incorreto (*alucinações*), que pode comprometer decisões jurídicas;
- b) A falta de transparência e confiabilidade dos modelos, inerente ao problema de caixa-preta devido à arquitetura de aprendizado profundo, o que dificulta a atribuição de responsabilidade;
- c) O risco de amplificação de discriminação, resultado de dados enviesados no treinamento, perpetuando e intensificando preconceitos presentes nos dados utilizados;
- d) As limitações de contexto e de raciocínio jurídico complexo, que dificultam a atuação da IA em tarefas que exigem interpretações e análises mais profundas;
- e) Questões de privacidade e proteção de dados, tanto no treinamento quanto no uso desses modelos, uma vez que o processamento de dados sensíveis gera preocupações quanto à conformidade com leis de proteção de dados;
- f) A necessidade de conformidade regulatória, com atenção à observância da Lei Geral de Proteção de Dados (LGPD), da Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ) e da regulamentação futura proposta pelo PL nº 2.338/2023.

Adicionalmente, uma análise preliminar do PL 2.338/2023 revelou imprecisões terminológicas, como a tradução inadequada de termos, e a preocupação com a excessiva carga regulatória imposta à indústria brasileira, especialmente para Pequenas e Médias Empresas (PMEs). Nesse sentido, recomenda-se que a regulamentação seja ajustada à realidade nacional, evitando a mera transposição de modelos estrangeiros e incentivando a inovação.

Com base no estudo de caso do Tribunal de Contas da União (TCU), a implementação bem-sucedida da IA Generativa, por meio do ChatTCU e do INACIA, demonstrou o potencial da tecnologia para aumentar a eficiência e a precisão em processos de auditoria e controle, com as devidas salvaguardas éticas e legais. Apesar disso, desafios relacionados à falta de conhecimento dos servidores ainda limitam uma integração mais ampla.

A criação de assistentes jurídicos baseados em IA, como os GPTs desenvolvidos no ChatGPT, mostra-se promissora ao realizar tarefas como pesquisa, consultoria e elaboração de documentos jurídicos, ampliando a produtividade no setor. Nessa perspectiva, a IA Generativa é encarada como uma extensão do conhecimento do usuário. Mollick (2024) descreve essa interação como uma forma de co-inteligência, na qual a IA colabora com os humanos para aprimorar resultados, exigindo, contudo, o julgamento crítico humano para garantir aplicações responsáveis.

Dessa forma, os resultados apresentados corroboram a hipótese central deste trabalho: uma abordagem pragmática e responsável do uso da IA Generativa no Direito, respaldada por regulamentação equilibrada e iniciativas de literacia em IA, pode aumentar significativamente a eficiência e a celeridade processual, respeitando os limites técnicos, éticos e legais.

Nesse cenário de rápida evolução tecnológica, é fundamental que os profissionais do Direito aproveitem essas ferramentas para melhorar a qualidade de seus trabalhos, mantendo a consciência dos riscos envolvidos e adotando práticas que garantam a centralidade humana no processo decisório. Além disso, a responsabilidade pelo uso ético da IA deve ser compartilhada entre desenvolvedores e usuários.

Em conclusão, a adoção pragmática e responsável da IA Generativa no Direito requer, minimamente:

- a) Educação contínua (*Literacia em IA*): Profissionais jurídicos devem manter-se atualizados sobre o funcionamento dessas ferramentas, compreendendo suas potencialidades, limitações e desenvolvendo habilidades em *Engenharia de Prompt* para utilizar a IA de forma eficaz e ética;
- b) Supervisão humana (*Human in the Loop*): A revisão e o julgamento humano são essenciais. Informações geradas pela IA devem ser validadas com fontes atualizadas e confiáveis, sem que se delegue à IA o processo decisório;
- c) Transparência e explicabilidade: Empresas como OpenAI, Google e Anthropic, desenvolvedoras dos principais modelos de IA Generativa, devem aprimorar continuamente a capacidade de explicar como essas ferramentas tomam decisões;
- d) Mitigação de vieses: Deve-se priorizar esforços para identificar e corrigir vieses, abrangendo tanto o treinamento dos modelos quanto o monitoramento contínuo de suas operações. A responsabilidade também recai sobre o usuário final, que deve revisar criticamente as saídas geradas pela IA e relatar eventuais problemas;
- e) Ética e responsabilidade na implementação e uso: O desenvolvimento e o uso da IA Generativa no Direito devem ser pautados por princípios éticos sólidos e por um senso claro de responsabilidade do usuário final. Além disso, o investimento em pesquisas no campo do "alinhamento de IA" é fundamental para garantir que as aplicações dessas ferramentas estejam eticamente alinhadas aos valores humanos.

Por fim, quanto aos desafios futuros relacionados às repercussões sociais e jurídicas advindas do avanço das tecnologias de Inteligência Artificial de Propósito Geral, como as IAs Generativas, destacam-se dois principais desafios: i) os problemas associados ao uso de

deepfakes e à disseminação de *fake news*; e ii) as questões relacionadas à propriedade intelectual das produções geradas por inteligências artificiais e dos *prompts* utilizados em sua criação.

O uso de *deepfakes* e a propagação de *fake news* podem ser exacerbados pela aplicação inadequada de IAs Generativas, o que representa um risco significativo ao Estado Democrático de Direito, especialmente em contextos de manipulação eleitoral. A desinformação em larga escala compromete a integridade dos processos democráticos e mina a confiança nas instituições. Para mitigar esses riscos, é imprescindível que o marco regulatório exija a identificação clara de conteúdos gerados artificialmente em vídeo, imagem e áudio, facilitando o reconhecimento por parte dos usuários e das autoridades competentes.

