

**INSTITUTO BRASILEIRO DE ENSINO, DESENVOLVIMENTO E PESQUISA
CURSOS DE PÓS GRADUAÇÃO LATO SENSU**

**GOVERNANÇA DA INTELIGÊNCIA ARTIFICIAL:
REGULAMENTAÇÕES E PRÁTICAS ÉTICAS**

Thiago Tomich Netto Guterres Soares¹

Isabel Maria Rosal Santos²

RESUMO

A Inteligência artificial (IA) está intimamente relacionada a capacidade de obter, guardar, manter e processar grandes volumes de dados através de máquinas, as quais conseguem tomar decisões informadas. A medida que a tecnologia avança, novos desafios aparecem, mormente quanto à questões de privacidade dos dados obtidos por intermédio dessa nova forma de tecnologia, bem como questões regulamentares, sociais e éticas, que foram o objeto de estudo da presente pesquisa. O objetivo principal é estudar sobre a governança da inteligência artificial, sua regulamentação e suas práticas éticas, analisando o papel dos comitês de expert na formulação de diretrizes imprescindíveis a preservação de direitos que surgem com a detenção de informações e dados pessoais. O problema central da pesquisa é saber se os comitês de expert auxiliam de forma eficaz no desenvolvimento de tecnologia de inteligência artificial. Foram utilizados como referenciais teóricos o direito comparado, com a análise de diferentes sistemas jurídicos, o que auxilia a identificar boas práticas e possíveis soluções para desafios éticos e legais relacionados à inteligência artificial, bem como doutrina, com abordagens sobre a governança de dados de maneira responsável e sustentável, considerando aspectos de privacidade e proteção de dados. Na metodologia foram utilizadas Revisão de literatura, estudo comparativo, estudo de casos e análise crítica. O estudo contribui para o campo da Inteligência Artificial ao fornecer uma compreensão abrangente das questões éticas e legais, propondo diretrizes e recomendações para o desenvolvimento e implementação responsável dessa tecnologia.

1

PALAVRAS-CHAVE: Inteligência artificial; Governança; Ética.

ABSTRACT

Artificial Intelligence (AI) is closely related to the capacity to obtain, store, maintain, and process large volumes of data through machines, which can make informed decisions. As technology advances, new challenges emerge, particularly regarding data privacy issues obtained through this new form of technology, as well as regulatory, social, and ethical issues, which were the subject of this research. The main objective is to study the governance of artificial intelligence, its regulation, and its ethical practices, analyzing the role of expert committees in formulating essential guidelines for preserving rights that arise with the possession of personal information and data. The central problem of the research is to determine whether expert committees effectively assist in the development of artificial intelligence technology. Theoretical frameworks such as comparative law were used, analyzing different legal systems, which helps identify best practices and possible solutions for ethical and legal challenges related to artificial intelligence, as well as doctrine, with approaches on data governance in a responsible and sustainable manner, considering aspects of privacy and data protection. The methodology included a literature review, comparative study, case study, and critical analysis. The study contributes to the field of Artificial Intelligence by providing a comprehensive understanding of ethical and legal issues, proposing guidelines and recommendations for the responsible development and implementation of this technology.

¹ *Bacharel em Direito.*

E-mail: thiagotomich@gmail.com

² *Bacharel em Direito, Mestre em Direito pela Universidade de Brasília. Pesquisadora na área de Direito Digital.*

E-mail: isabelamrosal@gmail.com

KEYWORDS: Artificial Intelligence; Health Governance; Ethics.

INTRODUÇÃO

A Inteligência Artificial (IA) emergiu como uma das forças motrizes mais transformadoras do século XXI, revolucionando diversos setores e remodelando a maneira como vivemos e trabalhamos. Desde a sua concepção, a IA tem oferecido uma vasta gama de oportunidades, proporcionando avanços significativos em áreas como saúde, educação, segurança, indústria e serviços públicos. A capacidade de máquinas para processar grandes volumes de dados, aprender com eles e tomar decisões informadas abriu novas possibilidades para a inovação e eficiência.

No entanto, a integração da IA no nosso dia a dia também traz consigo desafios substanciais. Questões éticas, regulamentares e de privacidade surgem à medida que a IA se torna mais prevalente, levantando preocupações sobre a transparência das decisões automatizadas, a equidade no acesso à tecnologia e a responsabilidade em caso de falhas. Além disso, a automação impulsionada pela IA tem o potencial de transformar o mercado de trabalho, exigindo uma adaptação das habilidades e competências dos trabalhadores para um novo panorama digital.

O artigo em seu capítulo inicial, se propõe a estudar o impacto profundo e multifacetado da Inteligência Artificial no mundo moderno. Ele busca explorar não apenas as oportunidades que a IA oferece para melhorar a eficiência, a precisão e a inovação em diversos setores, mas também os desafios éticos e sociais e regulatório que ela impõe. Posteriormente a pesquisa aborda sobre a importância dos Comitês de *Expert* na implementação da IA de maneira sustentável e equitativa, assegurando que os benefícios sejam amplamente distribuídos enquanto mitiga os riscos associados.

Capítulo 1: O ESTADO DA ARTE EM INTELIGÊNCIA ARTIFICIAL

Estamos imersos em um período marcado por avanços tecnológicos significativos, incluindo aprendizado de máquina, *blockchain*, *chat gpt*, criptomoedas, *big data*, *internet* e *bots*. Embora tais inovações tenham um impacto profundo em nossa vida cotidiana, compreender seu alcance total é complexo (Machado, 2021). Durante muito tempo, a Inteligência Artificial foi vista como um conceito de ficção científica, distante da realidade prática. No entanto, a evolução da tecnologia demonstrou o imenso potencial das máquinas para realizar tarefas antes consideradas intelectuais. Apesar da inteligência desses sistemas ser diferente da humana, a IA é agora uma realidade concreta. Sendo o desafio atual expandir nossa

compreensão de seus mecanismos e explorar suas várias possibilidades, especialmente para quem busca aumentar a produtividade (FIA – Business School, 2023).

A ideia de inteligência de máquina foi explorada inicialmente por *Alan Turing*³, conhecido por criar o famoso teste que leva seu nome. O objetivo do Teste de *Turing* era verificar se um determinado programa possuía inteligência. A avaliação consistia em uma interação textual entre o programa e um ser humano, onde o interrogador fazia diversas perguntas para determinar se estava conversando com uma máquina ou com uma pessoa. Se o programa conseguisse se passar por um ser humano em 30% das interações, ele seria considerado aprovado no teste (Russell, 2013).

Particularmente, o aprendizado de máquina versa sobre métodos que permitem aos computadores aprender a partir de dados. Esse campo se desenvolveu rapidamente nos últimos anos devido ao acesso a enormes quantidades de dados e ao aumento do poder de processamento analítico. Isso fez com que a IA contemporânea fosse vista como a capacidade computacional de interpretar grandes volumes de informações para tomar decisões, ao invés de focar na compreensão da inteligência humana ou na representação do conhecimento e raciocínio (Dignum, 2019).

Atualmente, não existe conceito estático e pré-definido de IA, mas existem definições bastante aceitas. Por exemplo, o Grupo de Peritos de Alto Nível da Comissão Europeia propôs uma definição abrangente que inclui a funcionalidade da IA em seu contexto sistêmico. Assim, os sistemas de Inteligência Artificial (IA) são definidos como *softwares* e, possivelmente, *hardwares* desenhados por humanos que, dado um objetivo complexo, atuam no ambiente físico ou digital percebendo seu entorno através da aquisição de dados. Eles interpretam os dados coletados, raciocinam sobre o conhecimento ou processam a informação derivada desses dados e decidem as melhores ações a serem tomadas para alcançar o objetivo dado. Os sistemas de IA podem usar regras simbólicas ou aprender um modelo numérico, e também podem adaptar seu comportamento analisando como o ambiente é afetado por suas ações anteriores (*High-Level Expert Group*, 2019).

³Alan Turing, um dos principais pioneiros da computação, é amplamente reconhecido por seus estudos inovadores e por desenvolver, nos anos 1930, o conceito de "máquina lógica da computação", que é considerado um precursor do computador moderno (ISAACSON, Walter. Os inovadores: uma biografia da revolução digital. São Paulo: Companhia das Letras, 2014. p. 55; 61). Recentemente, sua notoriedade aumentou com o lançamento do filme "The Imitation Game", que conta sua história.

