



Governança tecnológica e auditabilidade do alinhamento ético-valorativo (*alignment*) das inteligências artificiais generativas¹

Technological governance and auditability of the ethical-valorative alignment of generative artificial intelligences

Gobernanza tecnológica y auditabilidad del alineamiento ético-valorativo (alignment) de las inteligencias artificiales generativas

Fábio de Sousa Santos²

Pontifícia Universidade Católica do Paraná (Curitiba, PR, Brasil)

ORCID: <https://orcid.org/0000-0002-2590-5047>

E-mail: fabiosantos@pge.ro.gov.br

Pablo Ademir de Souza³

Universidade Federal do Paraná (Curitiba, PR, Brasil)

ORCID: <https://orcid.org/0000-0002-8082-8611>

E-mail: pablo.asouza@hotmail.com

Victor Guilherme Esteche Filho⁴

Universidade Federal do Paraná (Curitiba, PR, Brasil)

ORCID: <https://orcid.org/0000-0002-1678-5549>

E-mail: victorgesteche@gmail.com

¹SANTOS, Fábio de Sousa; SOUZA, Pablo Ademir de; ESTECHE FILHO, Victor Guilherme. Governança tecnológica e auditabilidade do alinhamento ético-valorativo (alignment) das inteligências artificiais generativas. **Suprema**: revista de estudos constitucionais, Brasília, v. 3, n. 2, p. 113-143, jul./dez. 2023. DOI: <https://doi.org/10.53798/suprema.2023.v3.n2.a225>.

²Procurador do Estado de Rondônia. Vice-Presidente do Instituto Rondoniense de Direito Administrativo. Doutor em Direito pela PUCPR. Currículo Lattes: <http://lattes.cnpq.br/7505335890817186>.

³Advogado. Mestre em Direito pela UFPR. Coordenador do ON – Observatório de Licitações e Contratos Administrativos da PUC-SP. Diretor-adjunto do Instituto de Direito Administrativo Sancionador Brasileiro (IDASAN). Currículo Lattes: <http://lattes.cnpq.br/3517412322369349>.

⁴Advogado. Membro do *Latin American Tax Policy Forum* (LATPF). Mestrando em Direito pela UFPR. Currículo Lattes: <http://lattes.cnpq.br/1370470272596953>.

Resumo

O presente artigo investiga o desafio do alinhamento ético-valorativo (*alignment*) em inteligências artificiais generativas, com ênfase no contexto dos tribunais brasileiros. A pesquisa procura elucidar a importância da adequação dessas tecnologias aos padrões éticos e morais da sociedade, especialmente quando aplicadas ao Poder Judiciário. A inter-relação entre o alinhamento da Inteligências Artificial (IA) e o conceito de governança tecnológica é examinada, levantando o seguinte problema de pesquisa: a auditabilidade do alinhamento ético-valorativo é uma condição *sine qua non* para sua integração aos tribunais brasileiros? Empregando uma metodologia lógico-dedutiva e apoiando-se em pesquisa bibliográfica direta, de caráter exploratório, o artigo aborda as nuances das IAs generativas, os dilemas inerentes à sua regulamentação e os fundamentos da governança tecnológica em IA. Conclui-se que a possibilidade de auditar o alinhamento ético-valorativo das IAs generativas é imprescindível para sua utilização no Poder Judiciário brasileiro, tornando-se um critério mandatório no plano de governança tecnológica de entidades que as adotam.

Palavras-chave

Governança tecnológica; alinhamento ético-valorativo; inteligência artificial generativa; auditabilidade; tribunais.

Sumário

1. Introdução. 2. Inteligência artificial generativa: conceito e breve panorama histórico. 3. Utilização de IA generativa nos tribunais brasileiros. 4. Conceito de alinhamento (“*alignment*”) e o problema dos vieses. 5. Governança tecnológica e a necessária auditabilidade do *alignment* das IAs. 6. Conclusão.

Abstract

This article investigates the challenge of ethical and value-based alignment in generative artificial intelligence, with a focus on the context of Brazilian courts. The research seeks to elucidate the importance of aligning these technologies with society’s ethical and moral standards, especially when applied to the Judiciary Branch. The interrelation between the alignment of AIs and the concept of technological governance is examined, raising the following research question: is the auditability of the ethical-value alignment a *sine qua non* condition for its integration into Brazilian courts? Employing a logical-deductive methodology and drawing upon direct bibliographical research of an exploratory nature, the article delves into the nuances of generative AIs, the dilemmas inherent in their regulation, and the fundamentals of AI technological governance. It is concluded

that the ability to audit the ethical-value alignment of generative AIs is essential for their use in the Brazilian Judiciary, becoming a mandatory criterion in the technological governance plan of entities that adopt them.

Keywords

Technology governance; Ethical-value alignment; Generative artificial intelligence; Auditability; Courts.

Contents

1. Introduction 2. Generative artificial intelligence: concept and brief historical overview 3. The use of generative AI in Brazilian courts. 4. The concept of alignment and the problem of bias. 5. Technological governance and the necessary auditability of AI alignment. 6. Conclusion.

Resumen

Este artículo investiga el desafío del alineamiento ético-valorativo en las inteligencias artificiales generativas, con énfasis en el contexto de los tribunales brasileños. La investigación busca dilucidar la importancia de la adecuación de estas tecnologías a las normas éticas y morales de la sociedad, especialmente cuando se aplican al Poder Judicial. Se examina la interrelación entre el alineamiento de la Inteligencia Artificial (IA) y el concepto de gobernanza tecnológica, planteando el siguiente problema de investigación: ¿la auditabilidad del alineamiento ético-valorativo es condición *sine qua non* para su integración en los tribunales brasileños? Empleando una metodología lógico-deductiva y basándose en una investigación bibliográfica directa de carácter exploratorio, el artículo aborda los matices de las IAs generativas, los dilemas inherentes a su regulación y los fundamentos de la gobernanza tecnológica en IA. Concluye que la posibilidad de auditar el alineamiento ético-valorativo de las IAs generativas es esencial para su uso en el Poder Judicial brasileño, convirtiéndose en un criterio obligatorio en el plan de gobernanza tecnológica de las organizaciones que las adopten.

Palabras clave

Gobernanza tecnológica; Alineación ética-valor; Inteligencia artificial generativa; Auditabilidad; Tribunales.

Índice

1. Introducción. 2. Inteligencia artificial generativa: concepto y breve reseña histórica. 3. Uso de la IA generativa en los tribunales brasileños. 4. El concepto

de alineación y el problema del sesgo. 5. La gobernanza tecnológica y la necesaria auditabilidad del alineamiento de la IA. 6. Conclusión.

1. Introdução

Em março de 2023, uma carta exigindo a pausa em experimentos com inteligência artificial, assinada pelos principais nomes do mundo do setor da tecnologia, como Elon Musk, levantou o debate sobre os riscos e a necessidade de regulamentação desta ferramenta.⁵

Estar-se-ia diante de um documento histórico, comparável aos avisos dos cientistas do século XX que protestavam contra o desenvolvimento de bombas nucleares, ou apenas de uma tentativa não confessada de concorrentes diminuir seu *gap* tecnológico em relação a quem hoje lidera o mercado?

No final de março, em 26.03.2023, o banco Goldman Sachs publicou projeções nas quais se estima que a IA generativa possa substituir até um quarto dos trabalhos atuais e expor até 300 milhões de empregos globais à automação, com projeção de aumento do PIB global em 7%.⁶ Já em julho de 2022, um engenheiro foi demitido da Google após alegar que o modelo de linguagem elaborado pela empresa, chamado LaMDA, havia adquirido *senciência*, ou seja, a capacidade de sentir sensações e sentimentos de forma consciente.⁷

Todos estes exemplos apontam para o fato de que o desenvolvimento da inteligência artificial, como se tem notícias, vem crescendo de forma exponencial na última década e, em 30 de novembro de 2022, ganhou os holofotes do mundo com o lançamento do ChatGPT, um modelo de linguagem de Inteligência Artificial (IA) criado e treinado pela empresa OpenAI.

O ChatGPT é um aplicativo de inteligência artificial que consegue gerar textos de alta qualidade em diversos estilos e formatos, incluindo artigos e

⁵ PAUSE giant AI experiments: an open letter. **Future of Life Institute**, Mar. 2023. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 4 abr. 2023.

⁶ GENERATIVE AI could raise global GDP by 7%. **Goldman Sachs**, 5 Apr. 2023. Disponível em: <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>. Acesso em: 6 abr. 2023.

⁷ LEMOINE, Blake. Is LaMDA sentient?: an interview. **Medium**, Jun. 2022. Disponível em: <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917>. Acesso em: 6 abr. 2023.

diálogos, além de ajudar a resolver problemas complexos de matemática ou programação. Desde o seu lançamento, ele conquistou mais de 100 milhões de usuários ativos em apenas dois meses – o que lhe rendeu o título de “*aplicativo de consumo com crescimento mais rápido da história*”, conforme constatado por estudo feito pela *Union de Banques Suisses* (UBS)⁸, até o lançamento da rede social Threads, que ultrapassou essa marca no dia 10 de julho de 2023, alcançando 100 milhões de usuários em apenas cinco dias.⁹ O ChatGPT, neste sentido, tem sido usado por empresas e organizações para melhorar seus negócios e serviços, bem como por indivíduos para aprimorar sua escrita, aprender novos idiomas e criar obras de arte.