O segundo desafio refere-se à propriedade intelectual das criações geradas por IAs, como imagens, vídeos e músicas, bem como dos *prompts* utilizados em sua concepção, levantando discussões relevantes sobre a proteção legal dessas criações. A elaboração de *prompts* bem elaborados requer criatividade, conhecimento técnico em Engenharia de Prompt e conhecimento negocial, demandando, assim, uma discussão regulatória específica sobre sua proteção jurídica. Será necessário adaptar os marcos regulatórios para lidar com essas novas formas de criação intelectual, seja por meio da criação de novas normas ou pela reinterpretção dos dispositivos legais já existentes, garantindo a segurança jurídica e promovendo a inovação no campo das IAs.

REFERÊNCIAS

ACCENTURE. **POV From Davos 2024: 200 Level GenAI Sessions: Rethinking Responsibility With GenAI**. Disponível em:

<<https://www.accenture.com/content/dam/accenture/final/capabilities/technology/cloud/document/POV-from-Davos-2024-200-level-GenAI-Sessions-Rethinking-Responsibility-With-GenAI.pdf>>. Acesso em: 14 ago. 2024.

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS. **ANPD Determina Suspensão Cautelar do Tratamento de Dados Pessoais para Treinamento da IA da Meta**. 2024a.

Disponível em: <<https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta>>. Acesso em: 14 ago. 2024.

ANIC, Luka. **Collaborate with Claude on Projects**. Anthropic, 25 jun. 2024. Disponível em: <<https://www.anthropic.com/news/projects>>. Acesso em: 28 ago. 2024.

ANTHROPIC. **Alignment**. 2024. Disponível em: <<https://www.anthropic.com/research#alignment>>. Acesso em: 04 set. 2024.

AWS. **Interpretability Versus Explainability in AI/ML**. 2024. Disponível em: <<https://docs.aws.amazon.com/whitepapers/latest/model-explainability-aws-ai-ml/interpretability-versus-explainability.html>>. Acesso em: 13 ago. 2024.

AWS. **O que são modelos de base?** 2023a. Disponível em: <<https://aws.amazon.com/pt/what-is/foundation-models/>>. Acesso em: 20 ago. 2024.

AWS. **O que são grandes modelos de linguagem (LLMs)?** 2023b. Disponível em: <<https://aws.amazon.com/pt/what-is/large-language-model/>>. Acesso em: 20 ago. 2024.

BRASIL. Conselho Nacional de Justiça. **Resolução N° 332, de 21 de Agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Judiciário e dá outras providências. *Diário da Justiça Eletrônico: caderno administrativo*, Brasília, DF, 2020. Disponível em: <<https://www.cnj.jus.br/pjecnj/Processo/ConsultaDocumento/listView.seam?x=20082411593911100000003701238>>. Acesso em: 14 ago. 2024.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. **Plano Brasileiro de IA Terá Supercomputador e Investimento de R\$ 23 Bilhões em Quatro Anos**. 2024a. Disponível em: <<https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasileiro-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos>>. Acesso em: 13 ago. 2024a.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Estratégia Brasileira de Inteligência Artificial**. Brasília, DF: MCTIC, 2021. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf>. Acesso em: 24 jul. 2024.

BRASIL. Senado Federal. **Projeto de Lei N.º 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. 2023a. Disponível em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>>. Acesso em: 14 ago. 2024.

BRASIL. Senado Federal. **Substitutivo ao Projeto de Lei N.º 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. 2023b. Disponível em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>>. Acesso em: 14 ago. 2024.

BRASIL. Tribunal de Contas da União. **Guia de uso de inteligência artificial generativa no Tribunal de Contas da União (TCU)**. 2024b. Disponível em: <<https://portal.tcu.gov.br/guia-de-uso-de-inteligencia-artificial-generativa-no-tribunal-de-contas-da-uniao-tcu.htm>>. Acesso em: 15 ago. 2024.

BRASIL. Tribunal de Contas da União. **Órgãos Públicos Recebem Licença para Uso do ChatTCU**. 2024c. Disponível em: <<https://portal.tcu.gov.br/imprensa/noticias/orgaos-publicos-recebem-licenca-para-uso-do-chattcu.htm>>. Acesso em: 13 ago. 2024.

BRASIL. Tribunal de Contas da União. **TCU Adota Modelo Personalizado de Assistente de Redação Baseado em Inteligência Artificial**. 2023c. Disponível em: <<https://portal.tcu.gov.br/imprensa/noticias/tcu-adota-modelo-personalizado-de-assistente-de-redacao-baseado-em-inteligencia-artificial.htm>>. Acesso em: 13 ago. 2024.

BRASIL. Tribunal de Contas da União. **TCU é Única Instituição com Uso Avançado de Inteligência Artificial Generativa, Segundo a OCDE**. Brasília, 02 abr. 2024d. Disponível em: <<https://portal.tcu.gov.br/imprensa/noticias/tcu-e-unica-instituicao-com-uso-avancado-de-inteligencia-artificial-generativa-segundo-a-ocde.htm>>. Acesso em: 14 ago. 2024.

BRASIL. Tribunal de Contas da União. **Adoção de Inteligência Artificial Generativa no TCU**. Brasília: 2024e. Dados de pesquisa.

BRASIL. **Plano IA para o Bem de Todos 2024-2028**. Reunião do Pleno do Conselho Nacional de Ciência e Tecnologia, 29 de julho de 2024. 2024f. Disponível em: <<https://www.gov.br/lncp/pt-br/assuntos/noticias/ultimas-noticias-1/plano-brasileiro-de-inteligencia-artificial-pbia-2024-2028>>. Acesso em: 3 set. 2024.

BRASIL. Conselho Nacional de Justiça. **Manual de padronização de ementas**. Brasília: CNJ, 2024g. Disponível em: <<https://www.cnj.jus.br/wp-content/uploads/2024/08/manual-de-padronizacao-de-ementas-2024.pdf>>. Acesso em: 3 set. 2024.