O atual desenvolvimento da Inteligência Artificial e da Aprendizagem de Máquina é marcado por uma rápida evolução tecnológica, impulsionada pelo aumento da disponibilidade e da qualidade dos dados, pelo aprimoramento dos algoritmos e das técnicas, pelo crescimento da capacidade de processamento e armazenamento dos computadores além do surgimento de novas plataformas e ferramentas. Esse cenário também é caracterizado por uma ampla difusão e aplicação dessas tecnologias em diversos setores da sociedade (FIA – *Business School*, 2023). Sendo capaz de criar sistemas e dispositivos que simulam ou ampliam a capacidade humana de resolver problemas, tomar decisões, aprender, criar e se comunicar, a Inteligência Artificial tem trazido benefícios e oportunidades para diversos setores e atividades, como a saúde, a educação, a segurança, a indústria, o comércio, os serviços públicos, entre outros (Kaufman, 2018).

Com o progresso das tecnologias, tornou-se viável transferir a delegação de atividades e decisões, anteriormente realizadas por pessoas físicas ou jurídicas, diretamente para sistemas de inteligência artificial. Essa evolução permite que a IA assuma papéis significativos na tomada de decisões e na execução de tarefas que antes eram exclusivamente humanas

No cenário político, a IA já vem transformando processos democráticos e políticas públicas, sendo uma ferramenta poderosa para aumentar a eficiência governamental, agilizando processos e melhorando a tomada de decisões. No entanto, essa mesma tecnologia levanta sérias preocupações sobre vigilância excessiva, manipulação de informações e ampliação das desigualdades sociais. Além disso, a IA pode ter um impacto profundo na segurança nacional e nas relações internacionais, exigindo uma abordagem cuidadosa e ética para garantir que seus benefícios sejam aproveitados sem comprometer os direitos e liberdades fundamentais.

No cenário jurídico, a IA desafia as estruturas legais existentes em áreas como privacidade de dados, direitos autorais e responsabilidade civil. Por exemplo, quem é responsável quando um carro autônomo causa um acidente? Como a IA deve ser regulamentada para garantir a privacidade dos dados do usuário? Como os direitos autorais se aplicam quando a IA cria uma obra de arte ou um texto?

Por outro lado, os estudos destacam a importância crucial da integração de tecnologias avançadas, especialmente a IA, no cotidiano jurídico brasileiro. Eles enfatizam o potencial transformador da IA para aprimorar a administração da justiça, reduzindo a morosidade processual, permitindo análises preditivas de decisões e automatizando tarefas. É crucial entender como a IA pode auxiliar a tomada de decisões e aplicações do direito. A diversidade de fontes de dados jurídicos e as facilidades proporcionadas, exigem novas pesquisas para

explorar todo o seu potencial, promovendo uma abordagem inovadora na gestão de informações e deliberações (Rosa, 2019).

No contexto brasileiros, as instituições já têm adotado sistemas de IA de diversas formas com o objetivo de otimizar processos e aumentar a eficiência. O Supremo Tribunal Federal (STF), em colaboração com a Universidade de Brasília, desenvolveu o "Projeto Victor", que utiliza IA para identificar temas de repercussão geral, agilizando a avaliação dos processos. No Estado de São Paulo, a *start-up* Nama criou o robô "Poupinha", que auxilia na emissão de documentos. Em Minas Gerais o Tribunal de Justiça do estado implementou o robô Radar, que automatiza decisões judiciais ao reconhecer recursos semelhantes e aplicar a jurisprudência correta, gerando minutas de voto padronizadas. No Distrito Federal a Procuradoria-Geral usa a IA "Dra. Luiza" para gerenciar processos jurídicos em massa e localizar informações relevantes sobre os envolvidos (Beckhauser, 2023). Essas iniciativas mostram o impacto positivo da IA na eficiência dos serviços públicos e na administração da justiça no Brasil.

O No campo jurídico, a IA apresenta desafios. Suas aplicações já incluem a análise de contratos, identificação de padrões em jurisprudência, sistemas de prevenção de crimes e automação de processos legais. Essas utilizações levantam questões legais complexas sobre responsabilidade, transparência, privacidade e ética. Adaptar as leis para abordar essas questões é crucial para garantir o uso ético da IA (Jonas, 2020).

Destaca-se, a título exemplificativo, alguns dilemas jurídicos relacionados à aplicação de Inteligência Artificial já identificados na doutrina brasileira e que são transmitidos, também, para o contexto jurídico:

Sendo capaz de criar sistemas e dispositivos que simulam ou ampliam a capacidade humana de resolver problemas, tomar decisões, aprender, criar e se comunicar, a IA tem trazido benefícios e oportunidades para diversos setores e atividades, como a saúde, a educação, a segurança, a indústria, o comércio, os serviços públicos, entre outros (Kaufman, 2018).

No campo jurídico, a IA apresenta desafios. Suas aplicações já incluem a análise de contratos, identificação de padrões em jurisprudência, sistemas de prevenção de crimes e automação de processos legais. Essas utilizações levantam questões legais complexas sobre responsabilidade, transparência, privacidade e ética. Sob essa perspectiva, adaptar as leis para abordar essas questões é crucial para garantir o uso ético da IA (Jonas, 2020).

Destaca-se, a título exemplificativo, alguns dilemas jurídicos relacionados à Inteligência Artificial já identificados na doutrina brasileira:

(a) Responsabilidade civil:

Os danos decorrentes do uso de sistemas de IA é um tema complexo e relevante. O ordenamento jurídico brasileiro prevê no Código Civil, especialmente nos artigos 186 e 927, a instituição da responsabilidade civil, a qual estabelece que qualquer pessoa que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem, fica obrigada a repará-lo (Brasil, 2002).

No contexto da IA, isso levanta questões sobre quem deve ser responsabilizado em caso de falhas ou danos causados por sistemas autônomos. Se um carro autônomo causar um acidente, por exemplo, a responsabilidade pode recair sobre diversos agentes: o fabricante do veículo, o programador do software, o proprietário do carro ou até mesmo o fornecedor dos dados utilizados pelo sistema.

Casos práticos, como o acidente com um *Tesla Model S* em modo *Autopilot*⁴ e práticas de *Geopricing* e *Geoblocking* no *e-commerce*⁵, evidenciam a importância de se analisar a responsabilidade tanto do fabricante quanto do usuário, além de avaliar a adequação e segurança dos sistemas de IA. Além disso, a complexidade dos sistemas de IA, que muitas vezes funcionam como "caixas-pretas"⁶ onde o processo de tomada de decisão não é inteiramente transparente, o que dificulta a atribuição de responsabilidade. Isso exige a criação de novos padrões de diligência e regulamentações específicas que possam abordar esses desafios, garantindo que os sistemas de IA sejam seguros, eficientes e alinhados com os direitos dos indivíduos.

A regulamentação específica e a criação de padrões mínimos de diligência são essenciais para garantir a segurança e a responsabilização adequada em casos de danos causados por IA. A análise de casos práticos mostra a necessidade de uma abordagem cuidadosa que considere aspectos tecnológicos e jurídicos, assegurando que a inovação tecnológica não comprometa a segurança e os direitos dos indivíduos.

(b) Proteção de dados

⁴https://cognitiojuris.com.br/autopilot-tesla-tendencias-regulatorias-e-desenvolvimento-continuo-baseado-na-inteligencia-artificial/#google_vignette Acesso em: 31 de Outubro de 2024.

⁵ MARTINS, Guilherme. **O Geopricing e Geoblocking e Seus Efeitos nas Relações de Consumo** In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade**. São Paulo (SP): Editora Revista dos Tribunais. 2020. Disponível em: <https://www.jusbrasil.com.br/doutrina/inteligencia-artificial-e-direito-etica-regulacao-e-responsabilidade/1196969611>. Acesso em: 31 de Outubro de 2024.

⁶ O'NEIL, Cathy. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. New York: Random House Audio, 2016.

No âmbito da proteção de dados, observa-se uma série de violações relacionadas ao uso de dados pessoais por sistemas de IA. Exemplos comuns no cotidiano brasileiro incluem ligações indesejadas de *bots* de empresas e mensagens no *WhatsApp* de números desconhecidos, tentando vender produtos ou serviços, muitas vezes sem o consentimento adequado dos destinatários. Além disso, variações abruptas de preços, justificadas pela taxa dinâmica⁷ de oferta e demanda dos aplicativos, levantam questões sobre a transparência e a ética no uso desses dados. São inúmeros os casos de uso de dados pessoais que resultam em discriminação para o titular.

