**UNIVERSIDADE DE BRASÍLIA**
**FACULDADE DE DIREITO**

**CLEIDEANE DOS SANTOS FARIAS**

# INTELIGÊNCIA ARTIFICIAL, ÉTICA E DIREITO: possíveis contribuições internacionais para o desenvolvimento responsável da IA no Brasil.

Brasília
2024

Universidade de Brasília
Faculdade de Direito

DIREITO

**CLEIDEANE DOS SANTOS FARIAS**

# INTELIGÊNCIA ARTIFICIAL, ÉTICA E DIREITO: possíveis contribuições internacionais para o desenvolvimento responsável da IA no Brasil.

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Direito da Universidade de Brasília, para a obtenção do título de Mestre em Direito.

**Orientador:** Prof. Dr. Fabiano Hartmann Peixoto

Brasília
2024

**Ficha Catalográfica**

Ficha catalográfica elaborada automaticamente, com os dados fornecidos pelo(a) autor(a)


Farias, Cleideane dos Santos
F224i INTELIGÊNCIA ARTIFICIAL, ÉTICA E DIREITO: possíveis contribuições internacionais para o desenvolvimento responsável da IA no Brasil. / Cleideane dos Santos Farias; orientador Fabiano Hartmann Peixoto. -- Brasília, 2024.
235 p.

Dissertação(Mestrado em Direito) -- Universidade de Brasília, 2024.

1. Inteligência Artificial. 2. Governança. 3. Responsável. 4. Legislações Internacionais. 5. Regulamentação. I. Hartmann Peixoto, Fabiano , orient. II. Título.

## FOLHA DE APROVAÇÃO

Nome do autor: FARIAS, Cleideane dos Santos

Título: INTELIGÊNCIA ARTIFICIAL, ÉTICA E DIREITO: possíveis contribuições internacionais para o desenvolvimento responsável da IA no Brasil.

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Direito da Universidade de Brasília, para a obtenção do título de Mestre em Direito.

Apresentada em: 25 de outubro de 2024

Banca Examinadora:

| | |
|---|---|
| Prof. Dr. Fabiano Hartmann Peixoto | Instituição: Universidade de Brasília (UnB) |
| Julgamento: | APROVADA |
| Profa. Dra. Ana Luisa Tarter Nunes | Instituição: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP) |
| Julgamento: | APROVADA |
| Profa. Dra. Cristina Mendes Bertoncini Corrêa | Instituição: Universidade Federal de Santa Catarina (UFSC) |
| Julgamento: | APROVADA |

## DEDICATÓRIA

*Minha pequena Liz, minha filha querida,*
*razão do meu empoderamento diário, da minha vida,*
*a você dedico este trabalho com todo meu amor,*
*para que você seja uma mulher extraordinária,*
*generosa, competente e feliz.*

*A educação é o desenvolvimento no homem*
*de toda a perfeição de que sua natureza é capaz.*

*Immanuel Kant'*

## AGRADECIMENTOS

Ao refletir sobre o dom da vida e a sua brevidade nesta terra, faz-se necessário agradecer ao Nosso Redentor, o Cristo, que nos guarda em sua infinita misericórdia, graça e nos sustenta a cada dia.

Agradeço ao meu marido, Elvio, que sempre me impulsionou para obter conhecimento relacionado a tecnologia e pela parceria. Aos meus pais, Cleide Farias e Jorge Farias, que nunca mediram esforços em nenhuma esfera de vida para que eu fosse atrás dos meus sonhos e de conhecimento; aos meus irmãos pela amizade; e principalmente a minha filha, Liz Santos Sousa, com quase dois anos de vida, razão do meu enfrentamento diário pelo conhecimento, trabalho e empoderamento feminino.

Agradeço nominalmente aos meus amigos que me incentivaram durante a caminhada do mestrado Janay Leandro, Carolina Rabelo e Marcos Leitão. Amigos são anjos vindos do céu, eu tenho certeza disto.

Na categoria de mestre, agradeço ao meu orientador, Prof. Dr. Fabiano Hartmann Peixoto que com toda a sua generosidade, simplicidade, educação e conhecimento tem me ajudado a alcançar o tão sonhado Mestrado em Direito. Ser orientada pelo professor Fabiano foi a realização de um grande sonho, pois sei que poucos profissionais do Direito têm a bagagem e experiência do direito aplicado à tecnologia que o meu orientador possui.

Agradeço a todos os professores do mestrado em Direito da UNB que ensinaram tanto sobre inclusão, diversidade, política, empoderamento feminino e sobre pensar  o Direito de uma forma maior e mais inclusiva. A Universidade de Brasília forma profissionais com uma abordagem realista e séria para o enfrentamento de problemas da nossa sociedade.

Agradeço aos membros da banca por fazerem parte deste momento tão marcante de concretização da minha produção intelectual.

## RESUMO

A inteligência artificial é uma tecnologia emergente e disruptiva que exige uma abordagem multifacetada que considere não apenas os aspectos técnicos, mas também os impactos sociais, econômicos e legais. Esta pesquisa estuda como se dá a aplicação da ética e do direito no desenvolvimento responsável de sistemas de inteligência artificial. Para tanto é feito um estudo comparativo das contribuições internacionais legislativas e de iniciativas políticas da União Europeia, Estados Unidos, Canadá e Brasil para o enfrentamento responsável, racional e gerador de valor agregado para as economias. O estudo destaca como diferentes jurisdições abordam a regulamentação da inteligência artificial, analisando a governança, melhores práticas e identificando lacunas a serem preenchidas para garantir o desenvolvimento de maneira ética e responsável, promovendo inovação sem comprometer os direitos fundamentais, privacidade e a segurança dos indivíduos. Pelo método dedutivo e bibliográfico de pesquisa, a dissertação é desenvolvida para trazer à tona discussões sobre a ontologia e deontologia da inteligência artificial e os assuntos que lhe são permeáveis, como as aplicações de direito e ética, e a explicabilidade desses processos. Numa forma complementar, apresentou-se as principais estratégias e iniciativas legislativas da União Europeia, Estados Unidos e Canadá a fim de contribuir para o desenvolvimento normativo da Inteligência Artificial, enquanto que o arcabouço de estratégias e iniciativas legislativas do Brasil teve uma análise ainda mais específica. Para sintetizar o estudo foi apresentado um quadro comparativo de semelhanças e contribuições de iniciativas entre Brasil e estes países internacionais. O objetivo é fornecer uma análise abrangente e crítica das abordagens legislativas e políticas, para buscar propor recomendações para o aprimoramento do arcabouço regulatório brasileiro. Dessa forma, a pesquisa pretende contribuir para o debate global sobre a governança da inteligência artificial, oferecendo ideias que possam guiar políticas públicas e práticas jurídicas capazes de equilibrar a inovação tecnológica com a proteção dos direitos humanos e a promoção do bem-estar social.



**Palavras-chave:** Inteligência Artificial; Responsável; Governança; Legislações Internacionais; Regulamentação.

## ABSTRACT

Artificial intelligence is an emerging and disruptive technology that requires a multifaceted approach that considers not only technical aspects, but also social, economic, and legal impacts. This research studies how ethics and law are applied in the responsible development of artificial intelligence systems. To this end, a comparative study is made of the international legislative contributions and political initiatives of the European Union, the United States, Canada, and Brazil for the responsible, rational, and value-added approach to the issue for economies. The study highlights how different jurisdictions approach the regulation of artificial intelligence, analyzing governance, best practices, and identifying gaps to be filled to ensure development in an ethical and responsible manner, promoting innovation without compromising fundamental rights, privacy, and the security of individuals. Using a deductive and bibliographical research method, the dissertation is developed to bring up discussions on the ontology and deontology of artificial intelligence and the issues that permeate it, such as the applications of law and ethics, and the explainability of these processes. In a complementary way, the main legislative strategies and initiatives of the European Union, the United States and Canada were presented in order to contribute to the normative development of Artificial Intelligence, while the framework of legislative strategies and initiatives of Brazil was analyzed even more specifically. To summarize the study, a comparative table of similarities and contributions of initiatives between Brazil and these international countries was presented. The objective is to provide a comprehensive and critical analysis of legislative and political approaches, in order to seek to propose recommendations for the improvement of the Brazilian regulatory framework. In this way, the research intends to contribute to the global debate on the governance of artificial intelligence, offering ideas that can guide public policies and legal practices capable of balancing technological innovation with the protection of human rights and the promotion of social well-being.



**Keywords**: Artificial Intelligence; Responsible; Governance; International Legislation; Regulation

## LISTA DE QUADROS

## LISTA DE SIGLAS E ABREVIATURAS

| | |
|---|---|
| ADM | Decisão Automatizada por Máquina |
| CE | Comissão Europeia |
| CPPA | Consumer Privacy Protection Act |
| | Lei de Proteção à Privacidade do Consumidor |
| EBIA | Estratégia Brasileira de Inteligência Artificial |
| FTC | Federal Trade Commission |
| GDPR | *General Data Protection Regulation* |
| | Regulamento Geral de Proteção de Dados |
| HLEG | *High-Level Expert Group on Artificial Intelligence* |
| | Grupo de Peritos de Alto Nível em Inteligência Artificial |
| IA | Inteligência Artificial |
| IoT | Internet das Coisas |
| JRC | Joint Research Centre (European Commission) |
| | Centro Comum de Investigação (da Comissão Europeia) |
| LGPD | Lei Geral de Proteção de Dados |
| OCDE | Organização para a Cooperação e Desenvolvimento Econômico |
| PLN | Processamento de Linguagem Natura |
| UE | União Europeia |

**SUMÁRIO**

# INTRODUÇÃO

A vida moderna é cheia de exigências. Das necessidades corporais mais básicas à natureza política do gênero humano, é notável o envolvimento da tecnologia como fator de conforto e comodidade na sociedade. Se computadores e celulares já são imprescindíveis há algum tempo, hoje se torna imprescindível um sistema bem específico nesses aparelhos: a Inteligência Artificial (IA).

Tudo que se confronta com a realidade humana, gera um impacto. Qual será o impacto da IA sendo tão íntima das pessoas? Será a demanda de supervisão, de mais intervenção humana, de robustez técnica e segurança, de proteção da privacidade e governança dos dados, de inclusão, equidade e diversidade, de influência no bem-estar social e ambiental, de responsabilidade? Uma dessas formas de resolução de problemas decorrentes desses impactos se dá pela aplicação do direito e da ética à IA.

Este capítulo inicial trata dos fundamentos, vocabulário e primeiras problematizações quanto à natureza da Inteligência Artificial, indo das primordiais bases filosóficas (ontologia e deontologia) até aos conceitos mais modernos e usuais em desenvolvimento.

No capítulo segundo, após está de posse do conhecimento “da coisa” que é a IA, esclarecemos a necessidade que surge em meio a tantos dados: garantir que os sistemas deles gerados apresentem um equilíbrio de interesses com os princípios que permeiam toda a sociedade, a ética, e que na falha dessa aplicação de forma eficiente haja um arcabouço de direito pronto para interferir nas tecnorregulações.

No capítulo três, é abordado de maneira mais profunda o movimento para construção de regulações eficientes, incentivando a justificativa das decisões de IA, dito de modo mais técnico desde já, a explicabilidade. Nesta parte os temas IA, ética e direito se interrelacionam, também como pontos abordados por mecanismos de organizações internacionais.

O capítulo quatro vai em exploração do que a União Europeia, os Estados Unidos e o Canadá já realizaram na tentativa de resolver os problemas legais da IA; o capítulo cinco faz o mesmo com relação ao Brasil.

O capítulo final faz um compilado comparativo de diversos assuntos em comum entre as três legislações internacionais com as existentes no Brasil. Neste âmbito, o esforço é pelas técnicas do direito comparado.

# 1 Ontologia e Deontologia da Inteligência Artificial

A ontologia é o estudo do que é o real e a certeza das existências no mundo, essa área aplicada à Inteligência Artificial (IA) fornece uma estrutura essencial para a organização e representação do conhecimento, facilitando a interoperabilidade entre sistemas e aprimorando a capacidade de raciocínio das máquinas (Tzimas, 2021). Através da definição clara de conceitos e suas relações, as informações ontológicas permitem que sistemas de IA compreendam e processem dados de maneira mais precisa e eficiente, promovendo avanços em diversas áreas de aplicação (Corrêa, 2013; Da Silveira, 2021).

Além de seu papel na filosofia, o conceito de ontologia foi adaptado e é amplamente utilizado em áreas como a ciência da computação, onde se refere a uma estrutura formal para representar conhecimento sobre um domínio específico (Salatino, 2020). Nesta forma de uso, uma ontologia define os tipos de coisas que existem em um determinado contexto e as relações entre elas, facilitando a organização, compartilhamento e análise de informações (Salatino, 2020; Da Silveira, 2021). É esta ideia bem estruturada que permite aos sistemas de IA compreender e utilizar informações de maneira consistente e lógica (Tzimas, 2021).

Em resumo, a ontologia aplicada à IA fornece uma estrutura essencial para a organização e representação do conhecimento, facilitando a interoperabilidade entre sistemas e aprimorando a capacidade de raciocínio das máquinas. Assim, a ontologia, na filosofia, é o estudo das categorias fundamentais do ser e da realidade, enquanto em ciência da computação, refere-se a uma representação estruturada de conhecimento sobre um domínio específico (Corrêa, 2013; Salatino, 2020).