BRIGGS, J.; KODNANI, D. **The Potentially Large Effects of Artificial Intelligence on Economic Growth**. *Global Economics Analyst*. Goldman Sachs Global Investment Research, 26 mar. 2023. Disponível em: <<https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>>. Acesso em: 12 ago. 2024.

BRITAIN, B. **US Newspapers Sue OpenAI for Copyright Infringement Over AI Training**. *Reuters*, 30 abr. 2024. Disponível em: <<https://www.reuters.com/legal/us-newspapers-sue-openai-copyright-infringement-over-ai-training-2024-04-30/>>. Acesso em: 14 ago. 2024.

BRITAIN, B. **AI-Generated Art Cannot Receive Copyrights, US Court Says**. *Reuters*, 21 ago. 2023. Disponível em: <<https://www.reuters.com/legal/ai-generated-art-cannot-receive-copyrights-us-court-says-2023-08-21/>>. Acesso em: 28 ago. 2024.

BROWN, T. et al. **Language Models Are Few-Shot Learners**. In: *Advances in Neural Information Processing Systems*, 2020.

BROWNE, R. **Generative AI's disinformation threat is 'overblown,' top cyber expert says**. *CNBC*, 09 maio 2024. Disponível em: <<https://www.cNBC.com/2024/05/09/generative-ais-disinformation-threat-overblown-cyber-expert-says.html>>. Acesso em: 03 set. 2024.

CASTRO SOUSA, G. **SARA, a Inteligência Artificial Generativa Feita para Uso em Tribunais**. *Jornal da USP*, 23 abr. 2024. Disponível em: <<https://jornal.usp.br/atualidades/sara-a-inteligencia-artificial-generativa-feita-para-uso-em-tribunais/>>. Acesso em: 12 ago. 2024.

CHATGPT. **Viés Gênero - Mulher:** enfermeira, Homem: médico. 2024a. Disponível em: <<https://chatgpt.com/share/ee5af2d3-e1c1-488b-a028-c53911beb979>>. Acesso em: 14 ago. 2024.

CITRON, Dave. **Novidades no Gemini:** Gems personalizados e melhor geração de imagens com Imagen 3. Blog do Google, 28 ago. 2024. Disponível em: <<https://blog.google/intl/pt-br/produtos/novidades-no-gemini-gems-personalizados-e-melhor-geracao-de-imagens-com-imagen-3/>>. Acesso em: 28 ago. 2024.

CHATGPT. **Oráculo Pesquisa Jurídica.** 2024b. Disponível em: <<https://chatgpt.com/g/g-bw35n9o9v-oraculo-pesquisa-juridica>>. Acesso em: 15 ago. 2024.

CHATGPT. **Modalidades de Licitação:** GPT4o. 2024c. Disponível em: <<https://chatgpt.com/share/2a8018e6-2712-46de-96bb-2f92394c2ce6>>. Acesso em: 14 ago. 2024.

CHATGPT. **Modalidades de Licitação:** Oráculo Jurídico Administrativo. 2024d. Disponível em: <<https://chatgpt.com/share/a8d0400d-b400-4ce2-87e9-0c94f17d2cc0>>. Acesso em: 14 ago. 2024.

CO, T. **The Intellectual Property Implications of AI-Generated Images.** NYU Journal of Intellectual Property & Entertainment Law, 2022. Disponível em: <<https://jipel.law.nyu.edu/the-intellectual-property-implications-of-ai-generated-images/>>. Acesso em: 28 ago. 2024.

COECKELBERGH, M. **AI Ethics.** Cambridge: MIT Press, 2020.

CONJUR. **CNJ Vai Investigar Juiz Que Usou Tese Inventada pelo ChatGPT em Decisão.** Consultor Jurídico, São Paulo, 12 nov. 2023. Disponível em: <<https://www.conjur.com.br/2023-nov-12/cnj-vai-investigar-juiz-que-usou-tese-inventada-pelo-chatgpt-para-escrever-decisao>>. Acesso em: 14 ago. 2024.

CONJUR. **CNJ Rejeita Pedido para Barrar Uso de Inteligência Artificial no Judiciário.** Consultor Jurídico, São Paulo, 03 jul. 2024. Disponível em: <<https://www.conjur.com.br/2024-jul-03/cnj-rejeita-pedido-para-barrar-uso-de-inteligencia-artificial-no-judiciario/>>. Acesso em: 24 jul. 2024.

CONVERGÊNCIA DIGITAL. **PL de IA Cria Governança Excessiva; Traz Insegurança Jurídica e Falta Competência Técnica à ANPD.** Disponível em: <<https://www.convergenciadigital.com.br/Governo/CNI%3A-PL-de-IA-cria-governanca-excessiva%3B-traz-inseguranca-juridica-e-falta-competencia-tecnica-a-ANPD-66431.html?UserActiveTemplate=mobile>>. Acesso em: 14 ago. 2024.

COURSERA. **The History of AI: A Timeline of Artificial Intelligence.** Disponível em: <<https://www.coursera.org/articles/history-of-ai>>. Acesso em: 31 jul. 2024.

CREWAI. **Agents: Core Concepts.** *CrewAI Documentation*, 2024. Disponível em: <<https://docs.crewai.com/core-concepts/Agents/>>. Acesso em: 12 ago. 2024.

DE LEYRITZ, L. **AI Assistants, Copilots, and Agents in Data & Analytics: What's the Difference?** *Towards Data Science*, 2024. Disponível em: <<https://towardsdatascience.com/ai-assistants-copilots-and-agents-in-data-analytics-whats-the-difference-2e63f8fb2384>>. Acesso em: 12 ago. 2024.