Recentemente, a OpenAI lançou uma versão mais avançada do aplicativo, o GPT-4. Uma de suas principais capacidades é gerar respostas mais completas e coesas, bem como analisar e gerar imagens. Além disso, a probabilidade de produzir conteúdo ofensivo foi reduzida. O GPT-4 também tem um limite maior de caracteres, permitindo interações mais longas e análise de conteúdos mais extensos. Essas melhorias mostram o enorme potencial do GPT-4 para revolucionar diversos aspectos da vida moderna.

Durante o ano de 2023, outros modelos de linguagem concorrentes foram lançados, como Bard, da Google, Llama, da Meta, e Claude, da Anthropic. Estes modelos de linguagem, neste sentido, podem ser utilizados para ajudar a resolver problemas em diversas áreas, como saúde, finanças, educação e muitas outras. Essa ferramenta é uma das mais notáveis conquistas da inteligência artificial nos últimos anos, e tem um enorme potencial para revolucionar a maneira como a sociedade se estrutura.

Com a velocidade do desenvolvimento da IA, aumenta-se a expectativa da chegada de uma nova etapa de sua evolução, chamada de Inteligência Artificial Geral (ou *Artificial General Intelligence*, AGI) ou “superinteligência”, que, por definição seria uma categoria de inteligência artificial que se refere a sistemas capazes de realizar tarefas intelectuais de maneira geral, sem a necessidade de programação específica para cada tarefa. Em outras palavras, a AGI seria uma inteligência

⁸ GORDON, Cindy. ChatGPT is the fastest growing app in history of web applications. *Forbes*, Feb. 2023. Disponível em: <https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/?sh=6a692910678c>. Acesso em: 6 abr. 2023.

⁹ SRIRAM, Akash; KACHWALA, Zaheer. Rival do Twitter, Threads bate recorde e atinge 100 milhões de usuários em cinco dias. Edição de Sriraj Kalluvila, Rashmi Aich. *CNN Brasil*, 10 jul. 2023. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/rival-do-twitter-threads-bate-recorde-e-atinge-100-milhoes-de-usuarios-em-cinco-dias>. Acesso em: 10 ago. 2023.

artificial que possui habilidades cognitivas semelhantes ou superiores às dos seres humanos, como aprendizado, raciocínio, compreensão da linguagem natural e tomada de decisões em ambientes complexos.

Diferentemente das abordagens de (IA), que visam resolver problemas específicos e limitados, a AGI tem como objetivo desenvolver sistemas capazes de aprender e lidar com uma ampla gama de tarefas cognitivas, de forma semelhante a um ser humano. Ou seja, a AGI teria a capacidade de aprender de forma autônoma e se adaptar a novas situações, o que permitiria que ela realizasse uma variedade de tarefas diferentes sem a necessidade de programação específica para cada uma delas.

Neste contexto, Elon Musk e outras figuras públicas do setor de tecnologia assinaram uma carta aberta intitulada “*Pause Giant AI Experiments: an Open Letter*”, pedindo uma pausa no uso de experimentos de inteligência artificial. A carta destaca os potenciais riscos que os sistemas de IA com inteligência competitiva aos humanos podem representar à sociedade e à humanidade, citando princípios amplamente endossados, como os Princípios de IA de Asilomar.¹⁰

A carta pede que os laboratórios de IA interrompam o treinamento de sistemas mais poderosos que o GPT-4 por seis meses, permitindo que especialistas desenvolvam protocolos de segurança e governança robustos. A ideia não seria a interrupção do desenvolvimento de IAs em geral, mas a melhora na precisão, segurança, transparência e robustez dos sistemas atuais.

Além disso, os desenvolvedores de IA deveriam, segundo a carta, trabalhar com formuladores de políticas públicas para acelerar o desenvolvimento de sistemas de governança de IA robustos, incluindo autoridades reguladoras, sistemas de rastreamento e auditorias. A proposta é ter um “verão de IA”, em que colheríamos os benefícios da IA atual, melhorariamos os sistemas para benefício de todos e daríamos à sociedade tempo para se adaptar.¹¹

Em resposta a esta carta, em 29.03.2023, o CEO da OpenAI, Sam Altman, publicou um tweet no qual lista três medidas que deveríamos nos atentar para termos uma “boa” AGI no futuro: **i)** habilidade técnica para realizar o alinhamento ético-valorativo das superinteligências; **ii)** coordenação entre os principais *players*

¹⁰ AI principles. **Future of Life Institute**, Aug. 2017. Disponível em: <https://futureoflife.org/open-letter/ai-principles/>. Acesso em: 8 abr. 2023.

¹¹ PAUSE giant AI experiments: an open letter.

do mercado da AGI; e **iii**) uma “estrutura regulatória global eficaz”, o que incluiria uma governança democrática.¹²

No mesmo sentido, as pesquisadoras Clara Iglesias Keller, Laura Schertel Mendes, Mariana Valente e Yasmin Curzi, em texto intitulado “Oito medidas para regular big techs”, publicado na Folha de S. Paulo em março de 2023, apontam a necessidade de se democratizar a governança e o funcionamento destas tecnologias, com mapeamento público de riscos, auditorias internas e externas para verificar o cumprimento de dever de diligência.¹³

Desta forma, nos últimos anos, o conteúdo gerado por inteligência artificial tem ganhado destaque não apenas na comunidade de ciência da computação, mas também em toda a sociedade, com um crescente interesse pelos diversos produtos de geração de conteúdo desenvolvidos por grandes empresas de tecnologia, como ChatGPT, Midjourney e DALL-E 2. Esses aplicativos são caracterizados pela geração de conteúdo utilizando técnicas avançadas de inteligência artificial generativa, em contraste com a criação humana, permitindo a automação e produção de grandes volumes de conteúdo em curtos períodos.

O ChatGPT e seus concorrentes exemplificam modelos de linguagem utilizados na construção de sistemas de inteligência artificial conversacional, capazes de compreender e responder eficientemente às entradas linguísticas humanas de maneira significativa. Este modelo representa um avanço notável no campo, com aplicações potenciais em diversos setores.

Muitos acreditam que, diante dos avanços significativos no setor, estamos ingressando em uma nova era da inteligência artificial, com impactos substanciais em todo o mundo.¹⁴ Essa perspectiva ressalta a importância de se compreender e explorar as potencialidades e implicações das tecnologias de geração de conteúdo

¹² ALTMAN, Sam. **Things we need for a good AGI future**: 1) the technical ability to align a superintelligence 2) sufficient coordination among most of the leading AGI efforts 3) an effective global regulatory framework including democratic governance. San Francisco, 29 Mar. 2023. X Twitter: @sama. Disponível em: <https://twitter.com/sama/status/1641229941131051008>. Acesso em: 4 abr. 2023.

¹³ KELLER, Clara Iglesias *et al.* Oito medidas para regular big techs. **Folha de S. Paulo**. fev. 2023. Disponível em: <https://www1.folha.uol.com.br/ilustrissima/2023/02/oito-medidas-para-regular-big-techs-garantindo-liberdade-de-expressao.shtml>. Acesso em: 3 abr. 2023.

¹⁴ POR que o ChatGPT, o Midjourney e o Dall-E são uma nova era na inteligência artificial? **Pequenas Empresas & Grandes Negócios**, Rio de Janeiro, 31 mar. 2023. Disponível em: <https://revistapegn.globo.com/tecnologia/noticia/2023/03/por-que-o-chatgpt-o-midjourney-e-o-dall-e-sao-uma-nova-era-na-inteligencia-artificial.ghtml>. Acesso em: 23 ago. 2023.

baseadas em inteligência artificial, tanto para o desenvolvimento tecnológico quanto para a sociedade em geral.

Um dos principais desafios para o funcionamento destas tecnologias é o seu alinhamento ético-valorativo (*alignment*), que consiste basicamente na adequação e conformação dos possíveis resultados das inteligências artificiais generativas aos padrões éticos e morais da sociedade.

Neste contexto, o objetivo do presente trabalho é investigar o funcionamento do mecanismo de alinhamento ético-valorativo (*alignment*) das IAs e sua relação com o conceito de governança tecnológica, a fim de responder ao seguinte problema de pesquisa: a auditabilidade do alinhamento ético-valorativo (*alignment*) é condição para sua utilização por tribunais brasileiros?

O artigo utilizou o método lógico-dedutivo, de caráter exploratório, a partir da pesquisa bibliográfica direta para sua produção. No primeiro capítulo, será abordado o conceito de inteligência artificial generativa e sua aplicação pelos tribunais brasileiros. Na sequência, o conceito de *alignment* será analisado, assim como os dilemas de sua regulamentação. Por fim, os imperativos da governança tecnológica em IA, e a conclusão pela imprescindibilidade da auditabilidade do alinhamento das IAs generativas.

2. Inteligência artificial generativa: conceito e breve panorama histórico

O conceito de inteligência pode ser definido como a capacidade mental de raciocinar, resolver problemas e aprender novas habilidades. Ela se relaciona, por sua vez, com funções cognitivas relativas à percepção, memória e linguagem e pode ser classificada em: **i)** inteligência natural (biológica), baseada nos cinco sentidos controlados pela mente humana; ou **ii)** inteligência não natural (mecânica), que consiste em simulações artificiais dos sentidos humanos controlados por uma rede neural mecânica.¹⁵

¹⁵ GREWAL, Dalvinder Singh. A critical conceptual analysis of definitions of artificial intelligence as applicable to computer engineering. *IOSR Journal of Computer Engineering*, v. 16, n. 2, p. 9-13, Jan. 2014.