Já quanto à discussão e à aplicação da deontologia a IA, foi tomada uma base kantiana (Kant, 2023). Fundamentada no imperativo categórico, essa abordagem exige que as ações de IA sejam universalmente aplicáveis e moralmente aceitáveis em qualquer circunstância similar (White, 2022). Isto implica que algoritmos de IA, como os utilizados em decisões de crédito ou recrutamento, devem operar de maneira justa, transparente e não

discriminatória (Wachter, 2022). Além disso, os sistemas de IA devem tratar os indivíduos como fins em si mesmos, respeitando sua dignidade e direitos, o que se traduz na proteção da privacidade dos dados e na prevenção de vieses prejudiciais (Da Silveira, 2020).

Desenvolver IA de forma responsável significa que os criadores e operadores desses sistemas devem buscar o bem-estar e a justiça, priorizando o benefício social e não apenas interesses econômicos. Isso envolve o cumprimento de deveres éticos e legais, garantindo que as decisões automatizadas sejam seguras, justas e transparentes (Nikolinakos, 2023). A autonomia humana deve ser respeitada, com os sistemas de IA operando de forma transparente para permitir decisões informadas pelos usuários. Embora a deontologia enfatize o dever sobre as consequências, na prática da IA, é essencial também considerar o impacto das ações dos sistemas, avaliando continuamente os resultados para evitar danos inadvertidos. Assim, a aplicação dos princípios kantianos à IA assegura que essa tecnologia promova valores éticos fundamentais e contribua positivamente para a sociedade (Corrêa, 2013; Da Silveira, 2021).

De forma prática e visual, a ontologia é uma dimensão representada de forma maior por um círculo com a finalidade de pensar os conceitos e desenvolvê-los de forma responsável para que a deontologia possa aplicar os princípios éticos no desenvolvimento da IA buscando sempre desempenhar um papel crucial na organização, representação e utilização do conhecimento de forma ética obedecendo a legislação e resguardando os direitos dos indivíduos.

## 1.1 Definições e Terminologia da Inteligência Artificial

Para Pei Wang (2019a), a definição de IA envolve mais do que apenas o uso da palavra, mas é sobre o conceito expresso por essa palavra. Destaca-se que o debate sobre o significado da IA frequentemente se refere ao conceito subjacente, independentemente do termo específico ou da língua usada. A escolha das palavras para expressar o conceito de IA é secundária em relação à ideia principal e que problemas associados à palavra “artificial”, como sua associação com “falso”, não são seu foco principal. Em vez disso, ele se concentra nos conceitos envolvidos. O termo “inteligência artificial” para o autor

poderia ser facilmente substituído por “inteligência computacional” ou até mesmo para “inteligência mecânica” uma vez que se concentra no conceito envolvido (Wang, 2019a).

Neste contexto, para o mesmo autor a definição de um conceito traça os seus limites. No entanto, mesmo com estas vantagens não podemos esperar que todos os conceitos sejam bem definidos desde o início, mesmo em se tratando de conceitos científicos. Os conceitos científicos passam por um conceito vago antes de se tornarem seguros para o embate dialético. Eles são definidos ao longo do tempo. Conclui-se que uma definição clara não é pressuposto para que o conceito seja utilizado tanto na investigação como nas discussões científicas, embora este fato seja desejado, no decorrer do caminho há o amadurecimento do significado do termo lexicográfico (Wang, 2019a).

Outra questão levantada sobre a definição da IA é que a definição do dicionário deve ser levada em consideração. Esta definição é útil para um revisor periódico ou conferência entender se aceita ou não um projeto de pesquisa de acordo com seu escopo de aceitação de trabalhos. Por outro lado, a definição funcional de IA define o objetivo de investigação para um projeto de IA, que quer dizer qual o sentido da IA, no contexto pesquisado, este conceito funcional por vezes não é igual ao do dicionário (Wang, 2019a).

De acordo com a obra Direito e Inteligência Artificial, que trata sobre os referenciais básicos com comentários à Resolução CNJ 332/2020 sobre a Inteligência Artificial, o conceito desta é de que pertence a um ramo da ciência da computação, que possui caráter interdisciplinar com outras áreas de conhecimento para tornar viável a capacidade de reprodução de ações cognitivas humanas. Assim, "a IA pode valer-se de diversas técnicas como estratégia de incremento de performance ou simplesmente de delegação de funções enfadonhas, repetitivas ou consideradas delegáveis e roboticamente praticáveis" (Hartmann Peixoto, 2020a, p. 17)

De acordo com a definição do Regulamento que cria regras harmonizadas do Parlamento Europeu e do Conselho da União Europeia sobre sistemas de IA, em vigor, é que uma característica central dos sistemas de inteligência artificial (IA) é sua habilidade de realizar inferências. Este processo de inferência envolve a obtenção de resultados, como previsões, conteúdos, recomendações ou decisões, que podem impactar tanto ambientes físicos

quanto virtuais. Além disso, essa capacidade abrange a habilidade dos sistemas de IA de derivar modelos ou algoritmos a partir de entradas ou dados fornecidos, permitindo que esses sistemas aprimorem suas funcionalidades com base nas informações recebidas (European Parliament, 2024). O Eurostat (Gabinete de Estatística da União Europeia), de forma complementar, considerou a definição da estratégia Europeia para IA, como sendo sistemas criados para atingir objetivos específicos que exibem comportamentos inteligentes e com um algum grau de autonomia (Samoili *et al,*. 2020).

De acordo com o *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* legislação dos EUA o conceito de IA é utilizado como:

> O termo “inteligência artificial” ou “IA” tem o significado estabelecido em 15 U.S.C. 9401(3): um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas de inteligência artificial utilizam dados baseados em máquinas e humanos para perceber ambientes reais e virtuais; abstrair tais percepções em modelos por meio de análise de forma automatizada; e usar a inferência de modelos para formular opções de informação ou ação (The White House, 2023).

Conforme o projeto de lei C-27 (Bill C-27) do Canadá o conceito de sistema de IA se perfaz através de um sistema tecnológico que ao processar dados por rede neural, aprendizado de máquina e algoritmo genético (ou alguma técnica alternativa de mesmo nível) consegue tomar decisões e estabelecer um resultado com conteúdo, seja de forma autônoma, seja parcialmente autônoma (Parliament of Canada, 2022).

Para chegar numa discussão proveitosa, é necessário distinguir uma definição funcional de uma definição de dicionário:

> Finalmente, uma definição funcional de IA deverá dar ao campo uma identidade adequada, especificando o seu objeto e âmbito, o que decidirá a sua relação com outros campos, como a ciência da computação e a psicologia cognitiva (WANG, 2019a).

Segundo uma ótica interdisciplinar de acordo com o livro “*Impact of Artificial Intelligence in Business and Society: Opportunities and Challenges*”, a IA é um campo interdisciplinar que integra conhecimentos de biologia, ciência da computação, filosofia, matemática, engenharia, robótica e ciência cognitiva.

Seu propósito é simular a inteligência humana por meio de tecnologias computacionais (La Torre *et al.*, 2023).

Um trabalho recente que traz características edificantes é o Relatório Técnico da *Joint Research Centre* (JRC) da Comissão Europeia (CE)[1], “AI Watch”[2], cujo objetivo foi dado pela busca de “uma definição operacional e taxonomia (classificação) de inteligência artificial”, em que os sistemas de inteligência artificial (IA) são concebidos como sistemas de *software*, e possivelmente *hardware*, criados por humanos para atingir objetivos complexos. Esses sistemas operam em dimensões físicas ou digitais, percebendo seu ambiente através da coleta de dados. Eles interpretam tanto dados estruturados quanto não estruturados, raciocinam sobre o conhecimento obtido ou processam as informações derivadas desses dados e decidem as melhores ações a serem tomadas para atingir seus objetivos. Além disso, os sistemas de IA podem empregar regras simbólicas ou aprender com modelos numéricos, adaptando seu comportamento ao analisar o impacto de suas ações anteriores no ambiente (Samoili *et al.*, 2020).

A definição anterior é abrangente e altamente técnica e existem outras definições menos especializadas que são tão importantes como a anterior. Desse modo, uma outra definição importante preceitua que o termo IA é geralmente utilizado para designar qualquer máquina ou algoritmo que possa observar seu ambiente e, utilizando o conhecimento e a experiência adquiridos, realizar ações inteligentes ou sugerir decisões. De acordo com o relatório emblemático do JRC da CE, publicado em 2018, intitulado "*Artificial Intelligence: A European Perspective*"[3], várias tecnologias diferentes se encaixam nessa definição ampla de IA, sendo que, atualmente, as técnicas de aprendizado de máquina (*machine learning* - ML) são as mais empregadas. Outros abrangeram subdomínios como  veículos conectados e automatizados (CAVs), reconhecimento de fala e processamento de linguagem natural (PLN) e reconhecimento facial (Annoni *et al.*, 2018).

---

[1] “Centro Comum de Investigação [da Comissão Europeia]”, em tradução livre.
[2] “Relógio de IA”, em tradução livre.
[3] “Inteligência Artificial: Uma perspectiva Europeia”, em tradução livre.

O Inquérito comunitário sobre a utilização das TIC e o Comércio eletrônico nas empresas de 2021 conceitua a IA de uma forma mais aprofundada e abrangente com a finalidade de atingir objetivos específicos:

> Inteligência Artificial refere-se a sistemas que utilizam tecnologias como: mineração de texto, visão computacional, reconhecimento de fala, geração de linguagem natural, aprendizado de máquina, aprendizado profundo para coletar e/ou usar dados para prever, recomendar ou decidir, com níveis de autonomia, a melhor ação para atingir objetivos específicos (Samoili *et al*., 2020, p. 29).

Desta forma, é compreendido que os conceitos mencionados são essenciais para a definição de Inteligência Artificial (IA) ou, pelo menos, para a busca por uma definição, visto que a tecnologia avança mais rapidamente do que outras áreas, exigindo uma constante renovação da conceituação da IA devido a esse dinamismo.

Segundo o texto "*Artificial Intelligence Principles for Vulnerable Population in Humanitarian Contexts*", a IA é tratada como um campo amplo de pesquisa dentro da ciência da computação que ainda não possui uma definição formal (Wright e Verity, 2020). De forma inversa, Wang (2019a, p. 16-17) afirma que essa área não só não pode ser confundida com a ciência da computação, como não faz parte dela. Isso ocorre pois mesmo sendo implementada em sistemas computacionais, a IA apresenta diferenças essenciais em relação aos sistemas tradicionais, por exemplo, além de apenas resolver mais problemas, e de modo mais específico, a inteligência exige um método distinto para projetar e utilizar computadores, diferente do método tradicional capturado pela definição atual de computação (Wang, 2019a). No entanto, dado o estudo do design de agentes inteligentes — em que agentes se referem a unidades de processamento, como computadores — a dependência das decisões humanas e a incapacidade de pensar por conta própria, limitando-se à extração de padrões dos conjuntos de dados fornecidos a seus algoritmos, é clara (Wright e Verity, 2020). Isto parece ser uma ideia contraditória à inteligência de um sistema como sua capacidade de resolver problemas disposto por Wang (Wang, 2019a).

Neste mesmo sentido de conceituar a Inteligência Artificial, o pesquisador Luciano Floridi traz o conceito de IA à luz do consenso

Bruxelas-Washington: “Ambos os lados do Atlântico concordam que a IA é um artefacto (*software* ou sistema baseado em máquinas)” que pode influenciar o ambiente, seja qual for, a partir de suas predições guiadas pelo conjunto de propósitos do desenvolvedor. Para tanto estes mesmos “lados” devem estar atentos à importância do conteúdo gerado (Floridi, 2023).

Floridi (2023, p. 87) destaca que a OCDE (Organização para a Cooperação e Desenvolvimento Econômico) definiu sistemas de IA em 2018 como sistemas baseados em máquinas capazes de realizar previsões, recomendações ou decisões para alcançar objetivos definidos por humanos, influenciando ambientes reais ou virtuais. E continua: esses sistemas utilizam informações de máquinas ou humanos para perceber ambientes, transformar essas percepções em modelos por meio de análises automatizadas ou manuais, e usar esses modelos para formular opções de ação. Em 2023, a OCDE revisou essa definição, removendo a cláusula "definida pelo homem" e melhorando a descrição ao incluir a distinção entre ambientes físicos e virtuais. A nova definição descreve sistemas de IA como baseados em máquinas que inferem, a partir das informações recebidas, como gerar resultados que podem influenciar esses ambientes. Os sistemas variam, operando em diversos níveis de autonomia e adaptabilidade após a implantação (Floridi, 2023).

Assim,  Protásio, Faria e Hartmann (2022, p. 272) ressaltam que embora não tenha um consenso a respeito do significado da palavra inteligência artificial pela doutrina, que através de um esforço em entender esta nova área e sua respectiva influência no direito vem defendendo o significado da IA como "a possibilidade de que as máquinas, em alguma medida, 'pensem', ou melhor, imitem o pensamento humano aprendendo e usando as generalizações que as pessoas usam para tomar nossas decisões habituais". O mesmo pensamento esteve presente em obras do final do século passado, incorrendo, por exemplo, na obra “*Artificial Intelligence: A New Synthesis*”, na qual a IA é retratada como uma área que se preocupa com o comportamento inteligente, de modo que, num longo prazo, máquinas seriam desenvolvidas para exercer funções de percepção, raciocínio, aprendizagem, comunicação, melhor que os humanos, ou ao menos no mesmo nível. Além disso, a IA corroboraria com a operação em cenários complexos, tanto para compreender compreender comportamentos de humanos, de diversos animais, quanto de outras

máquinas; de modo a ter seus objetivos abrangendo fins de engenharia e de ciência (Nilsson, 1998).