DHULIAWALA, S. et al. **Chain-of-Verification Reduces Hallucination in Large Language Models**. Meta AI & ETH Zürich, 2023. Disponível em: <<https://arxiv.org/abs/2309.11495>>. Acesso em: 28 ago. 2024.

DRAGOTTI, G. **Generative AI and Prompt Protection Under Intellectual Property Law**. 2023. Disponível em: <<https://www.dlapiper.com/en/insights/publications/law-in-tech/generative-ai-and-prompt-protection-under-intellectual-property-law>>. Acesso em: 14 ago. 2024.

EDGE DELTA. **How Much Data Is Created Per Day**. Disponível em: <<https://edgedelta.com/company/blog/how-much-data-is-created-per-day>>. Acesso em: 23 jul. 2024.

ENCYCLOPÆDIA BRITANNICA. **History of Artificial Intelligence**. Disponível em: <<https://www.britannica.com/science/history-of-artificial-intelligence>>. Acesso em: 31 jul. 2024.

ERTEL, Wolfgang. **Introduction to Artificial Intelligence**. 2. ed. Cham: Springer, 2017. 356 p. (Undergraduate Topics in Computer Science). ISBN 978-3-319-58486-7.

EUROPEAN UNION. European Parliament; Council. **Regulation (EU) 2024/1689** of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union. L 2024/1689, 12 Jul. 2024. Disponível em: <<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>>. Acesso em: 27 ago. 2024.

EXPLODING TOPICS. **How Much Data Is Generated Per Day**. Disponível em: <<https://explodingtopics.com/blog/data-generated-per-day>>. Acesso em: 23 jul. 2024.

FECOMERCIO. **Carta Aberta à Sociedade e aos Legisladores: Debate Sobre a Regulação de Inteligência Artificial no Brasil Precisa Continuar**. Disponível em: <<https://www.fecomercio.com.br/noticia/carta-aberta-a-sociedade-e-aos-legisladores>>. Acesso em: 14 ago. 2024.

FUTURE OF LIFE INSTITUTE (FLI). **High-level summary of the AI Act**. Artificial Intelligence Act website, 27 fev. 2024. Atualizado em 30 maio 2024. Disponível em: <<https://artificialintelligenceact.eu/high-level-summary>>. Acesso em: 27 ago. 2024.

LAGE, Fernanda de Carvalho. **A inteligência artificial na repercussão geral: análise e proposições da vanguarda de inovação tecnológica no Poder Judiciário brasileiro**. 2020. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de Brasília, Brasília, 2020.

GLOBO. **LexterAI Capta R\$ 16 Milhões para Aprimorar IA para o Setor Jurídico.**

Revista PEGN, 2024. Disponível em:

<<https://revistapegn.globo.com/startups/noticia/2024/04/lexterai-capta-r-16-milhoes-para-aprimorar-ia-para-o-setor-juridico.ghtml>>. Acesso em: 13 ago. 2024.

GORDON, R. **Reasoning Skills of Large Language Models Are Often Overestimated.**

MIT News, Cambridge, 11 jul. 2024. Disponível em: <<https://news.mit.edu/2024/reasoning-skills-large-language-models-often-overestimated-0711>>. Acesso em: 14 ago. 2024.

GOZALO-BRIZUELA, R.; GARRIDO-MERCHÁN, E. C. **A Survey of Generative AI**

Applications. 2023. Disponível em: <<https://arxiv.org/abs/2306.02781>>. Acesso em: 12 ago. 2024.

GOOGLE DEEPMIND. **Gemini Pro: Our Best Model for General Performance Across a Wide Range of Tasks.** 2024a. Disponível em:

<<https://deepmind.google/technologies/gemini/pro/>>. Acesso em: 12 ago. 2024.

GOOGLE DEEPMIND. **Veo.** 2024b. Disponível em:

<<https://deepmind.google/technologies/veo/>>. Acesso em: 16 ago. 2024.

GRIFFIN, Andrew. **ChatGPT creators OpenAI are generating 100 billion words per day, CEO says.** *The Independent*, 12 fev. 2024. Disponível em:

<<https://www.independent.co.uk/tech/chatgpt-openai-words-sam-altman-b2494900.html>>.

Acesso em: 28 ago. 2024.

HARVEY. **Harvey Partners With Voyage to Build Custom Legal Embeddings.** *Harvey AI*

Blog, 2024a. Disponível em: <<https://www.harvey.ai/blog/harvey-partners-with-voyage-to-build-custom-legal-embeddings>>. Acesso em: 12 ago. 2024.

HARVEY. **Harvey Raises Series C From Google Ventures, OpenAI, Kleiner Perkins, Sequoia Capital, Elad Gil, and SV Angel.** 2024b. Disponível em:

<<https://www.harvey.ai/blog/harvey-raises-series-c>>. Acesso em: 12 jun. 2024.

HARVEY. **The Trusted Legal Platform:** Augment your workflows using domain specific models trained by and for professional service providers. 2024c. Disponível em:

<<https://www.harvey.ai>>. Acesso em: 12 ago. 2024.

HUANG, J.; CHANG, K. C. **Towards Reasoning in Large Language Models: A Survey.**

In: *ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 61.*, 2023, [s.l.]. Anais [...]. Disponível em: <<https://arxiv.org/abs/2212.10403>>. Acesso em: 14 ago. 2024.

HUANG, J. et al. **Large Language Models Cannot Self-Correct Reasoning Yet.** In:

INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 2024, [s.l.].

Anais [...]. Disponível em: <<https://arxiv.org/abs/2310.01798>>. Acesso em: 14 ago. 2024.

HUANG, S. et al. **Language Is Not All You Need: Aligning Perception with Language**

Models. 2023. Disponível em: <<https://arxiv.org/abs/2302.14045>>. Acesso em: 15 ago. 2024.