A coleta de dados para alimentar sistemas de inteligência artificial suscita preocupações significativas, especialmente porque esses dados frequentemente são obtidos sem o consentimento ou conhecimento dos cidadãos e consumidores. Esses dados brutos são processados para gerar novos dados, permitindo que a IA realize uma de suas funções primordiais: a predição⁸.

A predição, no âmbito da inteligência artificial, refere-se à capacidade de antecipar resultados ou comportamentos futuros com base em padrões identificados em grandes volumes de dados. Por exemplo, algoritmos de IA podem prever quais produtos um consumidor está mais propenso a adquirir ou identificar quais pacientes têm maior risco de desenvolver determinadas doenças. Embora essa capacidade preditiva seja fundamental para processos decisórios e possa trazer benefícios significativos, a forma como os dados são coletados e utilizados pode infringir a privacidade e os direitos dos indivíduos. A falta de transparência e consentimento no uso desses dados para predições ressalta a necessidade de regulamentações rigorosas para proteger a privacidade e garantir o uso ético da IA.

Essas problemáticas se agravam ainda mais quando as decisões algorítmicas são totalmente automatizadas e sem intervenção humana. A falta de supervisão pode resultar em decisões enviesadas e injustas, destacando a necessidade de implementar mecanismos de controle para garantir que a inteligência artificial opere de maneira ética e responsável, respeitando os direitos dos titulares dos dados pessoais.

(c) Ética:

⁷ <https://www.uber.com/us/es/marketplace/pricing/surge-pricing/>

⁸ <https://ri.ufs.br/bitstream/riufs/10428/2/RevisaoBibliograficaManutencao.pdf>

A ética na inteligência artificial (IA) diz respeito aos valores e princípios que devem orientar o desenvolvimento e a utilização dessas tecnologias. O objetivo central é garantir que a IA contribua para o bem comum, respeite a dignidade humana e promova justiça social, diversidade cultural e sustentabilidade ambiental. É igualmente imperativo que a inovação tecnológica seja acompanhada de uma sólida confiança digital. Para tal, é necessário estabelecer critérios rigorosos para avaliar a qualidade, eficiência, segurança, robustez, explicabilidade, auditabilidade, responsabilidade e governança dos sistemas de IA (Larson, 2020).

A prevenção de impactos negativos sobre indivíduos, grupos e a sociedade deve ser uma prioridade central no desenvolvimento e implementação de sistemas de inteligência artificial. É crucial garantir que esses sistemas operem de maneira ética, protegendo os direitos dos usuários e promovendo a justiça. Além disso, é importante lembrar que algoritmos são criados por seres humanos, carregando consigo seus vieses e falibilidades. Esses vieses podem desvirtuar os resultados, favorecendo interesses específicos, mesmo quando os dados e o processamento são de alta qualidade. Correlações encontradas pelos algoritmos nem sempre representam relações reais, destacando a necessidade de supervisão humana contínua e crítica no uso da IA (Frazão, 2017).

O famoso caso da *Cambridge Analytica*⁹, em particular, destaca a importância da ética no uso de dados e como a manipulação indevida pode levar a sérias consequências sociais e políticas, reforçando a necessidade de regulamentações robustas e práticas transparentes.

(d) Direitos humanos:

Relacionam-se à garantia e ao respeito aos direitos fundamentais dos indivíduos e dos povos que podem ser afetados pela IA, como o direito à vida, à liberdade, à igualdade, à não discriminação, à privacidade, à informação, à educação, à saúde, à cultura, à participação política, entre outros. A questão envolve assegurar que o desenvolvimento e o uso da IA estejam em conformidade com a Constituição Federal, a Declaração Universal dos Direitos Humanos, tratados e convenções internacionais de direitos humanos, bem como outras normas jurídicas nacionais e internacionais que protegem esses direitos. Além disso, envolve monitorar e relatar possíveis violações de direitos humanos cometidas por ou com a ajuda da IA, além de buscar formas de reparação ou responsabilização dos infratores.

⁹https://www.academia.edu/75847075/Cambridge_Analytica_Esc%C3%A2ndalo_Legado_e_Poss%C3%ADveis_Futuros_para_a_Democracia

No continente americano, diversas práticas de policiamento preditivo têm levantado sérias preocupações quanto à violação de direitos humanos. Um exemplo notável é o *software Beware*, que classifica indivíduos com base em dados coletados de suas atividades *online*, e é utilizado por entidades policiais¹⁰. Essas ferramentas podem resultar em discriminações seja pela má qualidade dos dados utilizados, que podem ser imprecisos ou desatualizados pelos vieses inerentes dos desenvolvedores da tecnologia. Os usos de tais *softwares* podem levar a ações policiais injustas e desproporcionais, afetando negativamente comunidades vulneráveis e minoritárias, e subvertendo os princípios de justiça e igualdade. Portanto, é crucial uma supervisão rigorosa e a implementação de regulamentações claras para garantir que essas tecnologias sejam utilizadas de forma ética e responsável, respeitando os direitos fundamentais dos indivíduos (Braga, 2020).

Em âmbito global, governos, organizações internacionais e partes interessadas estão empenhados no desenvolvimento de IA. A crescente incorporação dessas tecnologias nos processos decisórios destaca a necessidade de avaliar o impacto potencial dos sistemas de IA baseados em dados sobre os indivíduos e a sociedade como um todo. As implicações do processamento de dados transcendem as questões convencionais de privacidade e proteção de dados, abrangendo preconceitos contra grupos específicos de indivíduos e uma gama mais ampla de direitos fundamentais. Além disso, a tensão entre o uso extensivo de sistemas baseados em grandes volumes de dados e a demanda crescente por um uso ético e socialmente responsável desses dados evidencia a carência de um quadro regulatório apropriado para abordar as questões sociais suscitadas pelas tecnologias de IA (Mantelero, 2022).

1.1 O QUE É ALGORITMO: Impactos e Desafios no Contexto Jurídico

O conceito de algoritmo pode ser definido como uma sequência finita e ordenada de instruções ou regras que, dadas determinadas entradas, permitem alcançar um resultado ou solução para um problema específico. Algoritmos são essenciais para a programação e o funcionamento dos sistemas de Inteligência Artificial, particularmente no campo da aprendizagem de máquina profunda, que utiliza algoritmos avançados e complexos para aprender a partir de grandes volumes de dados. Esses algoritmos têm a capacidade de realizar tarefas que normalmente exigem inteligência humana, tais como o reconhecimento de padrões, a classificação de informações e a previsão de resultados (Quaresma, 2021).

¹⁰ [1] Justin Jouvenal, The New Way Police Are Surveilling You: Calculating Your Threat ‘Score,’ WASH. POST

O uso de algoritmos vem beneficiando diversas áreas, como saúde, educação, transporte e comunicação, otimizando processos, personalizando serviços e criando produtos. No entanto, também apresenta riscos técnicos, éticos, jurídicos e sociais. Dentre eles destacam-se algumas das preocupações já existentes:

- (I) A opacidade dessas tecnologias pode dificultar a compreensão de seu funcionamento, os dados utilizados, os critérios seguidos, os resultados produzidos e as consequências geradas. Essa falta de transparência pode impactar a confiança, responsabilidade, prestação de contas e contestação dos sistemas de IA, além da proteção dos direitos e interesses das pessoas ou organizações afetadas;¹¹
- (II) Enviesamento ou discriminação pode ocorrer devido à qualidade, representatividade, diversidade ou imparcialidade dos dados utilizados para treinar ou alimentar os sistemas de IA, ou devido à intencionalidade, subjetividade, ideologia ou cultura dos desenvolvedores ou usuários dos sistemas de IA. Esse enviesamento pode gerar resultados injustos, desiguais, estereotipados, preconceituosos ou ofensivos, que afetam negativamente as pessoas ou grupos vulneráveis ou minoritários, ou que favorecem interesses ou agendas particulares ou privilegiados;¹²
- (III) Inseguranças ou vulnerabilidades podem surgir de falhas, erros, defeitos, *bugs* ou ataques direcionados aos sistemas de IA comprometendo sua funcionalidade, integridade, disponibilidade, confiabilidade ou robustez. Tal insegurança pode resultar em danos, prejuízos, perdas, roubos, fraudes ou sabotagens de dados, informações, sistemas, infraestruturas, serviços ou dispositivos conectados à IA. Além disso, pode afetar a privacidade, segurança, saúde ou vida das pessoas ou organizações envolvidas no uso da IA;
- (IV) A tomadas de decisões automatizadas: refere-se à capacidade dos sistemas de IA de tomar decisões ou realizar ações sem a intervenção, supervisão, controle ou

¹¹ O'NEIL, Cathy. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. London: Penguin Books, p. 28-29

¹² No original “He described the problems at NOPD as “serious, wide-ranging, systemic and deeply-rooted within the culture of the department”. Disponível em:

[https://www.nola.com/crime/2011/03/yearlong_justice_department_pr.html]

consentimento humano. Estes sistemas podem operar de forma independente, auto-organizada, adaptativa ou com aprendizagem automática. Tal autonomia levanta questões éticas, morais, legais e sociais relativas à responsabilidade, autoridade, liberdade, dignidade e direitos humanos em relação aos sistemas de IA. Além disso, discute-se os limites, valores, princípios e finalidades da IA na sociedade.