Chollet (2020, p. 27), ao continuar a discussão a partir de Wang (2019a), argumenta que definir IA é, em grande parte, sinônimo de definir inteligência. Ele explica que, enquanto a inteligência natural é especializada em lidar com problemas do mundo físico, a inteligência artificial é direcionada para problemas selecionados por seus criadores. Chollet é simpático à definição dada por Wang que por fim, define inteligência como um processo de aprendizagem e adaptação ao longo da vida que ocorre dentro de um agente individual, impulsionado pela experiência incorporada e que se desenrola em tempo real, mesmo com informações e recursos insuficientes, como recursos computacionais (Chollet, 2020).

Para finalizar a definição de Inteligência Artificial, é importante citar novamente Hartmann Peixoto (2020a, p. 17). Pois este apresenta a natureza da IA como uma caracterização da capacidade de organizar informações de maneira a proporcionar soluções aceitáveis para problemas complexos, representando uma integração sofisticada de funções cognitivas artificiais.

Desse entendimento de contínua conformidade com a realidade, decorre que os meios de incremento de performance pela IA, nas diferentes aplicações, são sempre multifacetadas (Hartmann Peixoto, 2020a). Suas utilidades envolvem aprimoramentos de modelos, de qualidade dos dados, de técnicas de treinamento, na otimização dos recursos computacionais e no monitoramento de informações, em todas as áreas do conhecimento (Van Noorden e Perkel, 2023).

## 1.2 Especificidades da Inteligência Artificial

Tendo em vista as ideias apresentadas sobre IA, é necessário explicitar brevemente alguns termos do vocabulário da área de computação, tanto por representarem bases da IA, por serem um meio para a eficiência desta, quanto por expressarem uma quebra de paradigma.

Desse modo, toma-se o algoritmo como um fator elementar na programação computacional. De modo simples, um algoritmo é um conjunto finito e ordenado de instruções projetadas para resolver um problema

específico ou realizar uma tarefa. Ele atua como um modelo abstrato que descreve o processo lógico necessário para transformar a entrada de dados em uma saída satisfatória, os resultados (Gawiejnowicz, 2020).

A estrutura de um algoritmo inclui uma sequência de operações, como cálculos, comparações, decisões e ciclos de repetição, que devem ser realizadas de maneira precisa, evitando ambiguidades. A eficácia de um algoritmo é avaliada por sua capacidade de gerar o resultado correto, otimizando o uso de recursos computacionais. Em sistemas de IA, essa eficiência é crucial (Gawiejnowicz, 2020).

Há uma discussão há muito levantada sobre alguns problemas éticos dos algoritmos, como a questão do preconceito observado, quando em execução em sistemas de tomada de decisão. Determinar o preconceito em algoritmos e nos modelos que eles produzem é um desafio quando não se analisa o contexto de produção destes (Mittelstadt, 2016).

Outros problemas dizem respeito ao preconceito técnico pode surgir de limitações tecnológicas, erros ou decisões de design que favorecem determinados grupos sem uma justificativa subjacente. Nesse caso, não há distinção entre padrões justos e injustos nos dados (Gawiejnowicz, 2020).

Nesse contexto, um breve nivelamento de termos é dado a seguir, segundo o que se relaciona intimamente com IA e algoritmos: Aprendizado de Máquina, Aprendizagem Profunda, Processamento de linguagem natural e Inteligência Artificial generativa.

### 1.2.1 Aprendizado de Máquina (machine learning)

A aprendizagem de máquina, *machine learning* (ML), refere-se a um campo da computação, que se constitui em uma subárea IA, envolvendo o desenvolvimento de algoritmos capazes de identificar padrões, aprender, tomar decisões e melhorar baseados em um grande conjunto de dados. Isto se dá sem a necessidade de serem explicitamente programados para cada tarefa específica (Pramod, Naicker e Tyagi, 2021; Zhu e Goldberg, 2022).

Há dois tipos fundamentais de trabalho da ML, a aprendizagem supervisionada e a não supervisionada. Na aprendizagem supervisionada, o sistema é treinado com um conjunto de dados rotulados, onde as respostas corretas são fornecidas, permitindo ao algoritmo aprender a mapear entradas

para saídas específicas. Já na aprendizagem não supervisionada, o sistema recebe dados não rotulados e deve identificar padrões ou agrupamentos inerentes aos dados por conta própria, sem orientação externa (Zhu e Goldberg, 2022).

O fator unitivo no modo de funcionamento da ML inclui a capacidade dos algoritmos de generalizar a partir de exemplos, o que significa que, ao serem expostos a novos dados, eles podem aplicar o conhecimento adquirido para prever resultados ou tomar decisões. Além disso, a eficácia de um modelo de aprendizagem automática depende da qualidade e quantidade de dados disponíveis, bem como da capacidade do algoritmo de evitar o *overfitting*, ou seja, a adaptação excessiva ao conjunto de dados de treino, que comprometeria sua capacidade de generalização (Zhu e Goldberg, 2022).

De modo geral, o comportamento de um modelo de *machine learning* é influenciado pelas características dos conjuntos de dados oferecido para treiná-lo, o que pode acarretar em problemas éticos e jurídicos. Outros temas que fazem parte dessa discussão são: regressão, árvores de decisão, redes neurais e análise preditiva (Zhu e Goldberg, 2022).

No campo do direito, esta tecnologia emergente já foi aplicada em alguns países e até mesmo regulada, como se verá no Capítulo 4.

### 1.2.2 Aprendizagem Profunda (*deep learning*)

Aproveitando-se do tópico anterior, a aprendizagem profunda, *deep learning*, é uma extensão avançada da ML, considerada, portanto, uma subárea desta. Valendo-se dos algoritmos, a *deep learning* se constrói com múltiplas camadas de redes neurais artificiais por onde vão passar os dados para serem processados (com relações complexas e não lineares) e gerar os resultados esperados. Esse ramo foi projetado para imitar a forma como os neurônios processam informações e gerar ainda mais eficiência tecnológica (Pramod, Naicker e Tyagi, 2021).

Uma das principais vantagens da aprendizagem profunda está na sua capacidade de lidar com grandes conjuntos de dados e identificar padrões complexos que podem não ser facilmente identificáveis por algoritmos de aprendizado de máquina mais convencionais. Algumas aplicações diferenciadas tem sido em áreas como reconhecimento de imagem,

processamento de linguagem natural e jogos, nos quais a complexidade dos dados requer a extração de características em múltiplos níveis de abstração (Taye, 2023).

Os modelos de aprendizagem profunda são frequentemente comparados a "caixas-pretas" devido à dificuldade em interpretar as decisões que tomam (problemática a ser desenvolvida no Capítulo 3) (Hassija *et al.*, 2024). Além disso, exigem grandes volumes de dados para treinamento e uma capacidade computacional significativa, o que pode limitar a sua aplicabilidade em contextos nos quais estes recursos são escassos.

### 1.2.3 Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) combina elementos de linguística, ciência da computação e estatística para analisar e modelar textos , além de outros elementos discursivos, em linguagem natural — humanizada. Para tanto, se utiliza, em modelos mais simples, de abordagens baseadas em regras e algoritmos estatísticos, e, em modelos mais complexos, das redes neurais de ML e *deep learning* (Chowdhary e Chowdhary, 2020).

Para se ter noção, uma técnica simples é a vetorização de palavras e uma técnica mais avançada é encontrada em modelos como o BERT e o GPT. Essas aplicações interagem com uma pessoa capturando nuances linguísticas e gerando respostas precisas (Chowdhary e Chowdhary, 2020).

Em campos como a medicina, está claro o desenvolvimento da PLN até para suporte eficaz de pessoas com saúde mental comprometida (Swaminathan, 2023). No campo do direito, o PLN tem potencial de aplicações significativas, como a análise automatizada de documentos legais, a extração de informações relevantes de textos jurídicos extensos e a geração automática de resumos.

### 1.2.4 Inteligência Artificial Generativa

Valendo-se do que foi desenvolvido na ML, deep learning e PLN, eis que surge a Inteligência Artificial Generativa. É a subárea da IA que se distingue pela capacidade de criar conteúdos originais, como textos, imagens, sons e

vídeos, a partir de dados previamente analisados, de maneira a realmente emular a capacidade criativa humana (OECD, 2024a).

No entanto, a IA Generativa levanta sérias questões de direitos de propriedade intelectual, especialmente no que diz respeito ao uso de conteúdo não licenciado em dados de treinamento e à possível violação de direitos autorais, patentes e marcas registradas em criações geradas por IA. A propriedade de obras geradas por IA também é um ponto de controvérsia. Nos Estados Unidos e na Europa, há disputas legais sobre a legalidade de treinar modelos de *machine learning* em material protegido por direitos autorais sem permissão. Esses processos legais têm o potencial de estabelecer precedentes importantes e impactar significativamente a indústria de IA generativa, afetando tanto startups quanto grandes empresas de tecnologia (OECD, 2024a).

Afinal, como fazer a aplicação dessa tecnologia especialmente em áreas sensíveis como o direito, que exigem um equilíbrio cuidadoso entre inovação e responsabilidade, assegurando que as ferramentas criadas sejam utilizadas de maneira ética e eficaz ?

## 2. Aplicações de Direito e Ética à Inteligência Artificial

O domínio da IA é dividido em dois principais pilares: centrado em dados e centrado no ser humano. A IA centrada em dados envolve métodos e técnicas que necessitam de dados para treinamento e são capazes de fazer previsões. Já a IA centrada no ser humano envolve sistemas que amplificam e aumentam as habilidades humanas, em vez de substituí-las (Appio *et al*., 2024; GPAN, 2024). Esse tipo de IA busca manter o controle humano, garantindo que atenda às necessidades, opere de forma transparente e ética, respeite a privacidade e seja justa em suas previsões (Capel e Brereton, 2024). O principal objetivo da IA centrada no ser humano é projetar e usar tecnologias que apoiem e capacitem os humanos. Para isso, esses sistemas precisam ser transparentes e explicáveis, de modo que os humanos possam entender seu funcionamento e confiar nos resultados. As explicações devem fornecer razões que abordem as preocupações dos gestores e profissionais que utilizam o sistema em contextos específicos (Capel e Brereton, 2024; Appio *et al*., 2024).

Ora, para uma IA centrada no ser humano, é necessário garantir que a IA esteja a serviço do ser humano, para tanto, não é urgente discutir o Direito e a Ética nesse contexto?

Um problema de pesquisa se faz premente e foi o que motivou este trabalho, em duas linhas:

Como a utilização dos princípios e diretrizes da ética pode contribuir para o desenvolvimento de uma Inteligência Artificial mais responsável?

O Direito é capaz de regular sozinho a Inteligência Artificial sem se utilizar dos princípios e diretrizes da ética?

Nas discussões a seguir será apresentado os meios necessários para alcançarmos as respostas a estas perguntas.

## 2.1 Definição de ética

Não é objetivo deste trabalho problematizar o entendimento de “ética”. No entanto, é de se esperar que uma dissertação que traga essa palavra no título apresente alguma definição, para, então, poder aplicá-la ao que se diz. Para tanto, a definição a ser apresentada é abrangente o suficiente para captar os pormenores dos problemas que a IA pode provocar.

Após uma breve pesquisa na ética nos trabalhos de Aristóteles, Tomás de Aquino, Alasdair MacIntyre, Immanuel Kant, John Rawls e Jürgen Habermas, foi encontrado um outro autor, o John Finnis, cujo trabalho é analítico e verossímil o suficiente para exercer um trabalho sob estes termos (França, 2022):

A ética é um conjunto de princípios que orienta a conduta humana.

O que busca essa ética, em que contexto e para construir o que?

Busca o equilíbrio entre o dever moral, a realização da justiça e o desenvolvimento das virtudes, dentro de um contexto racional, comunitário e histórico, para promover a dignidade humana, a equidade e o bem comum (Finnis, 1983).

## 2.2 Definição de Direito

Para este trabalho o posicionamento quanto à definição de direito se deu de maneira semelhante à definição de ética. Há o reconhecimento de autores que tratam o sistema de normas de forma distinta da moralidade e há aqueles que os veem em conjunto. O contraste e complementação entre Hans Kelsen, H.L.A. Hart, Joseph Raz, Ronald Dworkin e Miguel Reale foi realizado para, de forma sucinta, dar prosseguimento ao texto, com a seguinte ideia:

O Direito é um sistema normativo composto por regras e princípios estabelecidos por uma autoridade legítima, cujo objetivo é regular a conduta humana dentro de uma sociedade. Embora o direito possa ser concebido como distinto da moralidade, ele se estrutura sobre a autoridade, a validade e a aplicação de normas que, em muitos casos, refletem valores éticos compartilhados pela comunidade. Isto requer, ainda, uma coerência interna e uma obediência às regras procedimentais para garantir sua legitimidade e

eficácia, promovendo a ordem social, a justiça e o bem comum (Ferraz Junior, 2003).

## 2.3 Pontos de inflexão da Inteligência Artificial

A era do *Big Data* potencializada pela IA, marcada pela explosão de dados gerados e coletados em uma escala sem precedentes, juntamente com o aumento da capacidade computacional e a evolução dos algoritmos, tem transformado profundamente diversos setores e a sociedade como um todo. A capacidade de processar vastas quantidades de dados em tempo real e extrair *insights* valiosos proporciona vantagens competitivas significativas, impulsiona a inovação e melhora a eficiência em áreas como saúde, finanças, *marketing* e segurança, no entanto também é capaz de produzir conclusões enviesadas como vimos anteriormente se não for utilizada de maneira adequada.