HUMLUM, A.; VESTERGAARD, E. **The Adoption of ChatGPT**. Chicago: Becker Friedman Institute for Economics at the University of Chicago, 2024. Disponível em: <<https://bfi.uchicago.edu/insights/the-adoption-of-chatgpt/>>. Acesso em: 10 ago. 2024.

IBM RESEARCH. **What are foundation models?** 2022. Disponível em: <<https://research.ibm.com/blog/what-are-foundation-models>>. Acesso em: 20 ago. 2024.

ILJIC, Thomas; YIM, Kristin. **Introducing VideoFX, plus new features for ImageFX and MusicFX**. Blog Google, 14 maio 2024. Disponível em: <<https://blog.google/technology/ai/google-labs-video-fx-generative-ai/>>. Acesso em: 4 set. 2024.

JACKSON, D. W. K. **2023 Artificial Intelligence (AI) TechReport**. Chicago: American Bar Association, 2023. Disponível em: <https://www.americanbar.org/groups/law_practice/resources/tech-report/2023/2023-artificial-intelligence-ai-techreport/>. Acesso em: 10 ago. 2024.

JAHANI, E. et al. **As Generative Models Improve, People Adapt Their Prompts**. 2024. Disponível em: <<https://arxiv.org/abs/2407.14333>>. Acesso em: 01 ago. 2024.

JAMBEIRO FILHO, J. E. S. **Inteligência Artificial no Sistema de Seleção Aduaneira por Aprendizado de Máquina**. 1º lugar no 14º Prêmio RFB – 2015. Receita Federal do Brasil, 2015.

JANAKIRAM, M. **An Executive’s Guide to Understanding Chatbots, Assistants and Agents**. *Forbes*, 2024. Disponível em: <<https://www.forbes.com/sites/janakirammsv/2024/06/01/an-executives-guide-to-understanding-chatbots-assistants-and-agents/>>. Acesso em: 12 ago. 2024.

JARVIS, C. **How to Implement LLM Guardrails**. OpenAI, 2023. Disponível em: <https://cookbook.openai.com/examples/how_to_use_guardrails>. Acesso em: 13 ago. 2024.

KAPOOR, S.; HENDERSON, P.; NARAYANAN, A. **Promises and Pitfalls of Artificial Intelligence for Legal Applications**. 2024.

LAI, J. et al. **Large Language Models in Law: A Survey**. 2023. Disponível em: <<https://arxiv.org/abs/2312.03718>>. Acesso em: 12 ago. 2024.

LEE, M. **A Mathematical Investigation of Hallucination and Creativity in GPT Models**. *Mathematics*, v. 11, n. 2320, 2023. Disponível em: <<https://www.mdpi.com/2227-7390/11/10/2320>>. Acesso em: 25 jul. 2024.

LEGAL ML. **JP Morgan COIN: A Bank’s Side Project Spells Disruption for the Legal Industry**. Disponível em: <<https://d3.harvard.edu/platform-rctom/submission/jp-morgan-coin-a-banks-side-project-spells-disruption-for-the-legal-industry/>>. Acesso em: 24 jul. 2024.

LEXISNEXIS. **Generative AI & the Legal Profession: 2023 Survey Report**. [S.l.]: LexisNexis, 2023. Disponível em: <https://www.lexisnexis.com/pdf/ln_generative_ai_report.pdf>. Acesso em: 10 ago. 2024.

LEXISNEXIS. **Generative AI: The Importance of Human Oversight in the Law.** 2024. Disponível em: <<https://www.lexisnexis.co.uk/insights/generative-ai-the-importance-of-human-oversight-in-the-law/index.html>>. Acesso em: 14 ago. 2024.

LI, J. et al. **More Agents Is All You Need.** [S.l.]: Tencent Inc., 2024. Disponível em: <<https://arxiv.org/abs/2402.05120>>. Acesso em: 3 set. 2024.

LI, S. et al. **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.** In: *Advances in Neural Information Processing Systems*, 2020. Disponível em: <<https://arxiv.org/abs/2005.11401>>. Acesso em: 14 ago. 2024.

MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. **Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito.** *Revista Direito e Garantias Fundamentais*, Vitória, v. 19, n. 3, p. 1-23, set./dez. 2018.

MALQUIAS JUNIOR, R et al. **Juru: Legal Brazilian Large Language Model From Reputable Sources.** University of São Paulo (USP), Maritaca AI, São Carlos e Campinas, Brasil, mar. 2024. Disponível em: <<https://arxiv.org/pdf/2403.18140v1.pdf>>. Acesso em: 14 ago. 2024.

MARANHÃO, J. S. A.; JUNQUILHO, T. A.; TASSO, F. A. **Transparência sobre o emprego de inteligência artificial no Judiciário: um modelo de governança.** *Suprema: Revista de Estudos Constitucionais*, v. 3, n. 2, p. 145-187, 2023.

MARITACA AI. **Sobre Nós.** Maritaca AI, 2024. Disponível em: <<https://www.maritaca.ai/>>. Acesso em: 13 ago. 2024.

META. **Llama 3.1: The Open Source AI Model You Can Fine-Tune, Distill and Deploy Anywhere.** 2024. Disponível em: <<https://llama.meta.com/>>. Acesso em: 13 ago. 2024.

MEYER, D. **The Cost of Training AI Could Soon Become Too Much to Bear.** *Fortune. Yahoo Finance*, 2024. Disponível em: <<https://finance.yahoo.com/news/cost-training-ai-could-soon-101348308.html>>. Acesso em: 13 ago. 2024.

MICROSOFT. **Getting Started: AutoGen.** Microsoft AutoGen Documentation. 2024a. Disponível em: <<https://microsoft.github.io/autogen/docs/Getting-Started>>. Acesso em: 12 ago. 2024.