A inteligência artificial, neste sentido, pode ser conceituada como um “sistema de simulação mecânica de coleta de conhecimento, informações e inteligência de processamento do universo, compilando, interpretando e disseminando isso em forma de conhecimento prático”.¹⁶ Segundo a definição do PL nº 2338/2023 em seu art. 4º, I, um sistema de inteligência artificial é um:

... 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 real¹⁷

Já os modelos generativos são técnicas de inteligência artificial que geram resultados sintéticos realistas (texto, áudio, vídeo) a partir da apreensão de padrões de dados previamente treinados.¹⁸ Estes modelos, como sintetiza Matthew D. Byrne, geram resultados sintéticos basicamente tentando prever, a partir do treinamento de padrões, sequências lógicas de *pixels*, no caso de geração de imagem ou vídeos; palavras, no caso da geração de texto; ou notas musicais, em caso de músicas.¹⁹

A trajetória dos modelos generativos na inteligência artificial remonta à década de 1950, com a criação dos Modelos de Markov Ocultos (HMMs) e Modelos de Mistura Gaussiana (GMMs).²⁰ Tais modelos geravam dados sequenciais, como falas e séries temporais. Contudo, somente com o surgimento do aprendizado

¹⁶“Artificial Intelligence is the mechanical simulation system of collecting knowledge and information and processing intelligence of universe: (collating and interpreting) and disseminating it to the eligible in the form of actionable intelligence” (GREWAL, Dalvinder Singh. A critical conceptual analysis of definitions of artificial intelligence as applicable to computer engineering. p. 13)

¹⁷BRASIL. Congresso Nacional. Senado Federal. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, [2023]. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 10 ago. 2023.

¹⁸JOVANOVIĆ, Mladan; CAMPBELL, Mark. Generative artificial intelligence: trends and prospects. **Computer**, v. 55, n. 10, p. 107, Oct. 2022.

¹⁹BYRNE, Matthew D. Generative Artificial Intelligence and ChatGPT. **Journal of PeriAnesthesia Nursing**, v. 38, n. 3, p. 520, Apr. 2023.

²⁰YIHAN, Cao *et al.* A comprehensive survey of AI-Generated Content (AIGC): a history of generative AI from GAN to ChatGPT. **Association for Computing Machinery**, v. 37, n. 4, Article 111, Aug. 2018. Disponível em: <https://arxiv.org/pdf/2303.04226.pdf>. Acesso em: 22 ago. 2023.

profundo (“*deep learning*”), os modelos generativos alcançaram avanços expressivos em termos de desempenho.

Durante os estágios iniciais dos modelos generativos profundos, as diferentes áreas possuíam pouca interconexão. No âmbito do processamento de linguagem natural (NLP), o grande avanço para a elaboração de frases complexas adveio da introdução de Redes Neurais Recorrentes (RNNs). Posteriormente, foram desenvolvidas as Redes de Memória de Longo e Curto Prazo (LSTM) e Unidades Recorrentes com Portões (GRU), que empregavam mecanismos de controle de memória ao longo do treinamento. Essas técnicas conseguem processar cerca de 200 tokens por amostra, o que representa um avanço notável em comparação aos prévios modelos de linguagem.²¹

Paralelamente, no campo da visão computacional (“*computer vision*” ou CV), antes do advento de métodos baseados em aprendizado profundo, os algoritmos de geração de imagens convencionais recorriam a técnicas como síntese e mapeamento de texturas. Esses algoritmos dependiam de características projetadas manualmente e apresentavam limitações na capacidade de criar imagens diversificadas e complexas.²² Em 2014, surgiu a proposta das Redes Adversariais Generativas (“*generative adversarial network*” ou GANs), um marco importante nesta área, graças aos resultados notáveis obtidos em várias aplicações.²³ Estas redes adversariais, como ensina Mladan Jovanovic e Mark Campbell, basicamente operam com um par de redes neurais. Uma delas, conhecida como “geradora”, realiza a síntese do conteúdo (a partir da predição de sequências lógicas de *pixels*, por exemplo); e a segunda, chamada de “discriminadora”, avalia a autenticidade do conteúdo gerado, classificando-o como real ou sintético. A rede neural, então, repete este ciclo de geração e discriminação do conteúdo até que não seja mais possível distinguir o real do sintético.²⁴

²¹ KHANDELWAL, Urvashi *et al.* Sharp nearby, fuzzy far away: how neural language models use context. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 56th, 2018, Melbourne, Australia. **Proceedings** [...]. Melbourne, Australia: Association for Computational Linguistics, p. 284 –294, July 2018. (Long Papers, v.1). Disponível em: <https://aclanthology.org/P18-1027/>. Acesso em: 23 ago. 2023.

²² HECKBERT, Paul S. Survey of texture mapping. **IEEE Computer Graphics and Applications**. v. 6, n. 11, p. 56 –67, Nov. 1986.

²³ GOODFELLOW, Ian J. *et al.* Generative adversarial nets. In: GHARAMANI, Z. *et al.* (ed.). **Advances in neural information processing systems 27 (NIPS 2014)**. Montréal: Université de Montréal, Département d’informatique et de recherche opérationnelle, 2014. p. 1-8. Disponível em: https://papers.nips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html. Acesso em: 23 ago. 2023.

²⁴ JOVANOVIC, Mladan; CAMPBELL, Mark. Generative artificial intelligence: trends and prospects. p. 107.

Outros métodos, como *Autoencoders* Variacionais (VAEs) e modelos generativos de difusão, também foram desenvolvidos para proporcionar maior controle sobre o processo de geração de imagens e habilidade para gerar imagens de alta qualidade. O progresso dos modelos generativos em diversos domínios seguiu caminhos distintos, culminando no surgimento de uma interseção: a arquitetura *Transformer*.²⁵

A arquitetura *Transformer* foi posteriormente aplicada em CV, tornando-se a espinha dorsal dominante para diversos modelos generativos em várias áreas. No domínio do NLP, modelos de linguagem de grande porte, como BERT e GPT adotam a arquitetura *Transformer* como seu principal componente, proporcionando vantagens em relação aos blocos de construção anteriores, ou seja, LSTM e GRU. Na área de CV, o *Vision Transformer* (ViT)²⁶ expande ainda mais esse conceito, combinando a arquitetura *Transformer* com componentes visuais, permitindo sua aplicação em tarefas relacionadas a imagens. Além das melhorias que o *Transformer* trouxe às modalidades individuais, essa intersecção também possibilitou a fusão de modelos de diferentes domínios para tarefas multimodais.

Nos últimos anos, pesquisadores começaram a introduzir novas técnicas com base nesses modelos. Por exemplo, no campo do NLP, em vez de ajuste fino (“*fine tuning*”), às vezes as pessoas preferem a abordagem de “*few-shot prompting*”,²⁷ que consiste em incluir algumas amostras selecionadas do conjunto de dados no *prompt* para ajudar o modelo a compreender melhor os requisitos da tarefa. Já na linguagem visual, os pesquisadores frequentemente combinam modelos específicos de modalidade com objetivos de aprendizado contrastivo autossupervisionado para fornecer representações mais robustas. No futuro, à medida que o AIGC se tornar cada vez mais relevante, mais e mais tecnologias serão introduzidas, infundindo vitalidade nesta área.

Nesse sentido, devem ser destacados para o objetivo deste artigo, os grandes modelos de linguagem (ou LLMs) treinados para gerar texto legível com base em padrões e estruturas dos dados de entrada. Exemplos de LLMs incluem o GPT-3

²⁵ YIHAN, Cao *et al.* A comprehensive survey of AI-Generated Content (AIGC): a history of generative AI from GAN to ChatGPTJ.

²⁶ DOSOVITSKIY, Alexey *et al.* An image is worth 16x16 words: transformers for image recognition at scale. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 9, 2021. Annals [...]. ICLR, [2021]. Disponível em: <https://iclr.cc/virtual/2021/poster/3013>. Acesso em: 22 ago. 2023. p 1-9. 2021.

²⁷ QINGXIU, Dong *et al.* A survey on in-context learning. *arXiv*, 1 Jun. 2023. Disponível em: <https://arxiv.org/abs/2301.00234>. Acesso em: 22 ago. 2023.

e da OpenAI, que pode ser usado para tarefas como sistemas de diálogo, tradução e resposta a perguntas.

Também, pode ser destacado o modelo de linguagem generativo multimodal, que pode observar imagens e descrevê-las em texto, como o CLIP, desenvolvido pela OpenAI. O CLIP combina uma arquitetura de transformer com componentes visuais, permitindo que o modelo seja treinado em uma grande quantidade de dados de texto e imagem. Isso possibilita que o modelo aprenda a entender e descrever imagens em termos textuais, realizando tarefas como geração de legendas e identificação de objetos em imagens.