Diante desses riscos e desafios, é necessário estabelecer mecanismos de avaliação e de governança dessas tecnologias na sociedade, levando em conta seus possíveis impactos e implicações, e buscando garantir o desenvolvimento e o uso ético, legal, seguro, confiável, inclusivo, democrático e sustentável da IA.

Destaca-se, que a abordagem de classificação empregada pelos algoritmos contemporâneos não se limita apenas aos indivíduos e às categorias tradicionalmente associadas ao tratamento injusto ou prejudicial de diferentes grupos de pessoas. Os algoritmos formam agrupamentos de indivíduos com características que vão além dos motivos tradicionalmente protegidos, como os hábitos dos consumidores, estilo de vida, comportamento *online* e *offline*, rede de relações pessoais, entre outros. Por isso, a ampla aplicação de tecnologias preditivas baseadas nessas novas categorias e sua utilização em processos decisórios desafiam a maneira como a discriminação tem sido geralmente compreendida (Angeli et al., 2016).

A tecnologia geralmente envolve a definição de um objetivo determinado. No entanto, nem sempre o programador é capaz de definir os objetivos de maneira completa e correta. Isso resulta em uma situação em que a presença de objetivos determinados, porém imperfeitos, pode levar a uma IA incontrolável que persiste incessantemente até alcançar seu objetivo. Essas questões ressaltam a necessidade de uma abordagem jurídica cuidadosa para garantir que a IA seja usada de maneira justa e responsável, protegendo os direitos e interesses de todos os indivíduos (Mantelero, 2022).

A União Europeia aprovou recentemente o "IA Act", enquanto o Congresso Nacional brasileiro propôs um Projeto de Lei¹³ para estabelecer um quadro normativo para a IA. Essas iniciativas visam criar uma estrutura baseada em uma abordagem de risco proporcional e centrada no ser humano. O objetivo é definir regras e mecanismos de controle e fiscalização para desenvolvedores, fornecedores, usuários e aqueles afetados pelos sistemas de IA. Essas propostas têm como finalidade promover o uso seguro de tecnologias de IA. Contudo, é importante que essas medidas sejam consideradas pontos de partida para um debate contínuo e

¹³ <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

não soluções definitivas. A participação ativa, o monitoramento constante e a revisão periódica são necessárias para garantir que a regulamentação acompanhe as evoluções da tecnologia e as necessidades da sociedade.

Por exemplo, no setor de saúde, a regulamentação pode exigir que os sistemas de IA utilizados para diagnósticos médicos sejam auditados regularmente para garantir sua precisão e imparcialidade, além de assegurar que os dados dos pacientes sejam protegidos e utilizados de maneira ética. Na educação, as diretrizes podem incluir a necessidade de transparência nos algoritmos que personalizam o aprendizado dos estudantes, para evitar vieses que possam prejudicar determinados grupos de alunos. No setor de transporte, regulamentações podem ser aplicadas para garantir que veículos autônomos passem por testes rigorosos de segurança antes de serem liberados para uso público, além de estabelecer protocolos em caso de falhas ou acidentes.

Finalmente, na área de comunicação, a regulamentação pode focar na prevenção de desinformação e no controle de conteúdo gerados por IA assegurando que as plataformas digitais adotem medidas eficazes para detectar e mitigar a propagação de notícias falsas.

Dada a natureza dinâmica, complexa e diversificada dos algoritmos e dos contextos em que são aplicados, é imperativo que essas discussões sejam adaptáveis. Afinal, a IA é um campo em constante evolução, e as políticas e regulamentos que a governam tem o difícil dever de serem capazes de acompanhar esse ritmo de mudança, para que possamos navegar com sucesso no futuro da IA equilibrando inovação com proteção e consideração pelos direitos e interesses de todos os envolvidos (Mantelero, 2022).

2. QUADRO NORMATIVO DE INTELIGÊNCIA ARTIFICIAL NO BRASIL

A regulamentação da IA no Brasil ainda está em fase inicial, mas já existem algumas iniciativas que visam a estabelecer princípios, diretrizes e normas para o seu uso ético, responsável e transparente. O Projeto de Lei nº 2338/2023 (Brasil, 2023) e a Resolução nº 271/2020 do Conselho Nacional de Justiça (Brasil, 2020) são exemplos de iniciativas que buscam regulamentar a IA no Brasil. A primeira visa orientar ações públicas e privadas relacionadas à IA promovendo pesquisa, inovação, educação e proteção de dados pessoais. A segunda institui o Comitê de Inteligência Artificial do Poder Judiciário para propor e avaliar soluções de IA nos tribunais, seguindo princípios como legalidade, transparência e ética.

Além dessas iniciativas, destaca-se a Portaria nº 4.617/2021 do Ministério da Ciência, Tecnologia e Inovações (Brasil, 2021), que aprova a Estratégia Brasileira de Inteligência

Artificial, e o Decreto Nº 11.260/2022 (Brasil 2022), que estabelece a Política Nacional de Governo Digital. Ambos têm como objetivo melhorar a prestação de serviços públicos e estimular a participação social através do uso de IA garantindo segurança da informação e proteção dos dados pessoais.

Essas iniciativas mostram o esforço do Brasil em regulamentar a IA visando aproveitar suas potencialidades e mitigar seus riscos, conforme os padrões internacionais. No entanto, ainda existem desafios e lacunas a serem resolvidos, como a harmonização das normas existentes, a capacitação de profissionais e cidadãos para lidar com a IA, além da fiscalização e responsabilização dos agentes que desenvolvem e utilizam a tecnologia. O Projeto de Lei nº 2338/2023 (Brasil, 2023) uma vez aprovado e promulgado, tem potencial para enfrentar esses desafios.

O Brasil, por sua vez, ainda não possui uma legislação específica e abrangente sobre a IA, mas sim um conjunto de normas setoriais e dispersas, que tratam de aspectos parciais ou relacionados à IA, como a Lei Geral de Proteção de Dados Pessoais, o Marco Legal das Startups, O Marco civil da Internet, O Código do Consumidor etc. Além disso, o Brasil tem adotado uma abordagem baseada na autorregulação e na cooperação multissetorial, buscando envolver os diversos atores interessados no debate e na construção de consensos sobre os princípios e as boas práticas para a IA. A Estratégia Brasileira de Inteligência Artificial (EBIA), por exemplo, não tem força vinculante, mas sim orientadora, visando a estimular a adoção voluntária de medidas que promovam uma IA ética, responsável e inclusiva no país.

Análise do Projeto de Lei 2338/2023

O legislador brasileiro propôs regulamentar o uso de IA com base os níveis de risco que esses sistemas de podem apresentar para os direitos e interesses das pessoas e da sociedade. Esses níveis de risco, que podem variar de acordo com o tipo, contexto, propósito e impacto da IA, devem ser considerados tanto na fase de desenvolvimento quanto na de uso dos sistemas de IA. Critérios como complexidade, transparência, explicabilidade, auditabilidade, robustez, segurança, privacidade, discriminação, autonomia, interação, afetação e reversibilidade podem ser utilizadas para avaliar os níveis de risco da IA.

O projeto em questão propõe em seus artigos iniciais, a criação de um marco legal para a IA, fundamentado em princípios éticos, de direitos humanos e de proteção de dados pessoais (Brasil, 2023) que no Artigo 4º do Projeto de Lei demonstra a importância de definir claramente os atores envolvidos no desenvolvimento e uso de Inteligência Artificial e define Inteligência

Artificial como “sistemas computacionais com diferentes graus de autonomia, projetados para alcançar objetivos utilizando aprendizado de máquina e/ou lógica e representação do conhecimento, processando dados de entrada de máquinas ou humanos”(Brasil, 2023).