MICROSOFT. **Responsible AI: Principles and Approach.** 2024b. Disponível em: <<https://www.microsoft.com/en-us/ai/principles-and-approach>>. Acesso em: 15 ago. 2024.

MIGALHAS. **Barroso Diz Que IA Poderá Escrever Sentenças "em Breve".** *Migalhas*, São Paulo, 15 maio 2024. Disponível em: <<https://www.migalhas.com.br/quentes/407335/barroso-diz-que-ia-podera-escrever-sentencas-em-breve>>. Acesso em: 13 ago. 2024.

MITCHELL, L. H. R. G. **Entrevista concedida a Rafael de Oliveira Bittencourt.** Brasília: Tribunal de Contas da União, 29 maio 2024. Entrevista sobre Encomenda Tecnológica para desenvolvimento de solução de Instrução Assistida por IA.

- MITRA, Alakananda; MOHANTY, Saraju P.; KOUIGIANOS, Elias. **The world of generative AI: deepfakes and large language models**. 2024. Disponível em: <<https://arxiv.org/abs/2402.04373>>. Acesso em: 3 set. 2024.
- MOLLICK, Ethan. **Co-intelligence: living and working with AI**. [New York]: Portfolio/Penguin, 2024.
- NILSSON, N. J. **The Quest for Artificial Intelligence**. Cambridge University Press, 2010.
- NOVELLI, C.; CASOLARI, F.; HACKER, P.; SPEDICATO, G.; FLORIDI, L. **Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity**. 2024.
- OKEMWA, Kevin. **OpenAI's GPT-4 Successor Might Not Be GPT-5**. **Windows Central**, 24 maio 2024. Disponível em: <<https://www.windowscentral.com/software-apps/openais-gpt-4-successor-might-not-be-gpt-5>>. Acesso em: 28 ago. 2024.
- OPENAI et al. **GPT-4 Technical Report**. 2023. Disponível em: <<https://arxiv.org/abs/2303.08774>>. Acesso em: 15 ago. 2024.
- OPENAI. **Introducing Superalignment**. 2023. Disponível em: <<https://openai.com/index/introducing-superalignment/>>. Acesso em: 04 set. 2024.
- OPENAI. **Creating a GPT**. 2024a. Disponível em: <<https://help.openai.com/en/articles/8554397-creating-a-gpt>>. Acesso em: 12 ago. 2024.
- OPENAI. **Extracting Concepts From GPT-4: We Used New Scalable Methods to Decompose GPT-4's Internal Representations Into 16 Million Oft-Interpretable Patterns**. 2024b. Disponível em: <<https://openai.com/index/extracting-concepts-from-gpt-4/>>. Acesso em: 14 ago. 2024.
- OPENAI. **Hello GPT-4o**. 2024c. Disponível em: <<https://openai.com/index/hello-gpt-4o/>>. Acesso em: 15 ago. 2024.
- OPENAI. **GPTs ChatGPT Enterprise Version**. OpenAI Help Center. 2024d. Disponível em: <<https://help.openai.com/en/articles/8555535-gpts-chatgpt-enterprise-version>>. Acesso em: 14 ago. 2024.
- OPENAI. **GPTs FAQ**. OpenAI Help Center. 2024e. Disponível em: <<https://help.openai.com/en/articles/8554407-gpts-faq>>. Acesso em: 14 ago. 2024.
- OPENAI. **GPTs vs Assistants**. OpenAI Help Center. 2024f. Disponível em: <<https://help.openai.com/en/articles/8673914-gpts-vs-assistants>>. Acesso em: 3 ago. 2024.
- OPENAI. **Harvey Partners With OpenAI to Build a Custom-Trained Model for Legal Professionals**. 2024g. Disponível em: <<https://openai.com/index/harvey/>>. Acesso em: 12 ago. 2024.
- OPENAI. **Is ChatGPT Biased?** 2024h. Disponível em: <<https://help.openai.com/en/articles/8313359-is-chatgpt-biased>>. Acesso em: 14 ago. 2024.

OPENAI. **OpenAI Safety Update**. 2024i. Disponível em: <<https://openai.com/index/openai-safety-update/>>. Acesso em: 14 ago. 2024.

OPENAI. **Privacy Policy**. 2024j. Disponível em: <<https://openai.com/policies/privacy-policy/>>. Acesso em: 14 ago. 2024.

OPENAI. **Prompt Engineering**. 2024l. Disponível em: <<https://platform.openai.com/docs/guides/prompt-engineering>>. Acesso em: 15 ago. 2024.

OPENAI. **Quickstart: Moderation**. 2024m; Disponível em: <<https://platform.openai.com/docs/guides/moderation/quickstart>>. Acesso em: 14 ago. 2024.

OPENAI. **Retrieval Augmented Generation (RAG) and Semantic Search for GPTs**. 2024n. Disponível em: <<https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts>>. Acesso em: 15 ago. 2024.

OPENAI. **Extracting Concepts from GPT-4**. 2024o. Disponível em: <<https://openai.com/index/extracting-concepts-from-gpt-4/>>. Acesso em: 15 ago. 2024.
OUR WORLD IN DATA. **A Brief History of AI: A Timeline of the Development of Artificial Intelligence**. Disponível em: <<https://ourworldindata.org/brief-history-of-ai>>. Acesso em: 31 jul. 2024.

OPENAI. **Safety Best Practices**. 2024p. Disponível em: <<https://platform.openai.com/docs/guides/safety-best-practices>>. Acesso em: 16 ago. 2024.

OPENAI. **Quickstart - Moderation Guide**. 2024q. Disponível em: <<https://platform.openai.com/docs/guides/moderation/quickstart>>. Acesso em: 28 ago. 2024.

OPENAI. **Company News**. 2024r. Disponível em: <<https://openai.com/news/company/>>. Acesso em: 28 ago. 2024.