3. Utilização de IA generativa nos tribunais brasileiros

A Inteligência Artificial já é utilizada em alguns tribunais do Brasil, inclusive no Supremo Tribunal Federal (STF), que implementou o Victor, inteligência artificial que apoia a atividade de análise de admissibilidade recursal. O Robô Victor é treinado para a leitura e indicação da existência de algum tema de repercussão geral aplicável ao caso, com a validação posterior dos ministros. Para mais, o Victor realiza: a “conversão de imagens em textos no processo digital ou eletrônico; separação do começo e do fim de um documento (peça processual, decisão, etc); do fim de um documento (peça processual, decisão, etc)” e, conforme mencionado, “a identificação dos temas de repercussão geral de maior incidência”.²⁸

O Superior Tribunal de Justiça (STJ), por sua vez, criou o Sócrates 2.0, concebido como uma plataforma composta para várias funções, incluindo gerenciamento de normas, controvérsias, modelos, pesquisa automática de jurisprudência e doutrina, e gerenciamento de acervo por controvérsias.²⁹ O sistema trabalha exibindo uma “nuvem de palavras” das 30 palavras mais frequentes em uma petição de recurso especial, fornecendo uma ideia do conteúdo do recurso antes mesmo de sua

²⁸ PROJETO Victor avança em pesquisa e desenvolvimento para identificação dos temas de repercussão geral. **Supremo Tribunal Federal**, 19 ago. 2021. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=471331&ori#:-:text=O%20Victor%20C3%A9%20uma%20intelig%C3%Aancia,aplica%20ao%20caso%20dos%20autos>. Acesso em: 4 abr. 2023.

²⁹ REVOLUÇÃO tecnológica e desafios da pandemia marcaram gestão do ministro Noronha na presidência do STJ. **STJ**, 23 ago. 2020. Disponível em: <https://www.stj.jus.br/sites/porta/p/Paginas/Comunicacao/Noticias/23082020-Revolucao-tecnologica-e-desafios-da-pandemia-marcaram-gestao-do-ministro-Noronha-na-presidencia-do-STJ.aspx#:-:text=Not%C3%ADcias%20do%20STJ-,Revolu%C3%A7%C3%A3o%20tecnol%C3%B3gica%20e%20desafios%20da%20pandemia%20marcaram%20gest%C3%A3o%20do%20ministro,STJ%20no%20pr%C3%B3ximo%20dia%202027>. Acesso em: 23 ago. 2023.

leitura completa. A IA também identifica recursos potencialmente inadmissíveis, indicando a ausência de requisitos na interposição do recurso especial, reduzindo o tempo gasto na atividade de admissibilidade recursal.³⁰ Além disso, o Sócrates 2.0 elabora automaticamente documentos com informações relevantes sobre o recurso e cria modelos de decisão. A ferramenta tem como objetivo auxiliar na elaboração de minutas de decisão ou votos, e sua eficácia aumenta à medida que os servidores a utilizam. No futuro, a IA será capaz de identificar e classificar recursos especiais repetitivos e suas decisões, baseando-se em dados quantitativos e qualitativos do acervo de decisões do STJ. Essa inovação tem potencial para transformar o sistema jurídico e melhorar a eficiência na resolução de controvérsias legais.

No âmbito dos tribunais de justiça dos estados, há o Tribunal de Justiça de Minas Gerais, que desenvolveu o Radar, uma IA que identifica e separa recursos com pedidos idênticos, permitindo a elaboração de votos-padrão e julgamentos em conjunto de casos similares.³¹

No Tribunal de Justiça do Rio Grande do Norte, três robôs chamados Poti, Jerimum e Clara auxiliam na celeridade dos processos. Poti realiza buscas e bloqueios de valores em contas bancárias em apenas 35 segundos, enquanto Jerimum classifica e rotula processos e Clara lê documentos e recomenda decisões semelhantes.³²

Já o Tribunal de Justiça de Rondônia criou o Sinapse, que indica as etapas seguintes do processo e auxilia na elaboração de sentenças. Em Pernambuco, o Elis analisa certidões de dívida ativa, verificando dados, prescrição e competência. Em 2019, o Conselho Nacional de Justiça implantou um laboratório de inovação e um centro de IA para atender ao Judiciário.³³

³⁰ CHAVES, Guilherme; VEIGA, Elizabeth. A inteligência artificial na formação dos precedentes do STJ: Sistema Sócrates 2.0. *Migalhas*, 31 maio 2021. Disponível em: <https://www.migalhas.com.br/depeso/346278/a-inteligencia-artificial-na-formacao-dos-precedentes-do-stj>. Acesso em: 4 abr. 2023.

³¹ MELO, Jeferson. Judiciário ganha agilidade com uso de inteligência artificial. *Conselho Nacional de Justiça*, 3 abr. 2019. Disponível em: <https://www.cnj.jus.br/judiciario-ganha-agilidade-com-uso-de-inteligencia-artificial/>. Acesso em: 4 abr. 2023.

³² MELO, Jeferson. Judiciário ganha agilidade com uso de inteligência artificial.

³³ BRAGANÇA, Fernanda; BRAGANÇA, Laurinda Fátima da F.P.G. Revolução 4.0 no Poder Judiciário: levantamento do uso de inteligência artificial nos tribunais brasileiros. *Revista da Seção Judiciária do Rio de Janeiro*, Rio de Janeiro, v. 23, n. 46, p. 65-76, jul./out. 2019.

Vale destacar que o Conselho Nacional de Justiça (CNJ), em 2019, implantou um laboratório de inovação e um centro de IA para atender ao Poder Judiciário, visando criar ferramentas, pesquisas e experiências que auxiliem magistrados a decidir melhor, tornar mais céleres os fluxos processuais e encerrar processos mais rapidamente. A preocupação do CNJ é fazer com que os sistemas dos diversos tribunais do país “conversem” entre si, possibilitando a produção de métricas, diagnósticos e definição de políticas públicas.³⁴

Ainda, desde março de 2019, o Conselho da Justiça Federal (CJF) investe no programa FLUI JF, voltado à desburocratização e simplificação, incentivando práticas e casos de sucesso que facilitem procedimentos judiciais, extrajudiciais e de administração judicial. Em junho do mesmo ano, o CJF anunciou a criação da plataforma de IA Lia, destinada a responder dúvidas dos usuários por meio de um portal, adaptando-se às suas necessidades.

Em junho de 2022, o Centro de Inovação, Administração e Pesquisa do Judiciário da FGV Conhecimento publicou um relatório da 2ª fase de pesquisa sobre o uso de Inteligência Artificial (IA) nos tribunais brasileiros, intitulado “Tecnologias Aplicadas à Gestão de Conflitos no Poder Judiciário com ênfase no uso da Inteligência Artificial”. A pesquisa, coordenada pelo Ministro Luis Felipe Salomão, analisou quantitativa e qualitativamente o uso de IA no Judiciário e identificou que 44 tribunais, além do Conselho Nacional de Justiça, utilizam essa tecnologia.³⁵

Assim, a IA é considerada importante no contexto brasileiro, especialmente para racionalizar recursos e lidar com a judicialização expressiva, refletida em mais de 75 milhões de processos em tramitação, conforme o Relatório Justiça em Números 2021.³⁶

Conforme observado, a inteligência artificial, particularmente a generativa, tem demonstrado um enorme potencial na promoção do devido processo legal e no acesso à justiça no Brasil em razão da celeridade promovida em atenção à qualidade

³⁴ MELO, Jeferson. Judiciário ganha agilidade com uso de inteligência artificial.

³⁵ ESTUDO revela que 44 tribunais, além do Conselho Nacional de Justiça, usam inteligência artificial. FGV, 20 jul. 2022. Disponível em: <https://portal.fgv.br/noticias/estudo-revela-44-tribunais-alem-conselho-nacional-justica-usam-inteligencia-artificial>. Acesso em: 4 abr. 2023.

³⁶ SALOMÃO, Luis Felipe (coord.). **Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito ao Poder Judiciário brasileiro**. 2. ed. Rio de Janeiro: FGV Conhecimento, 2021. Disponível em: https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf. Acesso em: 4 abr. 2023.

dos atos processuais e da possibilidade de interconexão entre os sistemas, o que pode proporcionar economia processual.

Nesse sentido, a aplicação de IA em tribunais, como o STF com o Victor e o STJ com o Sócrates 2.0, tem melhorado significativamente a eficiência na análise de admissibilidade recursal, gestão de normas e pesquisa automática de jurisprudência. Além disso, a IA tem auxiliado na elaboração de minutas de decisão ou votos, otimizando o tempo gasto pelos servidores e magistrados.

Nos tribunais estaduais, a IA tem sido igualmente crucial na agilização dos processos. Exemplos como o Radar em Minas Gerais, Poti, Jerimum e Clara no Rio Grande do Norte, e Sinapse em Rondônia demonstram a versatilidade e eficiência da tecnologia na identificação, classificação e resolução de casos similares. O Elis em Pernambuco evidencia ainda mais a capacidade da IA em analisar e validar dados em certidões de dívida ativa.³⁷

A criação de laboratórios de inovação e centros de IA pelo CNJ e o programa FLUI JF são iniciativas que visam integrar os sistemas dos diversos tribunais do país, facilitando a produção de métricas, diagnósticos e a definição de políticas públicas. Essas ações têm como objetivo promover melhores decisões, conceder maior celeridade à marcha processual a fim de encerrar processos mais rapidamente em conformidade com o devido processo legal.

A crescente adoção dessa tecnologia demonstra um compromisso com a inovação e a busca por soluções eficientes para enfrentar os desafios do sistema judicial, todavia é necessário se atentar para potenciais riscos na adoção das novas tecnologias, a fim de que representem apenas ganhos para o sistema judiciário e, de modo geral, para a sociedade brasileira.