### 2.3.1 Avanço do *Big Data* e Internet das Coisas

O *big data* representa uma revolução tecnológica; as empresas que estão de posse destes dados coletados pelo *Big Data* (informações ricas em volume, velocidade e variedade) devem ser capazes de se tornarem altamente competitivas após analisar os dados derivados da Internet das Coisas (*Internet of Things* - IoT) podendo desenvolver novos negócios e produtos para o público alvo de consumo (Sestino *et al*., 2020).

Assim, neste contexto existem várias maneiras interessantes pelas quais a IoT[4] pode transformar serviços tradicionais em conjunto com o *big data*. Assim, segundo o texto de Sestino *et al*.(2020, p. 105), a ampla disponibilidade de dados sobre as atividades dos funcionários pode permitir que as empresas orientem e capacitem melhor seu pessoal. Em serviços, dispositivos IoT, como

[4] “IA, IoT e robótica compartilham muitas características. Elas podem combinar conectividade , autonomia e dependência de dados para executar tarefas com pouco ou nenhum controle ou supervisão humana. Sistemas equipados com IA também podem melhorar seu próprio desempenho aprendendo com a experiência. Sua complexidade se reflete tanto na pluralidade de operadores econômicos envolvidos na cadeia de suprimentos quanto na multiplicidade de componentes, peças, *software*, sistemas ou serviços, que juntos formam os novos ecossistemas tecnológicos. Somado a isso está a abertura para atualizações e *upgrades* após sua colocação no mercado. As vastas quantidades de dados envolvidas, a dependência de algoritmos e a opacidade da tomada de decisão de IA tornam mais difícil prever o comportamento de um produto habilitado para IA e entender as causas potenciais de um dano. Finalmente, a conectividade e a abertura também podem expor produtos de IA e IoT a ameaças cibernéticas” (European Commission, 2020b).

assistentes de voz, podem ser integrados em serviços tradicionais, como quartos de hotel e locais de trabalho, melhorando seu planejamento e prestação. Dispositivos IoT e análise de dados podem também fortalecer a gestão de segurança e serem integrados em máquinas de produção para monitorar carga e descarga de mercadorias, evitar paradas de produção, reduzir defeitos ou permitir modelos de negócios de servitização. Na área da saúde, dispositivos IoT podem conectar pacientes a plataformas monitoradas por médicos, permitindo intervenções rápidas em casos de anomalias de saúde, como na telemedicina ou sistemas de vida assistida. Governos locais poderiam usar esses dispositivos para proteger profissionais de saúde e conter epidemias, como a COVID-19 (Sestino *et al*., 2020).

*Big Data* e IoT também podem levar empresas a reconsiderar suas necessidades de recursos humanos, exigindo investimentos em cientistas de dados para gerenciar e extrair valor dos dados coletados. Esses dispositivos podem fortalecer plataformas digitais tradicionais, facilitando maior monitoramento, otimização, controle e avaliação de desempenho, até mesmo em tempo real. (Sestino *et al*., 2020).

De modo concomitante, é possível abstrair que  com o uso de IoT e *big data* os consumidores podem apresentar desconfiança em relação a segurança social e de dados. Assim, as empresas devem estar preocupadas com padrões éticos e de privacidade no recolhimento destes dados, se não for feito de forma ética as consequências como a inibição da vontade dos consumidores de interagir com plataformas online e dispositivos móveis será uma medida a cada dia mais implementada por esta categoria. Desta maneira, as empresas tem o difícil conflito de sopesar os próprios interesses no recolhimento destas informações com a privacidade e ética dos dados coletados através destas ferramentas de IA (Sestino *et al.,* 2020).

Em outro sentido, e não menos importante, o texto de Gutierrez *et al.* (2019, p. 305) aborda o fato de que os indivíduos estão gradualmente mais tendenciosos a adaptar-se a partilha de dados pessoais e aos benefícios conexos de abdicar de parte do direito à privacidade em troca de vantagens econômicas no momento da compra de seus produtos. As recompensas monetárias podem ser em troca de moeda ou vouchers, desconto e certificados

de presente — o peso dessas vantagens fazem os consumidores mais aptos ao compartilhamento de informações.

Assim, Gutierrez *et al.* (2019, p. 301) ressalta que a disposição dos consumidores para compartilhar informações é influenciada negativamente pelo risco percebido e positivamente pelos benefícios percebidos. Quando os consumidores fornecem informações de localização aos profissionais de marketing, eles podem receber anúncios personalizados que correspondem aos seus interesses, atividades, locais frequentes e horário do dia, além de obterem comunicações sobre recompensas monetárias. Por outro lado, as questões sobre a privacidade e ética pulsam neste caso, já que existe uma diferença muito grande de como as suas informações pessoais estão sendo usadas e a realidade da sua utilização.

Numa outra análise, Ekbia *et al.* (2015, p. 1531) trata sobre os “Dilemas do *Big Data*’’ e ressalta que a produção de dados envolve múltiplos agentes sociais com interesses distintos. Além disso, os métodos de geração de dados são frequentemente ambíguos e subdocumentados, como evidenciado por estudos em diferentes contextos históricos e etnográficos, possibilitando, desse modo, afirmar geração de dados distorcidos. O *big data* não chega às mãos dos analistas prontos para a análise, isto envolve trabalho humano mecanizado e interpretativo e opiniões subjetivas que pode “estragar os dados”. E, ainda, o processo de analisar quais os “melhores” dados a serem condicionados e utilizados podem “estragar os dados”, pois é uma análise subjetiva dos dados que pode estar eivados, por vezes, de vieses discriminatórios.

Assim, Ekbia *et al.* (2015, p.1531) preceitua que essas questões tornam-se especialmente críticas quando dados pessoais são utilizados em larga escala e a "desidentificação" torna-se uma preocupação central, exacerbada pelo potencial de "reidentificação". Esse potencial prejudica o avanço da pesquisa de *Big Data* em direção à "liquidez de dados". Existe um dilema entre a partilha de dados, liquidez e transparência, e os riscos à privacidade e anonimato decorrentes da reidentificação. Esse dilema é evidente em diversas áreas como medicina, pagamentos com localização geográfica, localização geográfica de dispositivos móveis e mídias sociais.

O mascaramento, ou criptografia, visa proteger os dados pessoais dos cidadãos, mas a possibilidade de reidentificação, exacerbada pelos

supercomputadores atuais, levanta questões sobre privacidade. Outra proposta para a privacidade envolve o uso de software que mede e monitora o uso de dados individuais, exigindo que os cidadãos marquem seus dados com preferências de privacidade. Segundo o Fórum Econômico Mundial, isso sugere um futuro onde bancos, governos e prestadores de serviço coletam e fornecem dados pessoalmente identificáveis aos consumidores recolhidos sobre eles (Fórum Econômico Mundial, 2013, p.13). Contudo, ao usar metadados para proteger os dados, essa solução coloca a responsabilidade pela proteção da privacidade nos próprios cidadãos (Ekbia *et al.*, 2015).

### 2.3.2 Necessidade de um avanço ético

Existe uma convergência lógica ou ponto de inflexão da junção da IA e *big data* potencializando, neste contexto, o aumento da capacidade computacional capaz de gerar a evolução dos algoritmos. E como se não bastasse essa conjuntura, uma aplicação da ética neste contexto é o que pode evitar resultados enviesados.

Um conjunto de preocupações com a privacidade em várias áreas como saúde, governo, inteligências e dados do consumidor. Podemos citar o caso da Target, loja americana, que estava prevendo a gravidez de gestantes com o intuito de oferecer cupons de desconto, na mesma perspectiva, anunciantes começaram a fazer o perfil de compra das pessoas através de metadados publicamente disponíveis de localização geográfica das postagens em redes sociais. Em outra perspectiva, e não menos importante, no campo da saúde, visualiza-se que o armazenamento e compartilhamento de dados de genomas podem relatar algo não somente sobre a pessoa, mas sobre a família da pessoa a quem se estuda. Estes exemplos acima, são apenas alguns que acontecem frequentemente e que demandam discussão sobre o sopesamento da evolução tecnológica e a ética neste contexto. Visualiza-se neste contexto que a IA não teria crescido tão rapidamente senão fosse o “combustível’’ do *big data*. A IA é intensamente utilizada na coibição e prevenção de fraudes, no comércio eletrônico para prever padrões de compra e necessidade futuras dos consumidores (Ekbia *et al.*, 2015).

Passando a citação de grandes empresas, inclusive pelas quais se forma grande parte do *big data*, o *Google* (Alphabet Inc.) mapeia o

comportamento dos indivíduos ao longo dos anos com base em sua localização, utilizando dados de serviços como *Gmail* e aplicativos *Android* nos *smartphones*. O *Facebook* (Meta), por sua vez, coleta informações fornecidas pelos usuários, como curtidas, postagens e fotos, utilizando metadados de localização nas fotos e dados capturados pela câmera em seus aplicativos. Além disso, o *Facebook* utiliza o catálogo de endereços, registros de chamadas e SMS para sugerir contatos, e monitora diversos dados do dispositivo, como nível da bateria, intensidade do sinal e armazenamento disponível. No computador, o *Facebook* registra o tipo de navegador e seus *plug-ins*, rastreia se a janela está em primeiro ou segundo plano e os movimentos do *mouse*. Mesmo que os serviços de localização sejam desativados, o *Facebook* continua rastreando a localização dos indivíduos por meio de endereços IP, pontos de acesso *Wi-Fi* e torres de celular próximas. Parece irrelevante esta grande quantidade de dados coletados pelo *Google* e pelo *Facebook*, no entanto, esta grande quantidade de dados pode construir uma visão abrangente sobre as pessoas e fazer o perfil de consumo e outras preferências do indivíduo, o que pode afrontar direitos e garantias individuais. Tudo isso é afirmado no trabalho de Alben (2020), intitulado “Colisão da Inteligência Artificial e Big Data”.

Neste mesmo sentido, a discussão sobre a colisão da IA e *big data* se expande trazendo à luz as consequências para nossa privacidade e autonomia, observa-se em trechos da leitura do texto que as questões de vieses discriminatórios puderam ser observados na utilização do software Rekognition:

> Quando a Wired Magazine publicou um artigo em 2018 citando um estudo da ACLU sobre o software de reconhecimento facial da Amazon que combinou erroneamente 28 dos 435 membros do Congresso dos EUA com um banco de dados de fotos policiais, organizações de liberdades civis alegaram um forte preconceito racial no software Rekognition, dado que os indivíduos com tons de pele mais escuros tinham duas vezes mais probabilidade de serem comparados com o banco de dados de prisões com uma configuração de nível de confiança de 80%. (Alben, 2020).

Desta maneira, é importante que a tecnologia evolua a ponto de não cometer erros graves e tendenciosos a discriminação de pessoas. Por isto, o autor conclui que a solução não é coibir o uso de reconhecimento facial, mas sim evoluir esta tecnologia para que ela possa se ater às características físicas

das pessoas sendo possível uma identificação facial, através do *software* Rekognition, sem levar em consideração fatores raciais. A tecnologia deve avançar sempre levando em consideração fatores éticos na sua aplicação para o alcance de sua melhor versão sem a interrupção de sua aplicação em fases iniciais (Alben, 2020).

Neste contexto, é crucial que empresas, governos e a sociedade civil colaborem para estabelecer diretrizes éticas robustas e regulamentações que garantam a proteção dos direitos dos indivíduos. A implementação de práticas de privacidade por *design*, a transparência nos processos algorítmicos e a responsabilização por decisões automatizadas são passos essenciais para mitigar os riscos associados ao *Big Data* e à inteligência artificial. Assim, ao mesmo tempo que aproveitamos as oportunidades oferecidas pela explosão de dados e avanços tecnológicos, devemos garantir que esses desenvolvimentos sejam conduzidos de maneira ética e responsável, promovendo um equilíbrio entre inovação e proteção dos direitos humanos. Somente com uma abordagem consciente e ética poderemos construir um futuro digital que beneficie a todos, sem comprometer os valores fundamentais da nossa sociedade (Alben, 2020).

Nos entremeios de uma evolução tecnológica, dilemas legais e éticos com alta expectativa social demandam uma resposta. Uma forma de trazer clareza sobre tamanho volume de dados, tanto para o consumidor, para empresas, quanto usuário especializado, é por meio do *dataset*.

## 2.4 Conceito de *datasheet* e *dataset*

Na IA, um "datasheet" pode ser interpretado de forma análoga ao conceito tradicional usado em engenharia e eletrônica, mas adaptado para descrever modelos de IA, algoritmos ou conjuntos de dados. Um *datasheet* é um documento com informações detalhadas sobre um conjunto de dados usado em modelos de IA (descrição, informações, composição, qualidade, direitos e ética) do conjunto de dados. Semelhante aos *datasheets* de engenharia, eles fornecem uma visão abrangente e padronizada das propriedades e limitações do conjunto de dados. Esses datasheets promovem práticas responsáveis e padronizadas na comunidade de IA. *Dataset* "é uma

definição mais estrita de conjunto de dados, em formatos adequados para a realização dos treinamentos e testes de aferição de desempenho" (Hartmann Peixoto, 2020a; Hartmann Peixoto, 2020b; Gebru, 2021).

### 2.4.1 Problema deontológico de fazer *dataset*

No documento escrito por Gebru *et al.* (2021), *Datasheet for dataset* (“Folhas de dados para conjunto de dados”) propõe que cada "conjunto de dados seja acompanhado de uma folha de dados — *Datasheet* — que documente sua motivação, composição, processo de coleta, usos recomendados e assim por diante". Assim, um “*datasheet for dataset*” têm o potencial de aumentar a transparência e a responsabilização dentro da comunidade de aprendizado de máquina, mitigar vieses sociais indesejados em modelos, facilitar maior reprodutibilidade dos resultados dessa área e ajudar pesquisadores e profissionais a selecionar conjunto de dados mais apropriados para suas tarefas escolhidas possibilitando desenvolver um sistema de IA mais ético e responsável.