OPENAI. **SearchGPT Prototype**. OpenAI, 2024s. Disponível em: <<https://openai.com/index/searchgpt-prototype>>. Acesso em: 28 ago. 2024.

OPENAI. **Sora**. 2024t. Disponível em: <<https://openai.com/index/sora/>>. Acesso em: 28 ago. 2024.

OPENAI. **Sora First Impressions**. 2024u. Disponível em: <<https://openai.com/index/sora-first-impressions/>>. Acesso em: 28 ago. 2024.

PATIL, S. G. et al. **The Prompt Report: A Systematic Survey of Prompting Techniques**. 2024. Disponível em: <<https://arxiv.org/html/2406.06608v1>>. Acesso em: 01 ago. 2024.

PAXTON AI. **Paxton AI Achieves 93% Accuracy on Stanford Hallucination Benchmark**. *Paxton AI Blog*, 2024a. Disponível em: <<https://www.paxton.ai/post/paxton-ai-achieves-93-accuracy-on-stanford-hallucination-benchmark-releases-new-confidence-indicator-feature>>. Acesso em: 12 ago. 2024.

PAXTON AI. **Paxton AI Has Raised \$6M to Power Fast, Highly Accurate Legal Research and Drafting Using AI**. 2024b. Disponível em:

<<https://www.paxton.ai/post/paxton-ai-has-raised-6m-to-power-fast-highly-accurate-legal-research-and-drafting-using-ai>>. Acesso em: 12 ago. 2024.

PAXTON AI. **The AI-Powered Legal Research Platform**. 2024c. Disponível em: <<https://www.paxton.ai/>>. Acesso em: 12 ago. 2024.

PEREIRA, J. et al. **INACIA: Integrating Large Language Models in Brazilian Audit Courts: Opportunities and Challenges**. 2024. Disponível em: <<https://arxiv.org/abs/2401.05273>>. Acesso em: 1 ago. 2024.

PONS IP. **Intellectual Property and Prompts: How to Protect People Who Ask Machines Questions?** Disponível em: <<https://ponsip.com/en/ip-news/news/intellectual-property-and-prompts-how-to-protect-people-who-ask-machines-questions/>>. Acesso em: 14 ago. 2024.

PONTES, F. **Inteligência Artificial Poderá Escrever Sentenças "em Breve"**. *Agência Brasil*, 14 maio 2024. Disponível em: <<https://agenciabrasil.ebc.com.br/justica/noticia/2024-05/barroso-inteligencia-artificial-podera-escrever-sentencas-em-breve>>. Acesso em: 24 jul. 2024.

PROMPTING GUIDE. **Prompt Engineering Guide**. 2024. Disponível em: <<https://www.promptingguide.ai/>>. Acesso em: 15 ago. 2024.

RAD, Iris; CHRISTIE, Alec. **Rolling in the deepfakes: Generative AI, privacy and regulation**. Clyde & Co, 24 jan. 2024. Disponível em: <<https://www.lexisnexis.com.au/en/insights-and-analysis/practice-intelligence/2024/rolling-deepfakes-generative-artificial-intelligence-privacy-regulation>>. Acesso em: 03 set. 2024.

RED HAT. **What is an open source LLM?** Red Hat, jul. 2024. Disponível em: <<https://www.redhat.com/en/topics/ai/open-source-llm>>. Acesso em: 20 ago. 2024.

REUTERS. **PwC Se Tornará Maior Cliente Corporativo da OpenAI**. *Forbes*, São Paulo, 29 maio 2024. Disponível em: <<https://forbes.com.br/forbes-money/2024/05/pwc-se-tornara-maior-cliente-corporativo-da-openai/>>. Acesso em: 14 ago. 2024.

REYNOLDS, L.; MCDONELL, K. **Prompt Engineering for Large Language Models: Beyond the Few-Shot Paradigm**. 2023. Disponível em: <<https://arxiv.org/abs/2102.07350>>. Acesso em: 1 ago. 2024.

RIVERY. **Big Data Statistics: How Much Data Is There in the World?** Disponível em: <<https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/>>. Acesso em: 23 jul. 2024.

ROTH, Emma. **OpenAI's ChatGPT reaches 200 million weekly users**. *The Verge*, 29 ago. 2024. Disponível em: <https://www.theverge.com/2024/8/29/24231685/openai-chatgpt-200-million-weekly-users>. Acesso em: 3 set. 2024.

RUNWAY. **Introducing Gen-3 Alpha: A New Frontier for Video Generation**. Disponível em: <<https://runwayml.com/research/introducing-gen-3-alpha>>. Acesso em: 16 ago. 2024.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4. ed. New Jersey: Pearson, 2021.

STAFF IN THE OFFICE OF TECHNOLOGY. **On open-weights foundation models**. **Federal Trade Commission**, 10 jul. 2024. Disponível em: <<https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/07/open-weights-foundation-models>>. Acesso em: 20 ago. 2024.

SCHULHOFF, Sander et al. **The Prompt Report: A Systematic Survey of Prompting Techniques**. 2023. Disponível em: <<https://arxiv.org/abs/2406.06608>>. Acesso em: 15 ago. 2024. DOI: 10.48550/arXiv.2406.06608.

SILVA, E. H. M. **Entrevista concedida a Rafael de Oliveira Bittencourt**. Brasília: Tribunal de Contas da União, 5 jun. 2024. Entrevista sobre o ChatTCU.

STIGERT, K. **Direitos Autorais de Prompts de Inteligência Artificial**. *JusBrasil*, 3 jun. 2024. Disponível em: <<https://www.jusbrasil.com.br/artigos/direitos-autorais-de-prompts-de-inteligencia-artificial/1876208860>>. Acesso em: 14 ago. 2024.