4. Conceito de alinhamento (“*alignment*”) e o problema dos vieses

A IA é baseada em modelos que são representações abstratas e simplificadas de processos do mundo real. Para desenvolver esses modelos, os programadores

³⁷ MELO, Jeferson. Judiciário ganha agilidade com uso de inteligência artificial.

selecionam informações que serão fornecidas ao sistema de IA, o que pode levar a pontos cegos nos algoritmos, refletindo os objetivos e prioridades do criador. Esses pontos cegos, também conhecidos como “*blindspots*”, podem ser irrelevantes para os resultados pretendidos ou, em alguns casos, podem ignorar informações importantes que afetam negativamente as respostas do sistema.³⁸

O “problema do alinhamento” refere-se à busca por garantir que os modelos de IA capturem nossos valores e normas, entendendo nossas intenções e objetivos. Resolver esse desalinhamento é crucial para evitar resultados catastróficos, já que a utilização contínua de sistemas de IA sem revisão pode reforçar e potencializar problemas preexistentes, como vieses e tratamentos discriminatórios.³⁹

Assim, a qualidade dos dados fornecidos aos sistemas de IA é crucial para garantir resultados confiáveis, visto que dados coletados de uma sociedade permeada por desigualdades e discriminações podem levar a padrões discriminatórios nos algoritmos e, assim, ao desalinhamento. Estudos realizados por pesquisadores da Universidade de Oxford mostram que o aprendizado da máquina pode confirmar e reproduzir esses padrões discriminatórios presentes nos dados. Deste modo, a introdução de um sistema de IA desalinhado em alguma função na sociedade pode significar o agravamento de problemas sociais decorrentes de condutas discriminatórias.⁴⁰

Nesse sentido, os vieses cognitivos, fenômenos estudados pelos psicólogos cognitivos e comportamentais, representam desvios cognitivos decorrentes de simplificações realizadas pela mente humana diante de questões complexas e podem ser responsáveis por agravar o problema do alinhamento, visto que podem ser transferidos para os algoritmos da IA.

Tais vieses algorítmicos podem ser observados em diversos sistemas de IA, como o COMPAS, utilizado nos Estados Unidos para avaliar o risco de

³⁸ NUNES, Dierle José Coelho; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. **Revista de Processo**, São Paulo, v. 43, n. 285, p. 421-447, nov. 2018.

³⁹ ROQUETE, Felipe Leitão Valadares. Inteligência artificial, valores e incerteza: um problema insolúvel para a regulação? **Internet&sociedade**, São Paulo, v. 2, n. 2, p. 150-154, 2021. Disponível em: <https://revista.internetlab.org.br/wp-content/uploads/2022/03/Inteligencia-artificial-valores-e-incerteza-um-problema-insolovel-para-a-regulacao.pdf>. Acesso em: 23 ago. 2023.

⁴⁰ NUNES, Dierle José Coelho; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. p. 435.

reincidência de acusados, e sistemas de concessão de crédito na Europa e América do Norte. Ambos os exemplos demonstram como os vieses presentes nos dados e algoritmos podem levar a resultados discriminatórios e injustos.⁴¹

Portanto, é importante destacar que a qualidade dos dados inseridos nos sistemas de IA é mais relevante do que a quantidade de informações. Dados enviesados podem ensinar a máquina a desempenhar suas funções de forma enviesada, perpetuando desigualdades sociais e erros.

A discussão sobre vieses tem sido frequente no campo do Direito, especialmente em relação às decisões judiciais. Os vieses cognitivos são inerentes aos seres humanos, uma vez que nosso cérebro possui recursos cognitivos limitados e cria atalhos para tomadas de decisões mais eficientes. Entretanto, quando afetam as decisões dos juízes, esses vieses podem ser extremamente prejudiciais, pois fatores que não deveriam influenciar o julgamento são levados em consideração, mesmo que inconscientemente.

O viés de confirmação é um dos principais vieses cognitivos que impactam o Direito e se caracteriza pela tendência de buscar e interpretar informações que confirmam as concepções já formadas pelo indivíduo. Isso afeta o processo de tomada de decisão, pois o julgador favorecerá as evidências que confirmam sua hipótese, descartando as que apontam para soluções diferentes.

Além do viés de confirmação, existem outros vieses que influenciam nosso pensamento e decisões, como o viés social, efeito de ancoragem e viés de conformidade. Embora existam diversos estudos sobre os vieses cognitivos, ainda há muitas dificuldades em lidar com o tema, pois muitos julgadores se consideram imparciais e não desenvolvem técnicas para superar o enviesamento.

O mesmo fenômeno pode ser observado nas ferramentas de IA, que muitos consideram imparciais. Porém, há um agravante: as decisões tomadas por humanos podem ser contestadas, enquanto os algoritmos utilizados nas ferramentas de IA são obscuros para a maioria da população e, às vezes, até para seus programadores. Isso torna as decisões das IAs, de certa forma, incontestáveis, o que é especialmente problemático no campo do Direito.

⁴¹ NUNES, Dierle José Coelho; MARQUES, Ana Luiza Pinto Coelho Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. p. 440.

É crucial que a implementação de sistemas de IA no âmbito do Poder Judiciário seja cuidadosa e permeada por um amplo debate sobre as melhores formas de realizar o aprendizado de máquina, reduzindo ao máximo o envio indesejado das máquinas a fim de otimizar o alinhamento. Também é necessário implementar mecanismos para garantir a transparência algorítmica, essencial para compreender o processo de tomada de decisões dos sistemas de IA.

Desta forma, é crucial que sejam instaurados procedimentos a fim da redução de vieses indesejados e a promoção do alinhamento com os princípios constitucionais. O supervisionamento da produção realizada pela IA generativa é vital para esse objetivo, bem como sistemas que viabilizem a realização de auditoria dos dados utilizados para o treinamento da IA.

5. Governança tecnológica e a necessária auditabilidade do *alignment* das IAs

O conceito de governança diz respeito aos processos, sistemas e controles pelos quais organizações públicas e privadas operam. Segundo Anthony Tarantino, a origem latina da palavra governança remete a “direção” e a governança geralmente envolve o exercício de autoridade legal e regulatória, além do uso de recursos institucionais para gerir organizações privadas.⁴²

No contexto destas organizações privadas, a governança é adjetivada como “corporativa”, e diz respeito ao processo por meio do qual as empresas regulam as relações entre as partes relacionadas (*stakeholders*) de sua operação, o que inclui as relações entre diretores, acionistas, gestores, fornecedores, clientes, órgãos reguladores e a comunidade na qual a empresa se insere. No contexto político, a governança pode significar a relação entre governo e seus cidadãos, tanto no nível legislativo, quanto no Executivo ou Judiciário. Em ambos os casos, a governança envolve três áreas de tomada de decisão: **i)** quem está governando, **ii)** quem está sendo governado e **iii)** quais recursos/ativos devem ser implantados no processo.⁴³

⁴²TARANTINO, Anthony (ed.). **Governance, risk and compliance handbook**: technology, finance, environmental and international guidance and best practices. New Jersey: John Wiley & Sons, 2008.

⁴³TARANTINO, Anthony (ed.). **Governance, risk and compliance handbook**: technology, finance, environmental and international guidance and best practices. p.2.

Ao analisar os dilemas da governança em questões relativas à tecnologia, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) aborda o desafio de regular tecnologias emergentes em um contexto em que a necessidade de intervenção e as consequências das tecnologias nem sempre são claras desde o início. A organização cita o chamado dilema de Collingridge, segundo o qual, no início do processo de inovação, quando intervenções e correções de rumo ainda seriam simples e pouco onerosas, as consequências e a necessidade da mudança podem não estar totalmente evidentes. Em contrapartida, quando a necessidade de intervenção se torna aparente, a mudança de rumo pode ser cara, onerosa e demorada.⁴⁴

A governança tecnológica, segundo a OCDE, pode ser definida como o processo de exercício de autoridade política, econômica e administrativa no desenvolvimento, difusão e operação da tecnologia nas sociedades. A governança pode incluir normas, como regulamentações, padrões e costumes, mas também pode ser operacionalizada por meio de arquiteturas físicas e virtuais que gerenciam riscos e benefícios. A governança tecnológica abrange atividades governamentais formais, mas também atividades de empresas, organizações da sociedade civil e práticas de mercado. Em seu sentido mais amplo, representa a soma das várias maneiras pelas quais indivíduos e organizações moldam a tecnologia e como, inversamente, a tecnologia molda a ordem social.⁴⁵

A OCDE enfatiza que acertar na governança tecnológica é fundamental, mas também desafiador. Muitas barreiras para viabilizar tecnologias emergentes não estão na tecnologia em si, mas na governança da tecnologia. Para alguns, a governança é muito complexa e onerosa. Para outros, os sistemas de governança não protegem valores humanos fundamentais, levando a uma crise de confiança pública na tecnologia. Ainda para outros, a governança não consegue alinhar adequadamente o desenvolvimento tecnológico aos maiores objetivos humanos. Diante dessa incerteza, instrumentos regulatórios tradicionais tendem a se concentrar em consequências imediatas ou facilmente quantificáveis e na gestão dessas consequências, ou atuam apenas depois que decisões fundamentais sobre o *design* da tecnologia já foram tomadas. A OCDE cita a IA e a neurotecnologia como

⁴⁴ TECHNOLOGY governance. OECD, [2019]. Disponível em: <https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/>. Acesso em: 23 ago. 2023.