Apesar das planilhas para conjunto de dados não fornecerem uma solução completa para eliminar os vieses sociais indesejados, riscos ou danos em potencial, esta é capaz de trazer à tona questões obscuras de vieses presentes e os criadores de conjunto de dados poderão mitigar ou eliminar questões discriminatórias antes do sistema de IA estar em produção para benefício dos consumidores de conjuntos de dados. É importante, ao se fazer as planilhas de conjunto de dados, responder às perguntas principais do estágio do ciclo de vida do conjunto de dados e principalmente fazer a pergunta se houve um processo de revisão ética do conjunto de dados conduzido pelos interessados (Gebru *et al.,* 2021).

Os dados usados na aprendizagem de máquina são a "matéria prima" de desenvolvimento de sistemas de IA. É por este intuito que o Fórum Econômico Mundial preceitua que todas as entidades que desenvolvem sistemas de IA devem documentar, através de *datasheet*, a procedência, criação e utilização de conjuntos de dados de aprendizagem automática para evitar resultados discriminatórios e assim, incentivar a práticas éticas na escolha dos dados (Boyd, 2021; Gebru *et al.*, 2021). Em seguida, outras representações também passaram a indicar fortemente a produção de

*datasheet*, como a influente *Call For Datasets & Benchmarks (NeurIIPS)*[5] (Mcmillan-Major, Bender e Friedman, 2018).

O artigo "*Man is to Computer Programer as Woman is to Homemaker? Debiasing Word Embeddings*" ( traz à luz questões discriminatórias que podem surgir no uso da geometria de gênero e preconceitos em incorporações de palavras que estejam mais perto de ela do que ele, e questiona até que ponto esses preconceitos geométricos concordam com a noção humana de estereótipo em palavras ocupacionais e avalia se a incorporação produz analogias que são julgadas como refletindo estereótipos pelos humanos e se o uso destas palavras podem causar aspectos discriminatórios em modelos de IA posteriormente. A conclusão do artigo se perfaz no sentido de alterar as incorporações de palavras neutras em termo de gênero e removendo as suas associações de gênero e além disso descobriu-se que palavras específicas de gênero possuem preconceitos adicionais. Assim, ao igualar avô de avó fora do gênero, recomenda-se que o lado da neutralidade no que tange as palavras chaves contribua para a diminuição de preconceitos. Este raciocínio de base se faz muito importante na elaboração de um *datasheet* pois, este documento introdutório poderá mitigar ou eliminar muitos vieses discriminatórios relacionados a gênero, algo que pode resolvido por uma elaboração de um algoritmo de incorporação "suave" (Bolukbasi *et al*, 2016).

Levando em consideração os aspectos de deontologia apresentados segundo Kant, a elaboração de um *datasheet* deve considerar aspectos éticos que possam estar presentes no conjunto de dados. No artigo de Mohammad (2022), essa discussão ganha até mesmo uma espécie de trocadilho, uma vez que é proposta a apresentação de uma “folha [de dados] ética” (Mohammad, 2022). Se um desenvolvedor quiser, logo no início, desenvolver uma tarefa de IA deverá fazer o documento (*datasheet*) o qual constará as informações éticas e de viés no conjunto de dados utilizados para treinar o modelo de IA. *O* documento deve também possuir a consideração ética de teorias, metodologias, recursos e práticas comuns utilizadas na construção de sistemas de IA para a tarefa sempre visualizando o impacto social (Gebru *et al.,* 2021; Mohammad, 2022).

[5] Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track.

Neste mesmo sentido, para a elaboração do datasheet com aspectos éticos no conjunto de dados deve, segundo o autor, ao estruturar tarefas e fazer escolhas sobre dados, métodos e avaliações, é crucial revelar suposições ocultas, pois isso levanta considerações éticas únicas ou especiais. Essas escolhas devem apresentar dimensões e pontos de escolha relevantes, incluindo compensações para várias partes interessadas (Mohammad, 2022). Também é importante listar estratégias comuns de mitigação de danos e comunicar as implicações sociais a pesquisadores, desenvolvedores e ao público em geral (Nikolinakos, 2023). Portanto, entende-se que uma boa ficha ética deve considerar várias questões, como o motivo para automatizar uma tarefa e até que ponto o comportamento humano seria interessante nessa tarefa. Deve-se analisar os fundamentos teóricos e como a decisão automatizada afetará diferentes grupos de pessoas. Além disso, é necessário considerar a descrição geral e propósitos do conjunto de dados, especificações técnicas (arquitetura do modelo), curvas e gráficos do desempenho do modelo, condições de uso e notas aplicativas. Finalmente, deve-se avaliar se a tecnologia está beneficiando a todos ou apenas aqueles com poder e vantagem (Mohammad, 2022).

Os autores documentaram que as limitações nos dados de treinamento conduzem a problemas éticos nos sistemas de Processamento de Linguagem Natural (PLN) resultantes. Esses sistemas, ao serem treinados com dados de linguagem natural, acabam aprendendo os preconceitos já existentes entre os falantes desses dados. Representações vetoriais típicas da semântica lexical captam preconceitos sobre gênero, raça, etnia e religião (Bolukbasi *et al.*, 2016, p. 598). Além de captar esses preconceitos, os algoritmos de aprendizado de máquina podem amplificá-los. Dessa forma, entende-se que os dados linguísticos podem sempre conter aspectos discriminatórios, tornando difícil a construção de um sistema de Processamento de Linguagem Natural (PNL) completamente imune ao preconceito. Portanto, é necessário buscar estratégias adicionais para mitigar as deficiências éticas decorrentes de dados imperfeitos. Os documentos (*datasheet*) de dados são utilizados como uma ferramenta para combater preconceitos e injustiças em sistemas de dados, verificando a existência de preconceitos ou vieses pré-existentes. Quando um sistema é examinado pelas declarações dos dados nos quais foi treinado, é

possível avaliar se as populações representadas nesses dados estão sendo de fato representadas de forma precisa e sem preconceitos (Bender e Friedman, 2018, p. 589-594).

A tese de doutorado de Boyd (2021) que esclarece que grandes conjuntos de dados colhidos frequentemente refletem o que seus criadores pretendem deduzir de uma determinada situação padrão. No entanto, esses dados podem apresentar uma alta demanda de inconsistências e preconceitos. Assim, a sociedade frequentemente acredita que delegar problemas sociais complexos aos computadores pode evitar fraquezas humanas, como inconsistências, preconceitos ou limitações na velocidade de processamento.

Os engenheiros podem super mostrar informações raras importantes, subamostrar casos majoritários e manipular exemplos de forma a garantir que estes dados são suficientes para que os dados de treinamento sejam os mais precisos possíveis. No caso da ferramenta de prática e intervenções no produto existe uma lista de verificação para garantir que os dados de treinamento sejam representativos das classes minoritárias, é uma forma de garantir que os dados de treinamento sejam o menos enviesados possíveis e que considerem questões éticas ao serem colocados em produção, neste caso em concreto o *datasheet* deve constar as informações éticas e de vieses do conjunto de dados usados para treinar o modelo (Boyd 2021).

Os *datasheets* desempenham um papel crucial ao fornecer informações detalhadas e padronizadas sobre componentes, modelos de IA e conjuntos de dados, garantindo transparência e reprodutibilidade. Considerar fatores éticos em *datasheets* é essencial para identificar e mitigar vieses, promover a utilização responsável da tecnologia e assegurar que os impactos sociais e éticos dos modelos de IA sejam devidamente avaliados. Isso não apenas aumenta a confiabilidade dos produtos tecnológicos, mas também protege usuários e sociedade de potenciais consequências negativas, fomentando um desenvolvimento tecnológico mais justo e inclusivo.

## 2.3 Desafios cognitivos para uma decisão com IA confiável e responsável

Estudos revelaram que o *Google* cometeu erros graves ao rotular uma imagem de duas pessoas negras como "gorilas" e ao exibir anúncios de empregos com altos salários preferencialmente para homens em vez de mulheres. Da mesma forma, o *Facebook* foi encontrado aplicando vieses de gênero ao anunciar oportunidades de emprego. Esses incidentes destacam problemas significativos em áreas de alto risco, como serviços financeiros, saúde e justiça criminal, onde a IA, como no caso do COMPAS, demonstrou propensão a classificar de forma errônea réus negros como riscos futuros em comparação com réus brancos (Cheng, Varshney e Liu, 2021).

Estes desafios cognitivos para o alcance de uma IA confiável, se mostram palpáveis no que tange a complexidade de aprendizado de robôs no âmbito doméstico. Isto é, ao contrário dos robôs industriais que realizam tarefas repetidas, os robôs domésticos ou mordomos se deparam com uma gama de tarefas incertas ao longo do dia, cabendo ao sistema de ML da IA entender como as coisas funcionam e alimentar o sistema preditivo para que o sistema faça escolhas eficazes e seguras. Deste modo, entende-se, de maneira superficial, que os sistemas de IA poderão decidir de forma complexa podendo até mesmo convencer e interferir na vontade das pessoas sobre determinada situação (Hartmann Peixoto, 2020b).

Outra situação que demonstra um forte desafio cognitivo para a construção de uma IA confiável são exemplificados pela pesquisa do MIT-IBM Watson AI Lab sobre a funcionalidade NeurIPS. Este algoritmo analisa coleções de textos e estrutura a modelagem de livros por tópicos representativos, em vez de palavras-chave. Além de ser extremamente rápido no processamento, comparando 1.720 pares de livros em um segundo, o algoritmo oferece uma visão ampliada e precisa sobre um tema. A combinação de classificação de interesses e análise temática pode impactar significativamente a autodeterminação das vontades (Hartmann Peixoto, 2020b). Neste tocante, a intersecção da tecnologia e dos valores humanos no domínio da inteligência artificial  é uma área crítica de estudo pois envolve a

problemática de valores éticos, morais e sociais que orientam o comportamento humano e as capacidades dos sistemas de IA (Ayinla *et al*.,2024).

Neste mesmo sentido, é necessário aprofundar no tema e visualizar a pirâmide de responsabilidade social da IA para entender como estruturar uma IA confiável e responsável. Com este estudo, procura-se diminuir os desafios cognitivos e ajudar os profissionais da tecnologia e demais investigadores a cumprir com a solidificação destes conceitos. Os quatro elementos da pirâmide da responsabilidade social da IA são: responsabilidade funcional, responsabilidade jurídica, responsabilidade ética, e responsabilidade filantrópica (Cheng, Varshney e Liu, 2021).

A responsabilidade funcional de um sistema de IA é garantir que ele opere de maneira eficiente e maximize os lucros, conforme as expectativas legais e sociais. Legalmente, a IA deve obedecer às leis e regulamentos, enquanto eticamente deve agir de forma justa e prevenir impactos negativos. Para cumprir com as finalidades éticas, os sistemas de IA precisam funcionar de uma forma consistente com as expectativas sociais, sem comprometer as expectativas funcionais. Sob aspecto prestigiado, ao invés da IA aumentar os desafios sociais, espera-se que a contribuição dela seja para mitigá-los (Cheng, Varshney e Liu, 2021).

Existem outros autores que conceituam a IA socialmente responsável, que é de extrema importância para o correto entendimento de como seria a sua aplicação e desenvolvimento de forma ética e responsável. Para tanto, se faz necessário entender os conceitos fundamentais da IA confiável, que envolve sistemas que atingem seu potencial quando há confiança em seu desenvolvimento, implantação e uso (Thiebes, Lins e Sunyaev, 2021). Quanto à a IA robusta, que se refere à capacidade de lidar com erros e entradas incorretas. A IA confiável também inclui aspectos de verificabilidade, explicabilidade e segurança (Singh, Vatsa e Ratha, 2021). Em fim, a IA segura deve ser implantada de forma a não prejudicar a humanidade , enquanto a IA centrada no ser humano melhora continuamente com a contribuição humana, proporcionando uma experiência eficaz entre humanos e robôs (Cheng, Varshney e Liu, 2021).

Destarte, para mitigar os problemas advindos do desenvolvimento da IA em que há o aparecimento de vieses, os algoritmos de IA socialmente

Responsáveis (SRAs) tem o papel de priorizar as necessidades de todas as partes interessadas, especialmente os usuários marginalizados e desfavorecidos, para garantir decisões justas e confiáveis. Isso inclui proteger e informar os usuários, prevenir e mitigar impactos negativos, e maximizar os benefícios a longo prazo. Os algoritmos de IA socialmente responsáveis são projetados para receber *feedback* contínuo dos usuários, assegurando o cumprimento dos valores sociais esperados. Assim, os objetivos funcionais, como a maximização dos lucros, e sociais, como a transparência, são integrados nos algoritmos de IA, visando equilibrar responsabilidade social e objetivos empresariais (Cheng, Varshney e Liu, 2021). Neste diapasão, Cowls *et al.* (2019) escreve o texto sobre a importância de desenvolver a IA para bem social incluindo a ética e sete fatores que apoiam as melhores práticas no seu desenvolvimento (Cowls *et al,* 2019 )

Os desafios cognitivos para a estruturação de uma IA confiável e responsável está intimamente ligado a sua estruturação de forma ética. Nesta mesma perspectiva, mostra-se plausível a discussão a respeito da IA ligada ao direito, apesar de ser muito complexa, uma vez que pela quantidade de volume de dados e toda a questão de ML se faz prudente entender que uma matéria estruturada a partir da linguagem humana e das relações sociais complexas pode não ser suficiente na aplicação e regulamentação no que tange às demandas que a IA traz (Cowls *et al.,* 2019; Hartmann Peixoto, 2020b).