Define também quem são os agentes, indicando que os fornecedores de IA podem ser indivíduos ou entidades, públicas ou privadas, responsáveis pelo desenvolvimento e comercialização desses sistemas, como por exemplo, uma grande empresa como *Google* ou *IBM*, ou uma *startup* especializada em tecnologias de aprendizado de máquina. Para o referido documento, os operadores de IA são pessoas físicas ou jurídicas que utilizam esses sistemas em benefício próprio, exceto para atividades pessoais não profissionais. Um exemplo de operador de IA pode ser um hospital que utiliza um sistema de diagnóstico médico ou uma empresa que utiliza um algoritmo para selecionar candidatos em processos de recrutamento.

A legislação trata de implementar uma forma de discriminação indireta, que ocorre quando práticas aparentemente neutras prejudicam grupos específicos, baseado em dados como raça, gênero e orientação sexual. Logo, com o avanço do aprendizado de máquina e algoritmos sofisticados, esses desafios se tornam complexos, pois a IA pode perpetuar preconceitos de forma silenciosas e despercebidas. A atenção do Projeto de Lei 2338/2023 com as formas de discriminação indireta demonstram um entendimento das nuances envolvendo essas tecnologias, assim, oferecendo soluções concretas para criar um ambiente tecnológico justo e inclusivo, garantindo que os sistemas de IA operem dentro de um *framework* ético e transparente.

Em seu Capítulo III o PL 2338/2023 (Brasil, 2023) categoriza os níveis de risco da IA em Excessivo e Alto. O nível Risco Excessivo é atribuído quando a IA tem potencial para causar danos significativos ou irreversíveis aos direitos e interesses das pessoas ou da sociedade, como vida, saúde, segurança, liberdade, dignidade, igualdade e democracia. O nível de Alto Risco é atribuído quando a IA pode causar danos moderados ou reversíveis aos direitos e interesses das pessoas ou da sociedade, como educação, cultura, comunicação, trabalho, consumo e meio ambiente.

Cabe destacar que o PL 2338 prevê, em seus Capítulos IV e VI, que os agentes de IA estabeleçam estruturas de governança e processos internos capazes de realizar uma avaliação prévia de conformidade, verificando o cumprimento dos requisitos legais e éticos no desenvolvimento e utilização dessas ferramentas. Essa abordagem é similar ao modelo existente na Lei Geral de Proteção de Dados (LGPD), que também enfatiza a autorregulação. Ambos os

marcos regulatórios propõem que indivíduos ou organizações ajustem seu comportamento ou operações de forma autônoma, mas seguindo diretrizes ou regras estabelecidas externamente.

A autorregulação oferece flexibilidade e adaptabilidade difíceis de alcançar com regulamentações formais, permitindo que a indústria responda rapidamente a inovações tecnológicas e desafios éticos. No entanto, ela precisa ser monitorada por mecanismos independentes para garantir eficácia e integridade. Garantindo que o desenvolvimento dessas novas tecnologias não seja completamente livre ou caótico, mas sim orientado pelos princípios, diretrizes e normas existentes visando alcançar determinados objetivos ou manter um certo equilíbrio. Muitas empresas e organizações desenvolvedoras de IA já implementam seus próprios códigos de conduta, diretrizes éticas e comitês de revisão para garantir que seus sistemas atendam aos mais altos padrões de segurança, transparência e imparcialidade. Essa abordagem proativa não apenas ajuda a mitigar riscos e evitar potenciais abusos, mas também promove a confiança pública na IA.

Comparativo com o Cenário Internacional

Um número crescente de países no mundo todo está elaborando e implementando leis e políticas de governança de IA. Embora os Estados Unidos (EUA) tenham inicialmente adotado uma abordagem leniente em relação à IA, os apelos para regulamentação têm aumentado recentemente. A Casa Branca lançou o Projeto para uma Declaração de Direitos da IA, um conjunto de diretrizes para proteger os direitos do público americano na era da IA. As agências governamentais chinesas aprovaram algumas diretrizes sobre IA generativa, enquanto o Reino Unido anunciou uma abordagem pró-inovação para a regulamentação da IA, que regula amplamente a IA por meio de leis existentes. Além disso, há a recém-criada parceria tecnológica entre o Conselho de Comércio e Tecnologia da UE e os EUA, que estão buscando desenvolver um entendimento mútuo sobre os princípios que sustentam uma IA confiável e responsável.

Uma forma de comparar a regulamentação da IA no Brasil com as de outros países e organizações internacionais é analisar os princípios e as diretrizes que orientam o desenvolvimento e o uso da IA em diferentes contextos. Nesse sentido, pode-se observar que há uma convergência em torno de alguns valores fundamentais, como o respeito aos direitos humanos, à democracia, ao Estado de Direito, à inovação e à transparência. Esses valores são expressos, por exemplo, no *AI Act* aprovado pela Comissão Europeia (EU, 2023) , na

Comunicação da Comissão Europeia sobre a IA de confiança¹⁴, na Declaração de Montreal para uma IA Responsável¹⁵, na Norma ISO/IEC 38507¹⁶ e na própria Estratégia Brasileira de Ética para Inteligência Artificial¹⁷

No entanto, também existem diferenças e especificidades na forma como cada país e organização internacional aborda a regulamentação da IA levando em conta seus contextos históricos, culturais, políticos, econômicos e sociais. Por exemplo, a União Europeia tem se destacado por sua abordagem baseada nos direitos fundamentais e na proteção dos dados pessoais, buscando estabelecer um quadro jurídico comum e harmonizado para a IA, que garanta a sua conformidade com os valores europeus e as regras do mercado interno. A proposta de regulamento sobre a IA aprovada pela Comissão Europeia prevê uma classificação dos sistemas de IA de acordo com seu nível de risco, impondo obrigações e restrições proporcionais aos provedores e aos usuários de tais sistemas, bem como mecanismos de supervisão, fiscalização e sanção.¹⁸

Ciente da importância de se estabelecer mecanismos de avaliação e de governança dos sistemas de IA levando em conta seus possíveis impactos e implicações, a União Europeia aprovou a primeira lei de regulamentação de IA. O *AI Act* aprovado pela Comissão Europeia, visa estabelecer um quadro jurídico harmonizado para a IA na União Europeia, baseado em uma abordagem de risco proporcionado e centrado no ser humano.

De acordo com a legislação europeia, os sistemas de IA são classificados em quatro níveis de risco: inaceitável, alto, limitado e mínimo.

Os sistemas considerados inaceitáveis são aqueles que violam os valores e os princípios fundamentais da União Europeia, como a dignidade humana, a democracia, o Estado de direito e os direitos fundamentais, e devem ser proibidos. Exemplos incluem sistemas de vigilância

¹⁴ UNIÃO EUROPEIA. *Excellence and Trust in Artificial Intelligence*. Disponível em: <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_pt#:~:text=Proposto%20pela%20Comiss%C3%A3o%20em%20abril,de%20uma%20IA%20inovadora%20e>. Acesso em: 3 nov. 2024.

¹⁵ UDEM. *Declaração de Montreal para uma Inteligência Artificial Responsável*. Disponível em: <https://www.sbmec.org.br/wp-content/uploads/2021/02/Portugue%CC%82s-UdeM_Decl-IA-Resp_LA-Declaration_vf.pdf>. Acesso em: 3 nov. 2024.

¹⁶ ISO/IEC. *Governance of IT – Implementation Guide – ISO/IEC 38507:2020*. Disponível em: <https://webstore.iec.ch/preview/info_isoiec38507%7Bed1.0%7Den.pdf>. Acesso em: 3 nov. 2024.

¹⁷ BRASIL. Ministério da Ciência, Tecnologia e Inovações. *Inteligência Artificial - Estratégia Eixo 1*. Disponível em: <<https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial-estrategia-eixo1>>. Acesso em: 3 nov. 2024.