⁴⁵ TECHNOLOGY governance.

exemplos de áreas em que as questões levantadas são mais fundamentais e de longo prazo.⁴⁶

A governança “antecipatória” de processos inovadores desloca o foco da gestão de riscos dos produtos tecnológicos para a gestão do próprio processo de inovação: quem, quando, o quê e como. Seu objetivo é antecipar preocupações desde cedo, abordá-las por meio de processos abertos e inclusivos e direcionar a trajetória da inovação em uma direção desejável.⁴⁷

Para a OCDE, a governança tecnológica emergente pode ser agrupada em um arcabouço político composto por: **i)** valores compartilhados; **ii)** critérios de *design* e **iii)** ferramentas para colocar os valores compartilhados em prática.

Os valores compartilhados, como democracia, direitos humanos, boa governança, segurança, sustentabilidade e mercados abertos, formam a base da governança tecnológica emergente. Os critérios de *design* para a governança tecnológica emergente devem ser baseados na **i)** antecipação, **ii)** inclusão e alinhamento, e **iii)** adaptabilidade.⁴⁸

A antecipação é essencial para enfrentar o dilema de Collingridge, o qual afirma que a governança de tecnologias emergentes pode ser excessivamente restritiva se ocorrer muito cedo no processo de desenvolvimento e cara ou impossível se ocorrer mais tarde. A inclusão e alinhamento envolvem a participação de um amplo conjunto de grupos de *stakeholders*, incluindo atores geralmente excluídos do processo de inovação, como pequenas empresas, regiões remotas e certos grupos sociais. A adaptabilidade é crucial, pois as tecnologias emergentes podem ter consequências imprevistas, e eventos ou resultados adversos podem ocorrer, tornando necessário que o sistema de governança seja adaptável para construir resiliência e permanecer relevante.⁴⁹ As ferramentas, por sua vez, se subdividem em: **i)** avaliação de tecnologia prospectiva (“*forward-looking technology assessment*”); **ii)** engajamento

⁴⁶ TECHNOLOGY governance.

⁴⁷ EMERGING technology governance: towards an anticipatory framework. In: OECD Science, Technology and Innovation Outlook. Paris: OECD Publishing, 2023. Disponível em: <https://www.oecd-ilibrary.org/sites/d67d5380-en/index.html?itemId=/content/component/d67d5380-en>. Acesso em: 23 ago. 2023.

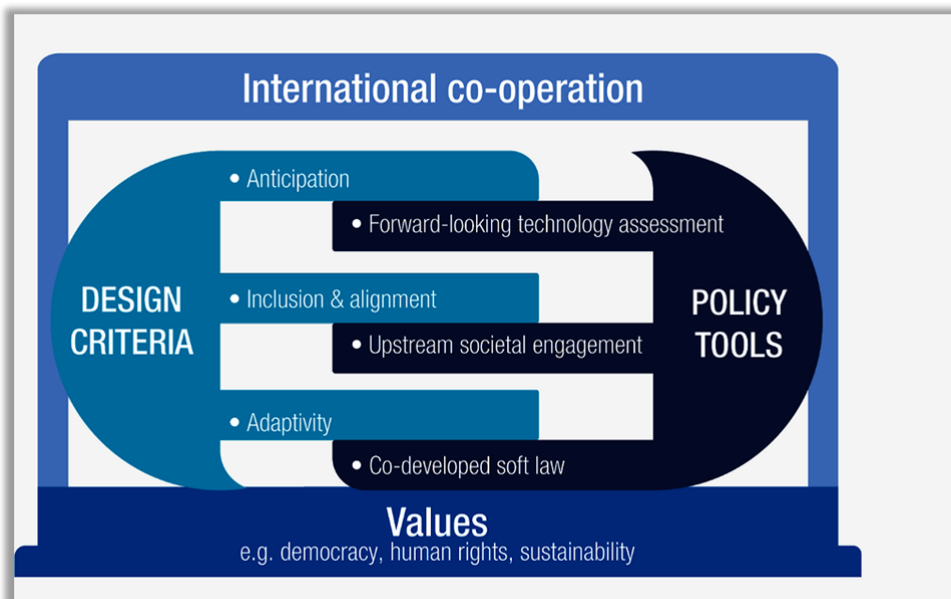
⁴⁸ EMERGING technology governance: towards an anticipatory framework.

⁴⁹ EMERGING technology governance: towards an anticipatory framework.

da sociedade (“*upstream engagement*”) e **iii**) codesenvolvimento de mecanismos de *soft law* (“*co-developed soft law*”).⁵⁰

A OCDE, ainda, destaca a relevância da cooperação internacional no âmbito da governança tecnológica. O desenvolvimento, uso e efeitos das tecnologias ultrapassam as fronteiras nacionais, o que demanda uma abordagem global para enfrentar os desafios apresentados por tais tecnologias emergentes. Essa abrangência global implica a necessidade de elaboração de sistemas de governança tecnológica tanto no âmbito nacional quanto internacional.⁵¹

Imagem 1 – Proposta de governança tecnológica emergente da OCDE⁵²



Fonte: EMERGING technology governance: towards an anticipatory framework. In: OECD Science, Technology and Innovation Outlook. Paris: OECD Publishing, 2023. Disponível em: <https://www.oecd-ilibrary.org/sites/d67d5380-en/index.html?itemId=/content/component/d67d5380-en>. Acesso em: 23 ago. 2023.

No caso específico da IA generativa, a compreensão e a gestão de seus riscos são cruciais para aproveitar os benefícios dessa tecnologia. A IA pode proporcionar

⁵⁰ EMERGING technology governance: towards an anticipatory framework.

⁵¹ EMERGING technology governance: towards an anticipatory framework.

⁵² EMERGING technology governance: towards an anticipatory framework.

avanços significativos no desenvolvimento sustentável e inteligente, como aumento da eficiência e qualidade na entrega de bens e serviços, maior autonomia e mobilidade para idosos e pessoas com deficiência, e melhoria da segurança em operações críticas, como saúde, transporte e resposta a emergências. Neste sentido, Araz Taeihagh enfatiza que, à medida que os sistemas de IA evoluem e se tornam mais complexos, também aumentam os riscos e a interconectividade com outros dispositivos inteligentes e sistemas. Essa evolução demanda a criação de mecanismos de governança específicos, como os voltados para saúde, transporte e armas autônomas, bem como um arcabouço global de governança para a IA em geral.⁵³

Segundo Taeihagh, os desafios impostos pela IA advêm de sua natureza altamente imprevisível, intratável e não linear, o que dificulta a formulação de políticas concretas pelos governos. Além disso, a opacidade e a imprevisibilidade inerentes aos sistemas de aprendizado de máquina (“*Machine learning*” ou ML) impõem desafios técnicos para garantir a responsabilidade e a governança da IA. A opacidade dos algoritmos complexos de ML limita a transparência, a explicabilidade e a responsabilização que podem ser alcançadas nos sistemas de IA.⁵⁴

Outra questão importante no debate sobre a governança da IA é a governança de dados. Segundo o autor, existem vários desafios organizacionais e tecnológicos que impedem o controle efetivo dos dados, bem como dificultam a atribuição de responsabilidade pelas decisões feitas pelos sistemas de IA orientados por dados. Os problemas relacionados à fragmentação de dados e à falta de interoperabilidade entre sistemas podem limitar o controle de uma organização sobre os fluxos de dados em todo o seu ciclo de vida. Além disso, a atribuição de responsabilidade e causalidade entre as decisões e eventos baseados em IA e as partes envolvidas na facilitação dessas decisões pode ser obscurecida devido a papéis compartilhados entre diferentes atores no compartilhamento de dados.⁵⁵

Taeihagh argumenta que os arcabouços regulatórios e de governança existentes são inadequados para gerenciar os problemas sociais introduzidos pela IA devido à falta de informações necessárias para compreender a tecnologia e aos atrasos regulatórios em relação aos desenvolvimentos da IA. A assimetria de

⁵³ TAEIHAGH, Araz. Governance of artificial intelligence. *Policy and Society*, Oxford, v. 40, n. 2, p. 137-157, 2021. Disponível em: <https://academic.oup.com/policyandsociety/article/40/2/137/6509315>. Acesso em: 23 ago. 2023.

⁵⁴ TAEIHAGH, Araz. Governance of artificial intelligence. p. 146.