Em sequência, dilemas a serem resolvidos no campo da ética surgem frente às possibilidades da IA, já que o direito está centrado nas ideias do antropocentrismo, principalmente nas decisões jurídicas (raciocínio jurídico), e a IA está estruturada em um raciocínio exato e, indo mais além os referenciais éticos ligados ao direito nem sempre são os mais atuais para a resolução dos impasses éticos advindos da IA (Cowls *et al.,* 2019; Hartmann Peixoto, 2020b). De acordo com Ayinla *et al.* (2024), os principais desafios éticos no desenvolvimento da IA são ligados à privacidade, preconceito e responsabilidade.

Assim, Madhavan *et al*. (2020) analisam os esforços do governo dos EUA para integrar a inteligência artificial nos quadros jurídicos, éticos e regulamentares existentes. O estudo destaca a importância das políticas e da governança na maximização dos benefícios da IA, enquanto aborda os riscos e

preocupações sociais. A participação dos bolsistas da Associação Americana para o Avanço da Ciência (AAAS) em questões de IA sublinha a necessidade de políticas de IA confiáveis e responsáveis, enfatizando a colaboração entre decisores políticos, pesquisadores e comunidades para criar políticas eticamente sólidas e socialmente benéficas.

## 2.4 Elementos para uma ética nos sistemas de IA no Direito

O autor Peters *et al.* (2020, p. 36) preceitua que para o desenvolvimento da ética na IA existem seis disciplinas centrais que se interconectam para que a IA seja desenvolvida na Filosofia, Sociologia, Psicologia, Design/HCI, Ciência da Computação e Engenharia. É cediço que, os novos requisitos para o desenvolvimento de uma IA cada vez mais forte, robusta e ética pressuponha a cada dia mais o envolvimento de conhecimentos multidisciplinares. Não existe uma única disciplina capaz de lidar de forma isolada com a matéria de ética. Entende-se que a experiência mais enriquecedora é aquela que envolve diversas matérias para ter resultados mais satisfatórios.

Além de se ter uma visão da interdisciplinaridade da matéria IA e ética é importante entender como a IA se relaciona com o Direito no contexto atual e os aspectos éticos levantados por esta matéria para a correta aplicação da IA (Carrillo, 2020). No âmbito do “direito público”, as tarefas realizadas por servidores públicos, magistrados, promotores de justiça e legisladores podem se utilizar da IA como ferramenta para obtenção de mais celeridade, redução de custos e segurança na execução de tarefas rotineiras destes (Dos Santos e Torres, 2024).

No Supremo Tribunal Federal (STF), o *software* chamado de “Victor’’, criado em parceria com a Universidade de Brasília (UnB), utiliza a tecnologia de aprendizado da máquina para identificar as teses de repercussão geral nos recursos extraordinários. Através deste sistema, o STF, afirma ter ganho uma eficiência na classificação destes recurso, que agora levam em média 5 segundos, acarretando a diminuição de 80% de recursos extraordinários no STF. Este sistema de IA repete padrões e regras estipuladas, não realizam atividades que dependem de análise minuciosa subjetiva do mérito da questão (Hartmann Peixoto, 2020a; Alencar, 2022).

O Conselho Nacional de Justiça (CNJ) na Resolução n.332/2020, que dispõe sobre a "ética, transparência e governaça na produção e uso de Inteligência Artificial no Poder Judiciário" é possível observar a preocupação com a necessidade de que os sistemas sejam compatíveis com os "direitos fundamentais" e atendam aos "critérios éticos de transparência, previsibilidade, possibilidade de auditoria e garantia de imparcialidade e justiça substancial" e ainda prevê no artigo 7 que as decisões judiciais apoiadas em ferramentas de IA devem preservar a pluralidade e a não discriminação, eliminando erros de julgamento decorrentes de vieses algorítmicos. E por fim, preceitua no artigo 7 que "a impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão" (Alencar, 2022).

É evidente que os sistemas que utilizam aprendizado da máquina por utilizarem um grande número de dados podem ter um resultado enviesado e muitas vezes discriminatório (Ferrer *et al.*, 2021). Pode-se ressaltar, questões de discriminação algorítmica de gênero em relação às mulheres, pesquisas mostram que as mulheres possuem mais dificuldade de obter empréstimos bancários junto a instituições financeiras uma vez que as instituições financeiras usam bancos de dados dos empréstimos concedidos ao longo dos anos para treinar os algoritmos; e a maioria dos empréstimos concedidos são para homens, assim o algoritmo poderá repetir este padrão rejeitando os pedidos de concessão de mulheres a empréstimos. Ainda que o banco de dados não traga justificativas que explique as decisões discriminatórias, o algoritmo poderá ainda se valer de outros dados com sugestivas lacunas de renda como, por exemplo, em períodos de gestação, criação de filhos, cuidados com parentes doentes dentre outros, circunstâncias estas que socialmente fazem parte da vida de mulheres e que se os algoritmos não forem bem treinados poderão desprivilegiar as mulheres neste tocante (Alencar, 2022, p. 37)[6].

[6] A Recomendação sobre a Ética da Inteligência Artificial n 90 e 91 da UNESCO preceitua que os Estados-membros devem assegurar que os estereótipos de gênero e os vieses discriminatórios não sejam traduzidos em sistemas de IA; em vez disso, tais elementos devem ser identificados e reparados de forma proativa(...). Os Estados-membros devem encorajar o empreendedorismo, a participação e o engajamento femininos em todos os estágios do ciclo de vida dos sistemas de IA, oferecendo e promovendo incentivos econômicos reguladores, entre

Assim, segundo o autor Alencar (2022, p. 39-44) é essencial que todos os sistemas de IA utilizados no Poder Judiciário brasileiro, ou fora dele, permitam auditoria, possuam testes satisfatórios, nível adequado de acurácia e sejam treinados por uma equipe diversa em termos de gênero, raça, etnia, cor e orientação sexual. Uma equipe diversa é crucial para identificar possíveis vieses na programação de algoritmos. Embora haja riscos, a IA no Poder Judiciário pode trazer maior eficiência, rapidez e redução de custos. Além disso, a implantação de algoritmos na Justiça brasileira pode aumentar a compreensão sobre as decisões judiciais no Brasil, fornecendo maior previsibilidade e segurança jurídica para as partes envolvidas e seus advogados. Desta maneira, no contexto de IA, ética e direito, é fundamental que a comunidade jurídica assegure que os algoritmos sejam desenvolvidos dentro de parâmetros ético-jurídicos, atuando como auxiliares na tomada de decisão. Para isso, é necessário implementar e defender mecanismos de auditoria, transparência, monitoramento e controle, incluindo a atuação de órgãos fiscalizatórios sobre esses sistemas.

O uso da IA no Poder Judiciário, tem um famoso caso no estado americano de Wisconsin, o caso de Paul Zilly, que revelou que o algoritmo de pontuação de risco considerou a prisão de um dos pais do acusado, um fator que promotores e juízes não utilizam para agravar penas. Além disso, ferramentas de mapeamento de crimes, que visam identificar áreas de alta criminalidade para alocação de recursos policiais, podem ser enviesadas por dados históricos. Esses dados, muitas vezes influenciados por práticas policiais preconceituosas, não refletem necessariamente a incidência real de crimes, mas sim a segmentação de grupos marginalizados (Angwin *et al.*, 2016). A IA deve ser vista como uma ferramenta para atingir fins humanos, e a sociedade deve definir seus objetivos (Alencar, 2022).

---

outros sistemas de apoio e incentivo, bem como políticas que visem a uma equilibrada participação de gênero na pesquisa em IA no meio acadêmico, representação de gênero em cargos de chefia, diretorias e equipes de pesquisa de empresas digitais e de IA. Os Estados-membros também devem assegurar que verbas públicas (para inovação, pesquisa e tecnologias) sejam canalizadas para programas e empresas inclusivas, com uma clara representação de gênero, e que investimentos privados sejam igualmente estimulados por meio de princípios de ação afirmativa (UNESCO, 2021).

Segundo o autor Elias (2017, p.10), é interessante observar o que acontece no Poder Judiciário nos Estados Unidos com relação a aplicação da IA:

> Como destacamos há pouco em relação ao Poder Judiciário, os estados de New Jersey e Wisconsin, nos Estados Unidos, por exemplo, já utilizam algoritmos para tomar decisões na Justiça Criminal sobre questões incluindo execução penal, condenação, fiança, reincidência, etc. Empresas privadas como é o caso da Northpointe desenvolveram as ferramentas com a utilização de algoritmos para essas finalidades. Algoritmos também são usados para analisar casos ou transações incomuns e investigar potenciais atividades fraudulentas (Elias, 2017, p. 10).

Neste mesmo sentido, o mesmo autor Elias (2017) preceitua que na esfera jurídica, já houve considerações sobre sistemas automatizados de raciocínio jurídico, utilizando algoritmos e inteligência artificial. Contudo, foi constatado que sua aplicação autônoma no Direito é inviável. A formação da convicção de um magistrado não se limita apenas a dados objetivos. A aplicação das leis transcende a simples aplicação de um conjunto de regras e jurisprudências. A autonomia de cada magistrado na formação de sua convicção em cada caso específico nunca deve ser ameaçada ou influenciada erroneamente por máquinas. Os juízes não interpretam a lei como robôs. A hermenêutica e a interpretação exercem um grande impacto no Direito. Os algoritmos ainda não conseguem replicar o raciocínio jurídico. É evidente que eles podem ser de grande auxílio para juízes, advogados, promotores e outros profissionais da Justiça. No entanto, nunca devem substituir completamente o elemento humano na equação. Os algoritmos são ferramentas e não devem ser considerados como substitutos integrais para o julgamento humano (Elias, 2017).

Assim, conclui-se que o direito, fundamentado em uma ética adequada, deve atuar como um mediador no processamento de dados e tecnologias, evitando tecnorregulação prejudiciais à humanidade. Nesse contexto, é crucial que o Direito oriente a criação e desenvolvimento de artefatos técnicos de maneira que sejam sensíveis a valores como privacidade, segurança e ética *by design*. Luciano Floridi compara o direito a uma meta-tecnologia que funciona como tubulações na era digital, conduzindo todo o conteúdo e ações (Magrani, 2019, p. 220).

## 2.5 A própria IA pode auxiliar no desafio ético

A IA pode auxiliar no enfrentamento de desafios éticos de várias maneiras, como na tomada de decisões, detecção e mitigação de viés, educação e conscientização. Ela pode oferecer sistemas de apoio à decisão, simulação de cenários, auditoria de algoritmos, programas de treinamento, análise de grandes volumes de dados, e monitoramento contínuo. No entanto, a implementação de IA também apresenta desafios éticos próprios, como responsabilidade, transparência, privacidade e equidade, que requerem uma abordagem multidisciplinar para serem adequadamente endereçados (Ayinla *et al.,* 2024; Olorunfemi *et al.*, 2024).

Em um cenário real no qual a IA pode auxiliar no desafio ético, pode-se citar o texto de Wang (2019b) que analisa o impacto da aplicação da IA na automação dos veículos e a consequente diminuição de impacto entre veículos. Neste contexto, apresenta uma abordagem de planejamento de movimento para veículos autônomos diante de situações de emergência onde a colisão é inevitável. O método proposto visa mitigar o acidente tanto quanto possível, assumindo que o planejamento de movimento recebe informações de faixa e velocidade desejadas, estados do veículo e informações de obstáculos e limites da estrada de módulos específicos. O algoritmo de controle preditivo do modelo é utilizado para o planejamento de movimento, integrando um fator de gravidade de colisão e um campo de potencial artificial para evitar obstáculos. Caso a evasão seja impossível, o objetivo é minimizar a gravidade da colisão. Além disso, a dinâmica do veículo é considerada um problema de controle ótimo. Simulações mostram que o algoritmo proposto pode evitar obstáculos e, se necessário, mitigar colisões e, além disso, em caso de colisão escolher qual seria menor a fatalidade, analisando os aspectos da vítima (ex: idoso ou criança) (Wang, 2019b).

É interessante notar que o planejador de movimento do algoritmo de mitigação de colisão proposto é capaz de gerar trajetórias com gravidade mínima de colisão, oferecendo uma técnica segura para veículos autônomos futuros. Em todas as simulações realizadas, o tempo de cálculo por intervalo foi

de aproximadamente 20 milissegundos, permitindo o uso do planejador em tempo real. Os algoritmos capazes de traduzir esta técnica segura para veículos autônomos devem ser construídos de acordo com os aspectos éticos do lugar em questão, como por exemplo no oriente os mais velhos são mais valorizados do que os mais novos, ao contrário do ocidente que tem uma cultura diferente nos quais a uma preferência aos mais novos. A construção da ética é desenvolvida através de fatores culturais (Wang, 2019b).

As tecnologias dos veículos autônomos têm o potencial de transformar os métodos de transporte convencionais, trazendo melhorias significativas em termos de segurança nas estradas. Isso se deve ao fato de que os erros cometidos por humanos são responsáveis por uma grande parcela, cerca de 94%, do total de acidentes registrados. Além disso, essas inovações podem aprimorar a experiência de deslocamento, permitindo que as pessoas usem seu tempo para trabalhar ou se entreter em vez de se concentrarem na direção do veículo. Adicionalmente, ao planejar a rota do tráfego ou ao estacionar de forma autônoma, os veículos autônomos têm o potencial de reduzir o tempo gasto em deslocamentos. Essas melhorias não apenas beneficiariam os motoristas, mas também tornariam o transporte mais acessível para uma gama mais ampla de pessoas, incluindo aquelas com diferentes habilidades físicas, contribuindo para sua independência e qualidade de vida (Wang, 2020).