¹⁸ [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

biométrica em tempo real em espaços públicos e sistemas que manipulam o comportamento humano de forma maliciosa, como brinquedos que incitam comportamentos perigosos em crianças.¹⁹

Já os de alto risco são aqueles que afetam direta ou indiretamente os direitos ou as obrigações das pessoas ou das organizações, ou que implicam uma intervenção significativa nas suas vidas ou atividades, e devem estar sujeitos a requisitos obrigatórios de qualidade, segurança, transparência, supervisão humana, entre outros. Por exemplo, sistemas de IA usados em processos de recrutamento que podem influenciar significativamente a vida profissional de uma pessoa, ou sistemas de diagnóstico médico que têm um impacto direto na saúde dos pacientes.

Já os sistemas de IA considerados de risco limitado são aqueles que interagem com as pessoas de forma a influenciar seu comportamento, suas escolhas ou suas emoções, e devem fornecer informações adequadas sobre sua natureza, seus objetivos e seus efeitos. Por exemplo, assistentes virtuais que recomendam produtos com base no histórico de compras do usuário, ou algoritmos que sugerem conteúdos em plataformas de streaming devem ser transparentes quanto à forma como operam e os dados que utilizam.

Por último, os sistemas de IA considerados de risco mínimo são aqueles que representam um baixo ou nenhum risco para os direitos ou interesses das pessoas ou das organizações, e devem estar isentos de qualquer obrigação específica. Exemplos incluem *chatbots* que fornecem respostas automáticas a perguntas frequentes em sites de serviço ao cliente, ou sistemas de recomendação de filmes baseados nas preferências do usuário que não envolvem decisões críticas.

O *AI Act* aprovado pela Comissão Europeia é um marco importante na regulação da IA na Europa e no mundo. Ele busca equilibrar o desenvolvimento e a inovação tecnológica com o respeito aos direitos humanos e aos valores democráticos. No entanto, o regulamento enfrenta desafios práticos e políticos para sua implementação e harmonização, considerando a diversidade e complexidade dos sistemas de IA, bem como os interesses e perspectivas dos diferentes Estados-Membros e setores econômicos e sociais afetados pela IA.

¹⁹ <https://d3.harvard.edu/platform-digit/submission/hello-barbie-ai-making-childrens-dreams-come-true/>

2. AUTORREGULAÇÃO E O PAPEL DOS COMITÊS DE *EXPERTS* NO DESENVOLVIMENTO E UTILIZAÇÃO DE IA

2.1 Breves considerações sobre Autorregulação

A autorregulação é um conceito central na governança moderna, especialmente relevante para áreas em constante evolução como a tecnologia e a IA. Este mecanismo refere-se ao processo pelo qual determinados setores da sociedade estabelecem normas e regulamentos internos para governar suas próprias atividades, sem a necessidade de intervenção direta do Estado e permite que profissionais do direito e entidades privadas criem suas próprias regras e princípios éticos, buscando aprimorar a qualidade dos serviços prestados e garantir confiança aos seus clientes (Hoffmann-Riem, 2019).

2.1.1 Características da Autorregulação:

- a) Autonomia: as entidades autorreguladoras têm a liberdade de definir suas próprias normas e procedimentos;
- b) Responsabilidade: as entidades são responsáveis por garantir a conformidade com as normas legais estabelecidas;
- c) Transparência: processos e decisões devem ser claros e acessíveis ao público;
- d) Flexibilidade: As normas podem ser ajustadas conforme necessário para atender às mudanças no setor.

2.1.2 Exemplos de Autorregulação:

- ✓ Conselho Nacional de Autorregulamentação Publicitária (CONAR): no Brasil, o CONAR é um exemplo de autorregulação no setor publicitário, onde empresas se comprometem a seguir normas de publicidade ética;
- ✓ Lei Geral de Proteção de Dados (LGPD): permite que controladores e operadores de dados pessoais formulem regimentos internos de gestão de dados, estabelecendo práticas que atendam às diretrizes normativas.

Assim, a prática da autorregulação é caracterizada pela autonomia das entidades que definem suas próprias normas e procedimentos. Essas entidades são responsáveis por assegurar a conformidade com as normas estabelecidas, promovendo processos transparentes e acessíveis ao público. A flexibilidade inerente à autorregulação permite ajustes conforme necessário, atendendo às mudanças e necessidades do setor regulado.

2.2 Uma Análise sobre o Papel dos Comitês de Experts no Desenvolvimento e IA

Partindo do pressuposto que o PL brasileiro será aprovado sem grandes modificações, será feito um estudo de caso sobre o papel crucial dos comitês de experts nas organizações, especialmente no âmbito da autorregulação. Esses comitês são responsáveis por formular diretrizes que garantem o desenvolvimento ético e a implementação responsável das tecnologias de IA. Compostos por especialistas de diferentes setores, esses comitês são fundamentais para a criação e a manutenção de *frameworks* de governança de IA promovendo a conformidade com normas e diretrizes estabelecidas e ajustando-as conforme necessário para atender às mudanças no setor.

O potencial dos Conselhos de Especialistas tem sido amplamente reconhecido por empresas de desenvolvimento de IA resultando na formação de comitês para fornecer orientação sobre os desafios do uso de sistemas intensivos em dados. Esses painéis, conhecidos como conselhos éticos, compartilham características essenciais para avaliar a eficácia e a adequação da integração de especialistas no processo de design e desenvolvimento de IA (Mantelero, 2022).

Um estudo de caso bem documentado é o *Facebook Oversight Board*, um conselho de supervisão criado pelo *Facebook* para promover a liberdade de expressão por meio de decisões independentes e baseadas em princípios sobre conteúdo no *Facebook* e *Instagram* e emitir recomendações sobre a Política de Conteúdo da Empresa. Para alcançar esse objetivo, o *Oversight Board* revisa um número seletivo de casos altamente emblemáticos, determinando se as decisões foram tomadas de acordo com os valores e políticas declarados do *Facebook* e emite decisões vinculativas.

Contudo, duas limitações que precisam ser consideradas. Primeiramente, um órgão supervisor criado internamente por uma empresa seleciona casos emblemáticos especificamente para destacar as fragilidades dos serviços ou produtos da própria companhia. Foi o que aconteceu com o conselho, onde os casos escolhidos abordaram questões cruciais de transparência na moderação de conteúdo e critérios, a quantidade de recursos disponíveis para essa moderação, a harmonização dos padrões de autorregulação da empresa, o papel da intervenção humana e a precisão dos sistemas automatizados de moderação de conteúdo (Mantelero, 2022).

A discussão sobre a eficácia dos comitês de expert e supervisão, como no caso do *Facebook*, revelam dois pontos importantes. Primeiro, comitês internos selecionam casos emblemáticos para destacar as fraquezas dos produtos ou serviços da empresa, que aborda

questões críticas de transparência na moderação de conteúdo, recursos disponíveis, harmonização de padrões de autorregulação, intervenção humana e precisão dos sistemas automatizados.

Essas decisões, que têm um impacto amplo na estrutura do produto ou serviço da empresa, levantam a questão sobre a eficácia de tais comitês. Poderia um processo de auditoria independente alcançar os mesmos resultados, abordando problemas de maneira abrangente e imparcial?

Nesse contexto, surgem questões relacionadas à transparência e harmonização, que poderiam ser adequadamente identificadas e analisadas por auditores verdadeiramente independentes. Eles revisariam o processo de tomada de decisões de moderação de conteúdo e dispensariam a necessidade de um método específico para cada caso, focando em uma revisão ampla e independente das práticas da empresa (Mantelero, 2022).

Em segundo lugar, o Conselho executa uma espécie de verificação de conformidade. Analisando a aplicação da autorregulação do *Facebook* em cada situação, não levando em conta o impacto geral e amplamente discutido da rede social e suas políticas. Essa limitação se torna ainda mais evidente, pois a competência do conselho se restringe ao conteúdo removido e não abrange todo o serviço como publicidade comportamental e exploração de dados pessoais, por exemplo.

Em vista disso, é difícil considerá-lo o modelo ideal para um comitê que visa incorporar valores sociais e éticos no design de sistemas de IA. O Conselho do *Facebook* não se concentra no impacto geral da aplicação, avaliando apenas seu funcionamento e sugerindo melhorias procedimentais. Este comitê não questiona o design da rede social baseada em IA nem seu impacto mais amplo sobre indivíduos e a sociedade (Mantelero, 2022).

Outro caso interessante é o da *Axon AI Ethics Board*, que tem um foco maior no design de produtos e serviços. Enquanto o *Oversight Board* revisa apenas uma parte específica dos produtos e serviços do *Facebook*, lidando com a moderação de conteúdo em casos polêmicos, a missão do *Axon Ethics Board* é fornecer orientação especializada à *Axon* sobre o desenvolvimento de seus produtos e serviços de IA, com atenção especial ao impacto nas comunidades.