⁵⁵ TAEIHAGH, Araz. Governance of artificial intelligence. p. 146.

informações entre as empresas de tecnologia e os reguladores agrava a dificuldade em compreender e aplicar novas ou antigas legislações às aplicações de IA.⁵⁶

Para enfrentar os desafios de governança apresentados pela incerteza e complexidade dos desenvolvimentos de IA, Taeihagh menciona que há crescentes apelos à adoção de abordagens inovadoras de governança, como a governança adaptativa e a governança híbrida ou ‘descentralizada’. Estas abordagens enfatizam o papel diminuído do governo no controle da distribuição de recursos na sociedade e a crescente importância dos atores não estatais na regulação da IA. Além disso, a governança adaptativa destaca a necessidade de mudar de medidas de ‘comando e controle’ para abordagens mais flexíveis, caracterizadas pelo ajuste iterativo e melhoria das regulamentações e políticas à medida que novas informações são coletadas.⁵⁷

Trata-se de sugestão que se aproxima à abordagem experimentalista. Segundo Charles F. Sabel e William H. Simon, o experimentalismo é um conjunto de práticas que visa acomodar a mudança contínua e a variação como desafios presentes nos problemas públicos atuais. Segundo os autores, as políticas devem ser experimentais no sentido de estarem sujeitas à observação constante e revisão flexível com base nas consequências observadas.⁵⁸

Sabel e Simon apresentam um modelo abstrato de experimentalismo, no qual os componentes básicos são um “centro” e um conjunto de “unidades locais”. Neste modelo, o centro e as unidades locais estabelecem e revisam metas e meios em um processo iterativo com quatro elementos básicos: definição de metas estruturais e medidas provisórias; concessão de ampla discricção às unidades locais; exigência de relatórios regulares de desempenho e participação em revisões por pares; e revisão periódica das metas, medidas de desempenho e procedimentos de tomada de decisão com base nas alternativas relatadas e avaliadas nas revisões por pares.⁵⁹

Taeihagh também trata da importância de abordagens de *soft law* que envolvam colaboração com as partes interessadas afetadas para desenvolver diretrizes, experimentação legal e *sandboxes* regulatórios para testar estruturas inovadoras

⁵⁶ TAEIHAGH, Araz. Governance of artificial intelligence. p. 146.

⁵⁷ TAEIHAGH, Araz. Governance of artificial intelligence. p. 147.

⁵⁸ SABEL, Charles F.; SIMON, William H. Minimalism and experimentalism in the administrative state. **Georgetown Law Journal**, Washington, v. 100, n. 1, p. 79, 2011. Disponível em: <https://heinonline.org/HOL/P?h=hein.journals/glj100&i=55>. Acesso em: 23 ago. 2023.

⁵⁹ SABEL, Charles F.; SIMON, William H. Minimalism and experimentalism in the administrative state. p. 80.

de responsabilidade e prestação de contas para a IA, que serão adaptadas em fases iterativas. O autor sugere que novos quadros de governança podem ser adaptados das abordagens adotadas para regular tecnologias emergentes anteriores, como a Internet, nanotecnologia, segurança da aviação e direito espacial.⁶⁰

Neste sentido, Aaron Mannes aborda os conceitos de *hard power* e *soft power* no gerenciamento de riscos associados à IA. Enquanto o *hard power* envolve a regulação governamental e políticas formais, o *soft power* diz respeito às normas e valores sociais. O autor enfatiza que ambos os tipos de abordagens precisam ser embasados por novos programas de pesquisa, que englobem tanto aspectos matemáticos quanto a compreensão de como organizações, pessoas e IA interagem.⁶¹

Ao discutir a avaliação e percepção de riscos, o autor observa que o gerenciamento de riscos requer uma avaliação cuidadosa das aplicações da IA, levando em consideração os tipos de danos que podem ocorrer quando a IA falha e como essas falhas podem acontecer. Além disso, Mannes aborda a importância do estudo da percepção de riscos, destacando que a maneira como as pessoas enxergam a IA influenciará diretamente sua adoção e implementação.⁶²

Segundo Mannes, os danos causados por falhas na IA podem ser físicos, como danos à propriedade, lesões ou mortes, ou não físicos, como perdas financeiras ou violações de privacidade. O autor também cita os conceitos de danos de alocação e danos de representação, que se referem, respectivamente, à discriminação na distribuição de recursos e oportunidades e ao reforço de atitudes e crenças discriminatórias por meio da IA.⁶³

Para gerenciar os riscos da IA, é crucial desenvolver abordagens interdisciplinares, envolvendo tecnólogos, especialistas em diferentes domínios, cientistas sociais e especialistas em políticas públicas, bem como considerar as percepções de riscos do público em geral. A análise de Mannes ressalta a necessidade de uma compreensão aprofundada dos riscos envolvidos no desenvolvimento e

⁶⁰TAEIHAGH, Araz. Governance of artificial intelligence. p. 147.

⁶¹MANNES, Aaron. Governance, risk, and artificial intelligence. *Ai Magazine*, v. 41, n. 1, p. 62, Spring 2020. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/5200>. Acesso em: 23 ago. 2023.

⁶²MANNES, Aaron. Governance, risk, and artificial intelligence. p. 63.

⁶³MANNES, Aaron. Governance, risk, and artificial intelligence. p. 63.

implementação da IA buscando estratégias para minimizar seus impactos negativos e maximizar seus benefícios para a sociedade.⁶⁴

Taeihagh, no mesmo sentido, destaca a importância do enquadramento conceitual da IA no processo de tomada de decisão. Segundo o autor, é essencial que a tomada de decisão em sistemas de IA seja cuidadosamente conceitualizada de acordo com o contexto de aplicação, e que esses processos de enquadramento estejam sujeitos ao debate público. O autor argumenta que a forma como os responsáveis políticos abordam a importância relativa dos diferentes princípios éticos, conflitos entre princípios éticos e a justificação da sua importância tem implicações críticas nos compromissos resultantes do projeto de sistemas de IA e na conformidade do público com diferentes diretrizes éticas.⁶⁵

Um dos pontos necessários em qualquer programa de governança tecnológica em inteligência artificial é justamente a auditabilidade dos processos de alinhamento ético-valorativo das IAs generativas, especialmente aquelas utilizadas para confeccionar minutas de atos jurisdicionais. Essa auditabilidade permite avaliar quais princípios éticos e valores sociais foram preestabelecidos para os sistemas de IA que serão utilizados pelos tribunais, além de identificar possíveis desvios ou falhas no sistema que possam resultar em consequências indesejáveis, como decisões enviesadas.

Trata-se de medida essencial para que o princípio da publicidade seja respeitado no funcionamento da Administração Pública. Sem que se possa auditar o alinhamento ético-valorativo das IAs utilizadas, não se poderá saber qual é a fonte em que a IA está utilizando para gerar seu conteúdo. É dizer: estar-se-á delegando a análise de mérito dos casos para ferramentas da qual não se tem conhecimento quanto ao real funcionamento.

A auditabilidade do alinhamento ético-valorativo das IAs generativas utilizadas pelo Poder Judiciário, portanto, constitui requisito obrigatório para sua implementação, e deve constar no plano de governança tecnológica de toda e qualquer entidade que a utilize.

⁶⁴ MANNES, Aaron. Governance, risk, and artificial intelligence. p. 62-63.

⁶⁵ TAEIHAGH, Araz. Governance of artificial intelligence. p. 146.

6. Conclusão

A crescente aplicação de IAs generativas no sistema judiciário tem potencial para melhorar a eficiência e a precisão das decisões, mas também levanta preocupações sobre possíveis vieses, discriminação e violações dos direitos fundamentais dos indivíduos.

Neste contexto, é necessário estabelecer mecanismos de auditoria robustos e transparentes que possam avaliar o alinhamento ético-valorativo das IAs generativas em decisões judiciais. Essa auditoria deve incluir a análise de diferentes aspectos do sistema de IA, como a qualidade e a representatividade dos dados utilizados no treinamento, a transparência dos algoritmos e a explicabilidade das decisões geradas pela IA.

Percebe-se que a regulamentação da IA ocupa um espaço de destaque cada vez maior para garantir que o desenvolvimento e o uso dessa tecnologia ocorram de maneira ética, segura e responsável. Algumas áreas-chave que precisam ser regulamentadas incluem a governança tecnológica, privacidade de dados, a transparência dos algoritmos, a responsabilidade pela tomada de decisões automatizadas, a segurança cibernética e a prevenção de discriminação algorítmica.

Assim, para enfrentar os desafios trazidos pela evolução da IA, parece evidente que serão necessários diálogos entre os agentes participantes no processo democrático, como auditorias internas e externas, a fim de averiguar se os procedimentos que visam ao desenvolvimento e uso responsável da tecnologia estão sendo respeitados, bem como parcerias com empresas do ramo de monitoria de conteúdo. Para isso também será necessária a criação de um espaço multissetorial, com transparência de dados, sob o intuito de pesquisa e desenvolvimento, direcionado para o auxílio de decisões acerca das práticas e políticas que possam impactar os direitos de modo geral.

Além disso, é essencial que haja uma coordenação entre os principais *players* do mercado da inteligência artificial, e uma estrutura regulatória global eficaz, incluindo autoridades reguladoras, sistemas de rastreamento e auditorias. A governança democrática e popular é fundamental para garantir que o desenvolvimento e uso da inteligência artificial ocorram de forma ética, segura e responsável, com foco no bem-estar da sociedade e na redução das desigualdades sociais e regionais.

Neste contexto, a auditabilidade do alinhamento ético-valorativo das IAs generativas constitui requisito obrigatório para sua implementação pelo poder público, e deve ser pauta de qualquer plano de governança tecnológica nas corporações.

Referências

AI principles. **future of life Institute**, Aug. 2017. Disponível em: <https://futureoflife.org/open-letter/ai-principles/>. Acesso em: 8 abr. 2023.

ALTMAN, Sam. **Things we need for a good AGI future**: 1) the technical ability to align a superintelligence 2) sufficient coordination among most of the leading AGI efforts 3) an effective global regulatory framework including democratic governance. San Francisco, 29 Mar. 2023. X Twitter: @sama. Disponível em: <https://twitter.com/sama/status/1641229941131051008>. Acesso em: 4 abr. 2023.