Em 2011, os Estados Unidos registraram mais de 5,3 milhões de acidentes de veículos, resultando em cerca de 2,2 milhões de feridos e 32 mil mortes, com prejuízos financeiros significativos. A maior parte desses acidentes foi causada por fatores humanos, como excesso de velocidade, distração ao volante e consumo de álcool. A introdução de veículos autônomos pode reduzir significativamente esses acidentes ao minimizar a participação humana. Tecnologias como sistemas de alerta de colisão e assistência de visão lateral podem diminuir em até 33% os ferimentos e fatalidades. Além disso, os veículos autônomos podem beneficiar pessoas com deficiências visuais, físicas ou jovens, melhorando sua independência e qualidade de vida. Assim, até mesmo na construção da tecnologia de veículos autônomos a ética pode auxiliar na mitigação dos embates trazidos por esta tecnologia em situações críticas como acidentes (Wang, 2020).

A IA pode auxiliar no desafio da promoção da ética na área da saúde, segundo os autores Bohr e Memarzadeh (2020, p. 29), assim a empresa *Deep Genomics*, que está desenvolvendo tecnologia para identificar o sequenciamento genético das pessoas, usa IA em todos os processos de descoberta de medicamentos e desenvolvimento, incluindo a descoberta de alvos, otimização de leads, avaliação de toxicidade e *design* de testes inovadores, a fim de identificar possíveis doenças futuras. Num exemplo de prognóstico de doenças de forma antecipada, a IA é usada para correlacionar sintomas e biomarcadores de registros médicos eletrônicos (EMRs) com a caracterização e prognóstico da doença, desde diagnósticos até tratamento. Numerosas condições hereditárias podem manifestar sintomas sem um diagnóstico preciso, e a interpretação de dados do genoma completo permanece desafiadora devido à diversidade de perfis genéticos. No entanto, a medicina de precisão pode oferecer abordagens para aprimorar a detecção de mutações genéticas através do sequenciamento completo do genoma e do uso de inteligência artificial (Bohr e Memarzadeh 2020).

Na mesma perspectiva, ressalta-se o desenvolvimento de medicamentos com a participação da IA e sua grande contribuição para a área da saúde, o processo de fazer um medicamento deve contar com processos rigorosos que obedecem a preceitos éticos. Com efeito, a descoberta e o desenvolvimento de medicamentos é um processo longo e caro, levando mais de 10 anos desde a identificação de alvos moleculares até a comercialização de um medicamento. Falhas durante esse processo têm grande impacto financeiro, e a maioria dos candidatos falha durante o desenvolvimento. Assim, a IA pode acelerar o processo de descoberta de medicamentos, reduzir custos e aumentar a precisão ao identificar novos compostos terapêuticos. Utilizando algoritmos de aprendizado de máquina, a IA pode analisar grandes volumes de dados biomédicos para prever a eficácia e a segurança de novos medicamentos. Além disso, a IA pode ajudar a minimizar vieses e garantir que os medicamentos desenvolvidos sejam seguros e eficazes para uma ampla variedade de populações, promovendo a transparência e a responsabilidade no processo de desenvolvimento de medicamentos. A IA generativa pode revolucionar a pesquisa farmacêutica ao criar e otimizar estruturas moleculares com propriedades específicas desejadas para novos medicamentos. Isso não

apenas enriquece o processo de descoberta de medicamentos, mas também acelera significativamente o desenvolvimento de novas terapias (OCDE, 2019).

Nesta mesma perspectiva, uma aplicação notável da IA e visão computacional na tecnologia cirúrgica é melhorar certos aspectos e habilidades na cirurgia, como suturas e nós. É crucial que os cirurgiões participem ativamente no desenvolvimento dessas ferramentas, assegurando assim sua relevância e qualidade clínica e facilitando a transição do laboratório para o ambiente clínico, bem como respeitando a ética (planejamento cirúrgico, assistência robótica, tomada de decisão, monitoramento intraoperatório e educação e treinamento) na promoção do desenvolvimento da IA nas cirurgias. Isso é especialmente importante para cirurgias de emergência, onde a precisão e eficácia das ferramentas podem ser decisivas para o resultado do procedimento (Bohr e Memarzadeh 2020).

O robô autônomo de tecido inteligente (STAR) desenvolvido pela Universidade Johns Hopkins é um exemplo disso, demonstrando capacidade de superar cirurgiões humanos em certos procedimentos cirúrgicos, como anastomose intestinal em animais. Embora um cirurgião robótico totalmente autônomo ainda seja um conceito distante, aprimorar diferentes aspectos da cirurgia por meio da IA tem sido objeto de interesse para os pesquisadores. Por exemplo, um grupo do Instituto de Tecnologia da Informação da Alpen-Adria Universität Klagenfurt está utilizando vídeos de cirurgias como material de treinamento para identificar intervenções específicas realizadas pelo cirurgião. O algoritmo desenvolvido reconhece a probabilidade de intervenção, assim como a região específica do corpo, quando um ato de dissecção ou corte é realizado nos tecidos ou órgãos do paciente. Tais algoritmos são naturalmente baseados em um amplo treinamento com vídeos e podem ser extremamente úteis em procedimentos cirúrgicos complexos ou em situações onde um cirurgião é inexperiente (Bohr e Memarzadeh 2020).

Em outra perspectiva não menos interessante, é importante analisar as casas inteligentes que utilizam a IoT. Isso exemplifica a utilização da IA para fins éticos, ampliada pelo uso de diferentes sensores de monitoramento que facilitam a vida dos moradores, como pacientes com demência. Esses sensores são usados para detectar comportamentos anormais em várias áreas da casa. Por exemplo, eles disparam um alarme caso o paciente esqueça de

desligar o fogão ou fechar uma janela em dias chuvosos. As informações são transmitidas para um aplicativo que processa os dados ou os carrega para a nuvem, onde são utilizados pelo aprendizado de máquina. Isso permite que, se necessário, parentes ou profissionais de saúde sejam alertados sobre os pacientes em estado de demência que utilizam esses aplicativos (Bohr e Memarzadeh, 2020, p. 47). Nesse contexto, "as interações entre os profissionais da saúde e as tecnologias de análise de dados e ML durante os atendimentos são um solo fértil para o desenvolvimento de modelos preditivos", estes modelos preditivos contribuem para a rotina médica e antecipação de tratamentos e diminuindo as taxas de pacientes que evoluem para estágios mais críticos. (Amaro Jr *et al*., 2020).

A Inteligência Artificial também pode auxiliar no desafio ético da construção da democracia, a democracia não pode se abster de ser sujeita às regras de uma sociedade regida por leis, principalmente em razão de vivermos tempos em que a democracia está sob pressão de populistas e ditaduras. Não se pode ter nenhum tipo de tecnologia que seja maior que as regras estabelecidas à uma democracia, isto se aplica a IA ou a internet (Nemitz, 2018).

Os trabalhos de ética para a IA podem ser precursores do direito, no entanto, não podem substituir as leis, pois não possuem legitimidade democrática e natureza vinculativa para tanto. Outro ponto importante é o fato de existir uma “ética que vai além do que a lei exige”: a ética intraempresarial, esta ética é vista pelos engenheiros e líderes de grandes empresas como uma coisa boa, se incluir o princípio integral da lei do país, e for além disso em termos de orientações de interesse público da empresa, por exemplo. Muito do que “os evangelistas” do Vale do Silício afirmam que as empresas de tecnologia estão fazendo em termos de coisas boas é bem-vindo e não exigido pela lei, mas nenhuma ética pode absolver da obrigação de cumprir a lei e de respeitar e apoiar o processo democrático e todas as outras regras da democracia constitucional (Nemitz, 2018).

O princípio do essencialismo, discutido acima, nos direciona em relação a quais as decisões criadas pela IA devem ser estabelecidas na lei. Uma vez que a IA cause um desafio ou problema a ser solucionado que afete algum direito fundamental, é necessário se perguntar se já existe uma lei para

salvaguardar este direito de forma proporcional e suficiente. A IA avança e o direito vem logo após dando a cobertura necessária para os casos em que não há a cobertura para um determinado direito de forma que não coíba o desenvolvimento de novas ferramentas tecnológicas responsáveis no tocante à questão ética.

Nesta mesma lógica da discussão da ética na IA, como um norteador e construtor de uma democracia forte, temos a discussão: uma máquina poderia emitir uma opinião sobre a democracia? Nemitz faz uma análise::

> Por outro lado, no discurso democrático, é importante saber se a contraparte na discussão é um ser humano ou uma máquina. Se as máquinas pudessem participar no discurso político sem serem identificadas como tal ou mesmo personificarem-se como seres humanos sem sanção, isso equivaleria a uma importante distorção do discurso, insustentável na democracia. Nenhuma lei garante que sejamos informados se as máquinas dialogarem conosco no contexto político. (...) [O] princípio da essencialidade prescreve que a transparência deve ser criada por lei quanto a saber se uma máquina ou um ser humano está a falar. (Nemitz, 2018, p. 11).

Neste sentido da correlação de IA, ética e democracia, Nemitz (2018, p.13) destaca que programadores de IA, desde o início do desenvolvimento de programas, devem considerar como seu trabalho pode afetar a democracia, os direitos fundamentais e o Estado de direito. Para Timmers (2021, p. 76) os desafios éticos relacionados à inteligência artificial na democracia, destaca a importância de reconhecer esses desafios e criar visibilidade sobre como a IA pode ser uma ameaça e como é implementada de ponta a ponta. Percebe-se que o desafio ético vai além de um único criador de IA antiético ou eticamente inconsciente, de uma única empresa presente nas redes sociais ou de um governo ou agência de segurança cibernética. Não há respostas fáceis, pois isto diz respeito à relação mal compreendida entre construções sociais e tecnológicas, a relação “código – lei”.E termina falando que pode parecer de interesse sobretudo acadêmico mas não é verdade, estamos construindo a soberania e a democracia na era da IA.

Ainda segundo Timmers (2021, p. 53) o texto “IA desafiando a soberania e democracia” o autor ressalta que apesar das discussões envolvidas serem cruciais para o amadurecimento de uma democracia, a identificação rápida de ameaças emergentes e o monitoramento dos componentes de *hardware* e

*software* em infraestruturas críticas dependem da IA. A IA ajusta firewalls, impede a disseminação de ataques e combate variações de *malware*. Em situações críticas, pode ser necessário desativar partes de serviços essenciais, como redes elétricas, em segundos, o que implica decisões delegadas à IA, mesmo com impactos significativos nos seres humanos (Nemitz, 2018; Timmers, 2021).[7]

A Inteligência Artificial pode auxiliar significativamente no desafio ético do seu desenvolvimento, promovendo práticas transparentes, responsáveis e justas. No setor da saúde, a IA aprimora diagnósticos e tratamentos, assegurando equidade e precisão. No sistema judicial, a IA aumenta a eficiência e minimiza vieses, promovendo justiça. Em segurança cibernética, identifica e mitiga ameças em tempo real, protegendo dados e privacidade. Para veículos autônomos, contribui para a segurança e acessibilidade. No fortalecimento da democracia a IA pode desempenhar um papel de promover práticas éticas, de transparência e de responsabilidade. Ao ser utilizada na administração pública, a IA pode melhorar a eficiência dos processos governamentais, reduzir a corrupção e aumentar a prestação de contas, sendo imprescindível que a IA seja implementada com rigorosos padrões éticos e regulatórios para salvaguardar os direitos fundamentais e consequentemente irá fortalecer a confiança pública na democracia. Em todos os setores, é fundamental que a IA seja desenvolvida com um foco robusto em transparência, responsabilidade e justiça, assegurando que suas aplicações beneficiem a sociedade de maneira eticamente responsável.

---

[7] **Monitoramento e Transparência**: Algoritmos de IA podem ser usados para monitorar processos governamentais e eleitorais, garantindo transparência e minimizando a corrupção. Isso inclui rastrear doações de campanha e despesas públicas, permitindo maior fiscalização e responsabilidade. **Detecção de Desinformação**: IA pode identificar e mitigar a disseminação de fake news, assegurando que os cidadãos recebam informações precisas e confiáveis. Isso é crucial para manter a integridade das informações que influenciam a opinião pública e as decisões eleitorais. **Inclusão e Acessibilidade**: Ferramentas de IA podem ser desenvolvidas para tornar informações governamentais mais acessíveis a todos, incluindo pessoas com deficiências ou diferentes níveis educacionais. Isso garante que todos os cidadãos tenham igual acesso às informações necessárias para participar ativamente da democracia. **Participação Cívica**: Plataformas baseadas em IA podem facilitar a comunicação entre cidadãos e seus representantes, promovendo uma participação mais ativa e informada no processo democrático. Isso pode incluir a organização de fóruns de discussão e a coleta de feedback dos cidadãos sobre políticas públicas. **Análise de Dados para Políticas Públicas**: A IA pode ajudar a analisar grandes volumes de dados para identificar necessidades sociais e avaliar o impacto de políticas públicas. Isso garante que as decisões sejam baseadas em evidências e atendam ao interesse público, promovendo uma governança mais ética e eficaz (Nemitz, 2018; Timmers, 2021).