O Conselho da *Axon* foi criado para fornecer orientação específica sobre produtos e serviços, com foco particular no design durante a fase de desenvolvimento. A empresa compromete-se a disponibilizar ao conselho informações detalhadas sobre a lógica utilizada na

criação de seus algoritmos, os dados usados para treinar os modelos e os insumos aplicados, além de explicar as medidas adotadas para mitigar efeitos adversos, como viés e uso indevido. Isso está alinhado com o "*Product Evaluation Framework da Axon*", que é uma avaliação de risco focada nos principais aspectos a serem considerados na avaliação dos benefícios e custos sociais durante o desenvolvimento do produto (Mantelero, 2022).

As principais preocupações, considerando a área de atuação da *Axon*, abrangem o uso inadequado da tecnologia, a criminalização de indivíduos, o tratamento de dados pessoais, possíveis preconceitos e transparência. Além disso, incluem categorias mais amplas, como "violação de direitos constitucionais ou outros direitos legais" e "potenciais custos sociais", o que pode expandir a análise. A autoavaliação é realizada pela empresa durante a fase de desenvolvimento e, posteriormente, revisada pelo Conselho de Ética.

Enfatiza-se, um ponto negativo que diz respeito à interação entre a empresa e o Conselho de Ética é apenas parcialmente documentada nos relatórios de atividades, o que limita a transparência. Contudo, as informações apresentadas mostram um diálogo contínuo sobre aplicações tecnológicas, com a empresa aceitando parcialmente as recomendações do Conselho, resultando em mudanças nos produtos e serviços. Uma questão crucial é o acesso completo às informações confidenciais dos produtos, considerando que os membros do Conselho não pertencem à empresa. Para isso, eles assinaram acordos de confidencialidade, que abrangem segredos comerciais e dados proprietários. Embora esses *Non-Disclosure Agreement* NDAs facilitem o diálogo e não interfiram na atividade do Conselho, levantam preocupações sobre a eficácia da interação com comunidades e partes interessadas potencialmente impactadas (Mantelero, 2022).

Além disso, a *Axon* designou dois membros para o Comitê de Especialistas: um funcionário da *Axon* que faz parte do Conselho de Ética e trabalha fora da cadeia de comando interna, e um membro do Conselho de Ética que não pertence à empresa. Esses conselheiros podem ser procurados por funcionários que tenham preocupações em casos específicos.

Semelhante ao Conselho de Supervisão do *Facebook*, o Conselho de Ética da *Axon* não adota uma abordagem participativa que inclua as vozes das comunidades e grupos potencialmente afetados. Embora as decisões do Conselho de Supervisão do *Facebook* sejam vinculativas para a empresa para as questões relacionadas a moderação de conteúdo, o Conselho de Ética da *Axon* apenas pode fornecer recomendações, assim como o Conselho de Supervisão

em relação à Política de Conteúdo do *Facebook*. Isso sugere que interesses comerciais podem facilmente prevalecer sobre quaisquer preocupações éticas levantadas pelo Conselho de Ética.

Embora possuam diferenças, tanto o *Facebook* quanto a *Axon* estabeleceram conselhos de experts que influenciaram o design ou o uso de produtos e serviços. Além disso, esses impactos foram devidamente documentados em seus relatórios ou decisões. Observa-se ainda a existência de diferentes modelos de Conselho Consultivo de Ética em IA. Em 2021, a *Arena Analytics*, uma empresa especializada em análise preditiva e aprendizado de máquina para processos de contratação, criou um conselho consultivo. Este conselho é composto por especialistas da academia, tecnologia, recursos humanos e ética, com o objetivo de desenvolver orientações que ajudem a *Arena* a gerenciar suas obrigações éticas. No entanto, não foram documentados resultados concretos ou impactos no modelo de negócios, produtos ou serviços, assim como os procedimentos envolvidos na seleção e no trabalho do conselho.

De maneira similar, a *SAP*, empresa de *software*, estabeleceu um Comitê Consultivo de Ética em IA, composto por acadêmicos, especialistas em políticas e profissionais da indústria, com o objetivo de aconselhar a empresa no desenvolvimento e implementação de seus princípios orientadores de IA. Este grupo externo interage com um Comitê de Supervisão de Ética em IA interno, formado por executivos da empresa de todas as áreas do conselho com supervisão de tópicos relevantes para orientar e implementar a Ética em IA, fornecendo orientação às equipes da empresa sobre como casos específicos são impactados por esses princípios.

O modelo da *SAP* inclui uma unidade interna focada em questões éticas, chamada Comitê de Supervisão de Ética em IA, que conta com a figura do *Chief Ethics Officer*. Este oficial tem um papel ampliado pela participação nas decisões que lidam com questões éticas. O Comitê de Supervisão de Ética em IA não substitui o *Chief Ethics Officer*, pois este comitê, centrado na posição executiva em uma área específica, não está necessariamente relacionado a um mandato ético. Este papel também pode ser desempenhado por um órgão externo, como o Painel Consultivo de Ética em IA da *SAP*, mas neste caso, o painel parece fornecer à empresa conselhos mais gerais sobre o desenvolvimento e operacionalização dos princípios éticos, em vez de orientações específicas para casos (Mantelero, 2022).

Em todos os casos, a falta de informações sobre o funcionamento desses órgãos ou a documentação de seu trabalho limita necessariamente nossa avaliação de sua eficácia na implementação de valores éticos nas práticas e produtos/serviços das empresas.

Uma outra abordagem adotada por empresas é a criação de conselhos de ética sem especificar objetivos concretos, composição ou procedimentos. Um exemplo é o Conselho de Ética em IA interno da IBM, que é composto por uma equipe interdisciplinar de executivos seniores da IBM, co-presidido pelo *Chief Privacy Officer* e pelo *AI Ethics Global Leader* da IBM, e reporta-se aos níveis mais altos da empresa. No entanto, a lista de seus membros não está disponível publicamente e seu impacto concreto na estratégia da empresa não é detalhado.

Com os estudos de caso envolvendo alguns dos principais atores de IA, é possível agrupar os conselhos de ética em três categorias distintas. A primeira categoria engloba conselhos que desempenham um papel ativo nos negócios das empresas, com membros nomeados cuja identidade é pública, procedimentos internos implementados, tarefas claramente definidas e um compromisso explícito da empresa em considerar os pareceres dos conselhos. A segunda categoria abrange conselhos cujas tarefas e membros são conhecidos, contudo, a interação concreta e o impacto nas decisões empresariais não são devidamente documentados. Finalmente, a terceira categoria inclui conselhos cuja identidade dos membros permanece desconhecida e que apenas possuem uma descrição geral e vaga do seu propósito principal (Mantelero, 2022).

Os Comitês de *experts* em IA nas corporações demonstram uma variedade de estruturas, incluindo órgãos internos e externos, os quais produzem diretrizes gerais até questionários de autoavaliação ética. A independência e a reputação dos membros são cruciais, e a transparência sobre a estrutura e o funcionamento desses conselhos é necessária para comprovar seu impacto nas decisões empresariais. A eficácia dos conselhos pode ser influenciada por decisões relacionadas ao pessoal e ao acesso às informações, e é fundamental promover a responsabilidade nas empresas em relação às recomendações dos conselhos. Apenas em casos limitados, como o do *Facebook*, as decisões desses conselhos são obrigatórias.

Além disso, é notável que os valores orientadores desses conselhos frequentemente se refiram a direitos humanos e liberdades fundamentais, enquanto as empresas especificam seus próprios princípios e valores para guiar as decisões do. Assim, a presença de conselhos consultivos externos pode complementar o trabalho desses órgãos. Contudo, a falta de informações públicas sobre a identidade dos membros e os procedimentos dos conselhos limita a avaliação de sua eficácia em implementar valores éticos nas práticas e produtos das empresas.

Desta forma, é possível concluir que não existe um modelo uniforme para conselhos de ética corporativa em sistemas de IA; em vez disso, há uma gama diversificada de abordagens.

No entanto, as diversas evidências destacadas nos estudos de caso oferecem importantes insights, identificando elementos essenciais à supervisão ética da IA, dentre eles destacam-se a independência e credibilidade do conselho, orientação sólida baseada em valores, eficácia operacional, transparência nas deliberações e responsabilidade das decisões tomadas.