BRAGANÇA, Fernanda; BRAGANÇA, Laurinda Fátima da F.P.G. Revolução 4.0 no Poder Judiciário: levantamento do uso de inteligência artificial nos tribunais brasileiros. **Revista da Seção Judiciária do Rio de Janeiro**, Rio de Janeiro, v. 23, n. 46, p. 65-76, jul./out. 2019.

BRASIL. Congresso Nacional. Senado Federal. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, [2023]. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 10 ago. 2023.

BYRNE, Matthew D. Generative Artificial Intelligence and ChatGPT. **Journal of PeriAnesthesia Nursing**, v. 38, n. 3, p. 519-522, Apr. 2023.

CHAVES, Guilherme; VEIGA, Elizabeth. A inteligência artificial na formação dos precedentes do STJ: Sistema Sócrates 2.0. **Migalhas**, 31 maio 2021. Disponível em: <https://www.migalhas.com.br/depeso/346278/a-inteligencia-artificial-na-formacao-dos-precedentes-do-stj>. Acesso em: 4 abr. 2023.

CORVALÁN, Juan Gustavo. Inteligencia Artificial GPT-3, PretorIA y oráculos algorítmicos en el Derecho. **International Journal of Digital Law**, Belo Horizonte, v.1, n. 1, p. 11-52, jan./abr. 2020. Disponível em: <https://journal.nuped.com.br/index.php/revista/libraryFiles/downloadPublic/24>. Acesso em: 22 ago. 2023.

DOSOVITSKIY, Alexey *et al.* An image is worth 16x16 words: transformers for image recognition at scale. *In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS*, 9, 2021. Annals [...]. **ICLR**, [2021]. Disponível em: <https://iclr.cc/virtual/2021/poster/3013>. Acesso em: 22 ago. 2023. p 1-9. 2021.

EMERGING technology governance: towards an anticipatory framework. *In: OECD Science, Technology and Innovation Outlook*. Paris: OECD Publishing, 2023. Disponível em: <https://www.oecd-ilibrary.org/sites/d67d5380-en/index.html?itemId=/content/component/d67d5380-en>. Acesso em: 23 ago. 2023.

ESTUDO revela que 44 tribunais, além do Conselho Nacional de Justiça, usam inteligência artificial. **FGV**, 20 jul. 2022. Disponível em: <https://portal.fgv.br/noticias/estudo-revela-44-tribunais-alem-conselho-nacional-justica-usam-inteligencia-artificial>. Acesso em: 4 abr. 2023.

GENERATIVE AI could raise global GDP by 7%. **Goldman Sachs**, 5 Apr. 2023. Disponível em: <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>. Acesso em: 6 abr. 2023.

GOODFELLOW, Ian J. *et al.* Generative adversarial nets. *In: GHAHRAMANI, Z. et al. (ed.). Advances in neural information processing systems 27 (NIPS 2014)*. Montréal: Université de Montréal, Département d'informatique et de recherche opérationnelle, 2014. p. 1-8. Disponível em: https://papers.nips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html. Acesso em: 23 ago. 2023.

GORDON, Cindy. ChatGPT is the fastest growing app in history of web applications. **Forbes**, Feb. 2023. Disponível em: <https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/?sh=6a692910678c>. Acesso em: 6 abr. 2023.

GREWAL, Dalvinder Singh. A critical conceptual analysis of definitions of artificial intelligence as applicable to computer engineering. **IOSR Journal of Computer Engineering**, v. 16, n. 2, p. 9-13, Jan. 2014.

HECKBERT, Paul S. Survey of texture mapping. **IEEE Computer Graphics and Applications**. v. 6, n. 11, p. 56 –67, Nov. 1986.

JOVANOVIC, Mladan; CAMPBELL, Mark. Generative artificial intelligence: trends and prospects. **Computer**, v. 55, n. 10, p. 107-112, Oct. 2022.

KELLER, Clara Iglesias *et al.* Oito medidas para regular big techs. **Folha de S. Paulo**. fev. 2023. Disponível em: <https://www1.folha.uol.com.br/ilustrissima/2023/02/oito-medidas-para-regular-big-techs-garantindo-liberdade-de-expressao.shtml>. Acesso em: 3 abr. de 2023.

KHANDELWAL, Urvashi *et al.* Sharp nearby, fuzzy far away: how neural language models use context. *In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 56th, 2018, Melbourne, Australia. **Proceedings [...]**. Melbourne, Australia: Association for Computational Linguistics, p. 284 –294, July 2018. (Long Papers, v.1). Disponível em: <https://aclanthology.org/P18-1027/>. Acesso em: 23 ago. 2023.

LEMOINE, Blake. Is LaMDA sentient?: an interview. **Medium**, Jun. 2022. Disponível em: <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>. Acesso em: 6 abr. 2023.

MANNES, Aaron. Governance, risk, and artificial intelligence. **Ai Magazine**, v. 41, n. 1, p. 61-69, Spring 2020. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/5200>. Acesso em: 23 ago. 2023.

MELO, Jeferson. Judiciário ganha agilidade com uso de inteligência artificial. **Conselho Nacional de Justiça**, 3 abr. 2019. Disponível em: <https://www.cnj.jus.br/judiciario-ganha-agilidade-com-uso-de-inteligencia-artificial/>. Acesso em: 4 abr. 2023.

NUNES, Dierle José Coelho; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. **Revista de Processo**, São Paulo, v. 43, n. 285, p. 421-447, nov. 2018.

PAUSE giant AI experiments: an open letter. **future of life Institute**, Mar. 2023. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 4 abr. 2023.

POR que o ChatGPT, o Midjourney e o Dall-E são uma nova era na inteligência artificial? **Pequenas Empresas & Grandes Negócios**, Rio de Janeiro, 31 mar. 2023. Disponível em: <https://revistapegn.globo.com/tecnologia/noticia/2023/03/por-que-o-chatgpt-o-midjourney-e-o-dall-e-sao-uma-nova-era-na-inteligencia-artificial.ghtml>. Acesso em: 23 ago. 2023.

PROJETO Victor avança em pesquisa e desenvolvimento para identificação dos temas de repercussão geral. **Supremo Tribunal Federal**, 19 ago. 2021. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=471331&ori#:~:text=O%20Victor%20C3%A9%20uma%20intelig%C3%Aancia,aplica%20ao%20caso%20dos%20autos>. Acesso em: 4 abr. 2023.

QINGXIU, Dong *et al.* A survey on in-context learning. **arXiv**, 1 Jun. 2023. Disponível em: <https://arxiv.org/abs/2301.00234>. Acesso em: 22 ago. 2023.

REVOLUÇÃO tecnológica e desafios da pandemia marcaram gestão do ministro Noronha na presidência do STJ. **STJ**, 23 ago. 2020. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/23082020-Revolucao-tecnologica-e-desafios-da-pandemia-marcaram-gestao-do-ministro-Noronha-na-presidencia-do-STJ.aspx>. Acesso em: 23 ago. 2023.

ROQUETE, Felipe Leitão Valadares. Inteligência artificial, valores e incerteza: um problema insolúvel para a regulação? **Internet&sociedade**, São Paulo, v. 2, n. 2, p. 150-154, 2021. Disponível em: <https://revista.internetlab.org.br/wp-content/uploads/2022/03/Inteligencia-artificial-valores-e-incerteza-um-problema-insolavel-para-a-regulacao.pdf>. Acesso em: 23 ago. 2023.

SABEL, Charles F.; SIMON, William H. Minimalism and experimentalism in the administrative state. **Georgetown Law Journal**, Washington, v. 100, n. 1, p. 53-93, 2011. Disponível em: <https://heinonline.org/HOL/P?h=hein.journals/glj100&i=55>. Acesso em: 23 ago. 2023.

SALOMÃO, Luis Felipe (coord.). **Inteligência artificial**: tecnologia aplicada à gestão dos conflitos no âmbito ao Poder Judiciário brasileiro. 2. ed. Rio de Janeiro: FGV Conhecimento, 2021. Disponível em: https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf. Acesso em: 4 abr. 2023.

SRIRAM, Akash; KACHWALA, Zaheer. Rival do Twitter, Threads bate recorde e atinge 100 milhões de usuários em cinco dias. Edição de Sriraj Kalluvila, Rashmi Aich. **CNN Brasil**, 10 jul. 2023. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/rival-do-twitter-threads-bate-recorde-e-atinge-100-milhoes-de-usuarios-em-cinco-dias>. Acesso em: 10 ago. 2023.

TAEIHAGH, Araz. Governance of artificial intelligence. **Policy and Society**, Oxford, v. 40, n. 2, p. 137-157, 2021. Disponível em: <https://academic.oup.com/policyandsociety/article/40/2/137/6509315>. Acesso em: 23 ago. 2023.

TARANTINO, Anthony (ed.). **Governance, risk and compliance handbook: technology, finance, environmental and international guidance and best practices.** New Jersey: John Wiley & Sons, 2008.

TECHNOLOGY governance. **OECD**, [2019]. Disponível em: <https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/>. Acesso em: 23 ago. 2023.

YIHAN, Cao *et al.* A comprehensive survey of AI-Generated Content (AIGC): a history of generative AI from GAN to ChatGPTJ. **Association for Computing Machinery**, v. 37, n. 4, Article 111, Aug. 2018. Disponível em: <https://arxiv.org/pdf/2303.04226.pdf>. Acesso em: 22 ago. 2023.