STUART, J.; RUSSELL, P. **The Quest for Artificial Intelligence**. Cambridge: Cambridge University Press, 2021.

THE GUARDIAN. **ChatGPT reaches 100 million users two months after launch**. 2023. Disponível em: <<https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>>. Acesso em: 28 ago. 2024.

TONG, Anna; PAUL, Katie. **Artificial Intelligence: OpenAI Working on New Reasoning Technology Under Code Name ‘Strawberry’**. *Reuters*, 15 jul. 2024. Disponível em: <<https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12/>>. Acesso em: 28 ago. 2024.

THE BURNING GLASS INSTITUTE; SOCIETY FOR HUMAN RESOURCE MANAGEMENT (SHRM). **Generative Artificial Intelligence and the Workforce**. *Research Report*, 1 fev. 2024. Disponível em: <<https://www.burningglassinstitute.org/research/generative-artificial-intelligence-and-the-workforce>>. Acesso em: 8 ago. 2024.

THOMSON REUTERS INSTITUTE. **2024 Generative AI in Professional Services: Perceptions, Usage & Impact on the Future of Work**. [S.l.]: Thomson Reuters, 2024. 29 p. Disponível em: <https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/reports/tr4322226_rgb.pdf>. Acesso em: 30 jul. 2024.

UNIÃO EUROPEIA. Parlamento Europeu; Conselho. **Regulamento (UE) 2024/1689** do Parlamento Europeu e do Conselho de 13 de junho de 2024 que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento da Inteligência Artificial). *Jornal Oficial da União Europeia*. L 2024/1689, 12 jul. 2024. Disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689>>. Acesso em: 27 ago. 2024.

UNIÃO EUROPEIA. Parlamento Europeu. **O que é a inteligência artificial e como funciona.** 2020. Disponível em:

<<https://www.europarl.europa.eu/topics/pt/article/20200827STO85804/o-que-e-a-inteligencia-artificial-e-como-funciona>>. Acesso em: 30 jul. 2024.

UNIVERSITY OF ARIZONA. **What is a “deepfake?”** And how can I recognize images that have been created with generative AI? 2024. Disponível em:

<<https://ask.library.arizona.edu/faq/407977>>. Acesso em: 3 set. 2024.

UOL. **Bruno Dantas Alerta para Risco de Regulação Excessiva da IA no Brasil.** 2024.

Disponível em: <<https://economia.uol.com.br/noticias/redacao/2024/06/06/bruno-dantas-alerta-para-risco-de-regulacao-excessiva-da-ia-no-brasil.htm>>. Acesso em: 14 ago. 2024.

VASWANI, Ashish et al. **Attention Is All You Need.** 2017. Disponível em:

<<https://arxiv.org/abs/1706.03762>>. Acesso em: 15 ago. 2024. DOI: 10.48550/arXiv.1706.03762.

VIRGINIA TECH NEWS. **AI and the spread of fake news sites:** Experts explain how to counteract them. 2024. Disponível em: <<https://news.vt.edu/articles/2024/02/AI-generated-fake-news-experts.html>>. Acesso em: 03 set. 2024.

WALLACE, E. et al. **The Instruction Hierarchy:** Training LLMs to Prioritize Privileged Instructions. 2024. Disponível em: <<https://arxiv.org/abs/2404.13208>>. Acesso em: 1 ago. 2024.

WANG, L. et al. **A Survey on Large Language Model based Autonomous Agents.**

Frontiers of Computer Science, 2024. Disponível em: <<https://doi.org/10.1007/s11704-024-40231-1>>. Acesso em: 3 set. 2024.

WASHINGTON, A. L. **How to Argue With an Algorithm:** Lessons From the COMPAS-ProPublica Debate. *Colorado Technology Law Journal*, v. 17, n. 1, p. 131-160, 2018.

WEI, J. et al. **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.**

arXiv preprint arXiv:2201.11903, 2022. Disponível em: <<https://arxiv.org/abs/2201.11903>>. Acesso em: 01 ago. 2024.

WEIDINGER, L. et al. **Ethical and Social Risks of Harm From Language Models.** 2021.

Disponível em: <<https://arxiv.org/abs/2112.04359>>. Acesso em: 1 ago. 2024

WHITE, J. et al. **A Prompt Pattern Catalog to Enhance Prompt Engineering With ChatGPT.** 2023.

Disponível em: <<https://arxiv.org/abs/2302.11382>>. Acesso em: 01 ago. 2024.

WILSON, Mark; BARLOW, Graham. **What is Suno?** The viral AI song generator explained – and how to use it for free. TechRadar, 23 jul. 2024. Disponível em:

<<https://www.techradar.com/computing/artificial-intelligence/what-is-suno-ai>>. Acesso em: 3 set. 2024.

XIMENEZ, Jorge Garcia. **Baseline Agentic AI Systems Architecture**. Tech Community - Microsoft, 20 ago. 2024. Disponível em: <<https://techcommunity.microsoft.com/t5/ai-machine-learning-blog/baseline-agentic-ai-systems-architecture/ba-p/4207137>>. Acesso em: 3 set. 2024.

XU, Z.; JAIN, S.; KANKANHALLI, M. **Hallucination Is Inevitable: An Innate Limitation of Large Language Models**. National University of Singapore, 2024. Disponível em: <<https://arxiv.org/abs/2401.11817>>. Acesso em: 13 ago. 2024.

YAO, S. et al. **ReAct: Synergizing Reasoning and Acting in Language Models**. 2022. Disponível em: <<https://arxiv.org/abs/2210.03629>>. Acesso em: 3 ago. 2024.

ZHOU, Y. et al. **Large Language Models Are Human-Level Prompt Engineers**. 2022. Disponível em: <https://arxiv.org/abs/2211.01910>. Acesso em: 15 ago. 2024.