Assim sendo, a melhor prática para a implementação eficaz de conselhos de ética em IA nas corporações parece depender da combinação de diferentes abordagens. Uma estrutura híbrida, que integra conselhos internos e externos, pode proporcionar uma supervisão mais abrangente e equilibrada. A combinação de conhecimentos especializados de diversos campos e a inclusão de representantes independentes aumentam a credibilidade e a aceitação das recomendações feitas. Adicionalmente, a transparência nas operações desses conselhos e a documentação pública de suas atividades são essenciais para garantir a responsabilidade e a confiança do público nas práticas empresariais relacionadas à IA.

Referências como a *Arena Analytics* e a *SAP* exemplificam como a integração de conselhos consultivos pode ser moldada para atender às necessidades específicas da organização e de seu ambiente regulatório. Ao mesmo tempo, a *IBM* destaca a importância de uma liderança ética forte e de uma equipe interdisciplinar dedicada a essa função. Em todos os casos, a clareza e a divulgação das diretrizes e resultados desses conselhos são vitais para avaliar seu impacto real e promover uma cultura de responsabilidade ética.

Portanto, os estudos de caso revelam que a eficácia dos conselhos de ética em IA nas corporações não depende apenas de sua existência, mas também de sua estrutura, transparência e integração com as práticas de negócios da empresa. A contínua evolução e adaptação desses conselhos são indispensáveis para acompanhar o rápido desenvolvimento da IA e assegurar que os valores éticos permaneçam centrais na inovação tecnológica das corporações.

CONSIDERAÇÕES FINAIS

Na primeira parte do estudo, examinou-se o desenvolvimento e o impacto da IA e das tecnologias relacionadas, como aprendizado de máquina e big data. A evolução da IA revelou seu potencial para executar tarefas complexas e antes consideradas exclusivas da inteligência humana, desafiando conceitos tradicionais de capacidade e raciocínio. A definição abrangente de sistemas de IA proposta pelo Grupo de Peritos de Alto Nível da Comissão Europeia destaca a importância de perceber, interpretar e agir com base em dados para alcançar objetivos específicos. No entanto, essa evolução também trouxe desafios significativos, especialmente no

contexto político e jurídico, em áreas como privacidade de dados, direitos autorais e responsabilidade civil.

Em seguida, a atenção se voltou para a autorregulação e a formação de comitês de especialistas com o objetivo de assegurar o desenvolvimento ético e responsável da IA. Foram discutidos exemplos como o *Facebook Oversight Board* e o *Axon AI Ethics Board*, ressaltando suas funções e limitações. Observamos que os conselhos podem desempenhar um papel vital na orientação das empresas sobre questões éticas e na implementação de práticas que promovam a confiança pública na IA. A transparência, a independência e a credibilidade dos membros são fundamentais para a eficácia desses conselhos. No entanto, a falta de documentação pública sobre suas atividades e a interação limitada com as comunidades afetadas constituem desafios significativos para a avaliação do seu impacto.

A eficácia dos conselhos de ética depende de sua estrutura, transparência e integração com as práticas empresariais. Uma abordagem híbrida, que combine conselhos internos e externos, parece ser a melhor prática para proporcionar uma supervisão mais abrangente e equilibrada. A contínua evolução e adaptação dessas estruturas são indispensáveis para acompanhar o rápido desenvolvimento da IA assegurando que os valores éticos permaneçam centrais na inovação tecnológica. Dessa forma, é possível promover um uso da IA que seja seguro, responsável e benéfico para toda a sociedade.

REFERÊNCIAS

ANGELI, P.H.; COLODETTE, L.; OLIVEIRA, P.H.S.; SILVA, A.B. A evolução da inteligência artificial e a substituição do trabalho humano. **Rev. Ambiente Acadêmico**, 2019; v.5, n.1. Disponível em: <https://multivix.edu.br/wp-content/uploads/2019/10/revista-ambiente-academico-v05-n01-artigo01.pdf>. Acesso em: 20 setembro de 2024.

BECKHAUSER, L.A. **Revolucionando a Prática Jurídica: O Papel Imperativo da Inteligência Artificial na Advocacia**. 2023.

BRAGA, Carolina. **Discriminação nas Decisões por Algoritmos: Polícia Preditiva** In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade**. São Paulo (SP): Editora Revista dos Tribunais. 2020. Disponível em: <https://www.jusbrasil.com.br/doutrina/inteligencia-artificial-e-direito-etica-regulacao-e-responsabilidade/1196969611>. Acesso em: 4 de Novembro de 2024.

BRASIL. Código Civil Brasileiro. **Lei nº 10.406, de 10 de janeiro de 2002**. Diário Oficial da União, Brasília, DF, 11 jan. 2002.

BRASIL. **Lei nº 12.965**, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União, Brasília, DF, 24 abr. 2014.

BRASIL. **Lei nº 13.709**, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da União, Brasília, DF, 15 ago. 2018.

BRASIL. Conselho Nacional de Justiça. Resolução nº 271, de 13 de outubro de 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3613>>. Acesso em: 3 nov. 2024.

BRASIL. **Lei Complementar nº 182**, de 1º de junho de 2021. Institui o marco legal das *startups* e do empreendedorismo inovador; e altera a Lei nº 6.404, de 15 de dezembro de 1976, e a Lei Complementar nº 123, de 14 de dezembro de 2006. Diário Oficial da União, Brasília, DF, 2 jun. 2021.

BRASIL. SENADO FEDERAL **Projeto de Lei nº 2338**, de 2023. Dispõe sobre o uso da Inteligência Artificial. Senado Federal, Brasília, DF, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em 26 de set 2024.

DIGNUM, V. Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, Cham: **Springer International Publishing**, 2019

FIA – *Business School*. **Inteligência Artificial**: o que é, como funciona e exemplos. 2023. Disponível em: <https://fia.com.br/blog/inteligencia-artificial-nas-empresas/#:~:text=A%20intelig%C3%A2ncia%20artificial%20nas%20empresas,de%20tecnologia%20em%20suas%20atividades>. Acesso em: 17 mai. 2024.

FRAZÃO, Ana. Dados, estatísticas e algoritmos: perspectivas e riscos da sua recente utilização. **Jota**, 2017. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial>. Acesso em: 3 de nov. 2024

HOFFMANN-RIEM, W. Autorregulação, autorregulamentação e autorregulamentação regulamentada no contexto digital. **Revista da AJURIS**, 2019; v. 46, n. 146, p. 529–554. Disponível em: <https://revistadaajuris.ajuris.org.br/index.php/REVAJURIS/article/view/1048>. Acesso em: 4 nov. 2024.

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. **A Definition of AI: Main Capabilities and Disciplines**: definition developed for the purpose of the AI HLEG’s Deliverables. *Brussels: European Commission*, 2019. Acesso em: 3 nov. 2024.

KAUFMAN, D. Entrevista com Davi Geiger. **Teccogs: Revista Digital de Tecnologias Cognitivas**, 2018; n. 17, p. 10-15

JONAS, V. **Inteligência artificial e o impacto nos empregos e profissões**. 2020. <https://agenciabrasil.ebc.com.br/geral/noticia/2020-08/inteligencia-artificial-e-o-impacto-nos-empregos-e-profissoes>. Acesso em 03 nov. 2024

LARSON, Stefan. On the Governance of Artificial Intelligence through Ethics Guidelines. **Asian Journal of Law and Society**, v. 7, p. 437-451, 2020.

MACHADO, C.M. **Novas tecnologias para uma velha advocacia**. 2021. Disponível em: <https://jus.com.br/artigos/92220/novas-tecnologias-para-uma-velha-advocacia>. Acesso em: 07 de jul. 2024.

MANTELERO, A. *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*: 36. Editora: T.M.C. **Asser Press**; 2022.

O'NEIL, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: **Random House Audio**, 2016.

QUARESMA, A. Inteligência artificial fraca e força bruta computacional. **TECHNO Review**, 2021; n. 1, v.10, n.1, Disponível em: <https://dialnet.unirioja.es/descarga/articulo/9199926.pdf>. Acesso em: 3 de nov. 2024

ROSA, A.M. **Guia do Processo Penal conforme a Teoria dos Jogos**. Florianópolis: E Mais, 2019.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2013. p. 1174.

UNIÃO EUROPEIA. PARLAMENTO EUROPEU E DO CONSELHO. **Regulamento (UE) nº 2024/1689**, de 13 de junho de 2024. Cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (EU). 2024.