# 3 Explicabilidade no processo de tomada de decisão

Dentre as atuações que perfazem a UNESCO, também está no seu âmbito de discussão a ética aplicada à IA. Tanto é que houve a publicação da "*Recommendation on the Ethics of Artificial Intelligence*" (Recomendação sobre a Ética da Inteligência Artificial), na qual conceitua-se a explicabilidade no processo de tomada de decisão como "a capacidade de tornar compreensíveis e fornecer informações sobre os resultados dos sistemas de IA". Assim, a UNESCO define a explicabilidade como a capacidade de descrever e justificar as decisões e comportamentos de um sistema de IA de maneira compreensível para os humanos (UNESCO, 2021).

A explicabilidade é um componente essencial da ética da IA, pois permite que os usuários e *stakeholders* entendam como e por que uma decisão foi tomada por um sistema de IA, promovendo transparência, confiança e responsabilidade. Ela envolve a comunicação clara dos processos internos do sistema, os dados utilizados, os critérios considerados, e as razões por trás das decisões, garantindo que os sistemas de IA operem de forma justa, não discriminatória e de acordo com valores éticos (UNESCO, 2021; Herzog, 2022).

## 3.1 Diferença entre transparência e explicabilidade nos sistemas de IA e sua aplicabilidade em casos práticos

Ainda segundo a Recomendação sobre a Ética da Inteligência Artificial, o propósito da transparência, de acordo com a recomendação número 39, é prover informações adequadas aos destinatários. Tais informações devem permitir a compreensão e promover a confiança. No contexto específico dos sistemas de IA, a transparência pode possibilitar às pessoas entender como um sistema de IA é implementado em cada estágio, observando o contexto e a sensibilidade do sistema. Além disso, a transparência pode englobar informações sobre os fatores que influenciam uma previsão ou decisão específica, incluindo se as garantias apropriadas, como medidas de segurança ou justiça, estão implementadas ou não. Em situações em que haja graves ameaças de impactos adversos aos direitos humanos, a transparência também

pode requerer o compartilhamento de códigos ou conjuntos de dados (Felzmann *et al.*, 2020; UNESCO, 2021).

No mesmo sentido, na recomendação da Unesco, a explicabilidade na IA tem o entendimento da entrada, saída e operação de cada componente dos algoritmos e como cada um contribui para os resultados dos sistemas. Assim, a explicabilidade está diretamente relacionada à transparência, pois os resultados e os subprocessos que levam a eles devem ser compreensíveis e rastreáveis, de acordo com o contexto. Os atores da IA devem garantir que os algoritmos desenvolvidos sejam explicáveis. Transparência e explicabilidade estão intimamente vinculadas a medidas adequadas de responsabilidade e prestação de contas, bem como à confiabilidade dos sistemas de IA (UNESCO, 2021).

A transparência na Inteligência Artificial refere-se à habilidade de compreender em profundidade o funcionamento dos algoritmos e das abordagens empregadas. Modelos que são intrinsecamente explicáveis são essenciais; contudo, é igualmente importante considerar aqueles que, embora não intrinsecamente explicáveis, auxiliam na identificação de vieses e na proteção da privacidade. A explicabilidade, portanto, torna os modelos, métodos e processos inteligíveis para os usuários, permitindo a validação dos resultados e proporcionando maior controle sobre o processo (Protásio, Faria e Hartmann, 2022). Isso não exige a revelação completa do código-fonte, mas sim a compreensão do processo realizado, gerando confiança nos resultados da IA. De forma semelhante é comum que algoritmos de aprendizado de máquina funcionem como "caixas-pretas", ou seja, as decisões, recomendações ou previsões alcançadas por esses algoritmos são difíceis de serem entendidas a partir de uma análise superficial (Hassija *et al.*, 2024). Portanto, é crucial a compreensão de como esses resultados são obtidos, de modo a garantir a proteção dos direitos fundamentais e a aplicação correta de um devido processo legal tecnológico. A transparência e a explicabilidade são meios para alcançar esses objetivos, permitindo o entendimento e a auditoria dos resultados obtidos por meio dos algoritmos de aprendizado de máquina (UNESCO, 2021).

Quando o assunto é a explicabilidade no processo de tomada de decisão, Olsen *et al.* (2019, p. 23) destacam que, mesmo em abordagens

algorítmicas, a explicabilidade não difere muito das tradicionais. Os diferentes métodos apresentados estabelecem limites que qualquer sistema de Decisão Automatizada por Máquina (ADM) deve superar sem revelar completamente o funcionamento interno. Por exemplo, em um modelo baseado em direitos humanos, o tomador de decisão deve esclarecer fatos relevantes, como a ausência de um atestado médico, diagnósticos similares e sistemas de ponderação aplicados. Por exemplo, na contestação de uma decisão desfavorável sobre uma doença grave. A abordagem do direito nacional e da UE exige uma explicação detalhada das razões factuais e jurídicas, especialmente quando há maior poder discricionário. Isso inclui critérios de avaliação de casos semelhantes e desvios de políticas, assegurando controle, transparência e proteção dos interesses individuais, aplicando-se também a decisões favoráveis (Olsen *et al.* 2019).

De acordo com Kirat *et al.,* (2023. p. 9), o termo "explicabilidade" refere-se à capacidade de tornar algo inteligível ou compreensível[8], abordando o "porquê" e o "como" de uma decisão ou processo. Isso implica fornecer explicações delimitadas ao objeto, destinatários específicos e determinar quem deve fornecer essas explicações e de que forma.

A legislação europeia e francesa exige que as decisões algorítmicas sejam explicáveis, sem definir modalidades específicas. Nos EUA, o direito à explicabilidade foi destacado em dois casos: na avaliação de professores, onde o juiz considerou essencial o acesso a equações e códigos-fonte, e no caso COMPAS, onde o Supremo Tribunal de Wisconsin pediu explicações gerais para avaliar a precisão das sentenças. No setor bancário, há uma obrigação de explicabilidade na concessão de empréstimos, exigindo razões específicas para recusas. Alguns estados americanos, como Washington, também

[8] “O significado daquilo a que a literatura académica em IA se refere como interpretabilidade ou explicabilidade de um modelo de IA é muito diferente do significado de uma explicação que é geralmente discutida noutros contextos sociais (...) De acordo com o nível de criticidade das aplicações e das ameaças aos sistema, deverão ser aplicados diferentes níveis de requisitos a determinar. Na verdade, a relevância de uma explicação está sujeita ao público-alvo: explicar a decisão ao utilizador final, a uma equipe técnica de engenharia ou a um organismo de certificação requer diferentes ferramentas e abordagens, e deve ser feito considerando tanto as limitações técnicas da interpretabilidade da IA e expectativas legítimas das partes interessadas. Para este fim, a definição de recomendações para conectar métodos técnicos de interpretabilidade e explicação daria um passo adiante na compreensão dos sistemas de IA” (Hamon, Junklewitz e Sanchez, 2020).

implementaram requisitos de explicabilidade em leis de reconhecimento facial (Hamon, Junklewitz e Sanchez, 2020).

Na União Europeia, o Artigo 22 do RGPD inclui requisitos de explicabilidade, especialmente para decisões automatizadas e criação de perfis, exigindo que as organizações informem os titulares dos dados sobre o uso de decisões automatizadas, fornecendo informações significativas sobre a lógica envolvida e as possíveis consequências. A abrangência dessa obrigação é limitada a sistemas que utilizam dados pessoais e a decisões individuais automatizadas.

Outros autores destacam que a explicabilidade também é diferenciada da interpretabilidade, o que pode gerar confusão na compreensão do significado das noções em diferentes domínios. A legislação francesa, por exemplo, exige explicações detalhadas para decisões administrativas negativas, mas permite amplas exceções para proteger segredos de Estado ou outros interesses. A exigência de explicabilidade tem como objetivo fortalecer a transparência e a confiança na administração, permitindo a revisão e contestação das decisões perante um tribunal. Em resumo, a explicabilidade é essencial para garantir que decisões automatizadas sejam transparentes e compreensíveis, permitindo que indivíduos e autoridades possam questionar e entender a lógica por trás dessas decisões, promovendo um maior controle e confiança nos sistemas algorítmicos (Kirat *et al.,* 2023).

Neste sentido, Kirat *et al.,* (2023. p. 11) explicita a diferença entre explicabilidade e interpretabilidade: alguns autores consideram as noções de "explicabilidade" e "interpretabilidade" intercambiáveis (Beaudouin *et al.*, 2020a, p. 8), enquanto outros fazem distinção. Ambos os termos visam tornar compreensíveis as decisões tomadas por algoritmos. Na IA, a interpretabilidade avalia globalmente o processo de decisão, como um algoritmo que toma uma decisão de maneira geral, enquanto a explicabilidade se divide em três níveis técnicos:

1) Descritiva aquela que reconstrói e rastreia cada passo que o algoritmo executa, útil para especialistas em codificação; é importante uma explicação descritiva do que o algoritmo fez ou faz e na maioria das vezes é interpretável apenas por especialistas em comunicação. 2) Lógica: Reconstrói as razões pelas quais um algoritmo executa certos passos, mostrando a

relação causal entre entrada e saída, essencial para verificação de segurança e robustez. 3) Argumentativa: fornece razões finais para as decisões do algoritmo, considerando dados e procedimentos utilizados, problemática em casos de "caixas pretas" onde a conexão lógica é difícil de estabelecer, como por exemplo (Herzog, 2022).[9]

Para finalizar, Kirat *et al.* (2023. p. 13) ressaltam que especialistas do HLEG, têm a explicabilidade como um dos quatro princípios éticos fundamentais para qualquer sistema de IA, juntamente com o respeito pela autonomia humana, a prevenção de danos e a justiça. Assim, de acordo com o HLEG, a explicabilidade está fortemente ligada à transparência, rastreabilidade e comunicação. O GPAN afirma que a explicabilidade é essencial para construir e manter a confiança dos usuários nos sistemas de IA, significando que os processos devem ser transparentes, e as capacidades e finalidades dos sistemas de IA devem ser comunicadas de forma clara às pessoas diretamente e indiretamente afetadas.

Neste sentido outras recomendações da OCDE sobre a IA responsável no que tange a explicabilidade são:

> i) crescimento inclusivo, desenvolvimento sustentável e bem-estar; ii) valores e justiça centrados no ser humano, iii) transparência e explicabilidade; iv) robustez, segurança e proteção; ev) responsabilização. No parágrafo 1.3. sobre “transparência e explicabilidade”, a OCDE considera que: “Os atores de IA devem comprometer-se com a transparência e a divulgação responsável em relação aos sistemas de IA. Para este fim, devem (...): i. promover uma compreensão geral dos sistemas de IA, . ii. conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho, iii. permitir que as pessoas afetadas por um sistema de IA compreendam o resultado e, iv. permitir que as pessoas afetadas negativamente por um sistema de IA desafiem o seu resultado com base em informações simples e fáceis de compreender sobre os fatores e na lógica que serviu de base para a previsão, recomendação ou decisão” (Kirat *et al.,* 2023, p. 13).

O autor Zuiderveen Borgesius (2018, p. 24), destaca que o RGPD impõe requisitos de transparência para decisões automatizadas. As entidades

[9] Um exemplo típico de explicabilidade argumentativa é o resultado de um procedimento de votação: o resultado depende dos votos emitidos pelos eleitores, mas também de como a maioria é computada com esse procedimento específico. Uma explicação argumentada sempre deve fornecer ambos. Este tipo de explicação torna-se problemático quando o algoritmo modifica a execução a cada vez dependendo do conhecimento acumulado de execuções anteriores ou quando o algoritmo inclui componentes para os quais é praticamente impossível fazer uma conexão lógica entre entrada e saída (caixas pretas) (Kirat *et al.* 2023. p. 11).

responsáveis pelo tratamento de dados devem informar os titulares sobre a utilização dessas decisões, explicando a lógica e as possíveis consequências. Deduz que, em certos casos, uma organização seria obrigada a explicar o uso de tomada de decisão por IA e a fornecer explicação de como este processo funciona. O RGPD tem sido objeto de muita análise acadêmica no tocante às suas regras sobre as decisões automatizadas, que prevê um “direito à explicação” das decisões individuais. No entanto, sua eficácia é questionada, pois muitas decisões automatizadas não são abrangidas, aplicando-se apenas a decisões baseadas exclusivamente em processamento automatizado.

Assim, baseando em um caso prático, se um funcionário de um banco recusar um empréstimo com base em uma recomendação de um sistema de IA, desde que esta decisão não seja inteiramente delegada a um sistema de IA, a disposição do RGPD[10] não se aplica. Por outro lado, A Convenção de Proteção de Dados (COE, 2018) é mais moderna no que tange a este assunto de explicação e concede um direito mais abrangente ao invés de se aplicar apenas a decisões baseadas exclusivamente em processamento automatizado, permitindo aos indivíduos solicitar o conhecimento do raciocínio subjacente ao processamento de dados quando aplicado a eles. No entanto, o autor finaliza dizendo que a extensão exata desse direito ainda não foi

[10] O Regulamento Geral de Proteção de Dados (RGPD) possui regras específicas sobre ‘’decisões individuais automatizadas’’, estas regras visam diminuir o risco de tomada ilegal de decisões com cunho discriminatório. ‘’O artigo 22 do GDPR, às vezes chamado de disposição Kafka, contém uma proibição em princípios de decisões totalmente automatizadas com efeitos legais ou similares significativos e se aplica, por exemplo, a práticas de recrutamento eletrônico totalmente automatizadas sem intervenção humana. A regra principal da disposição do RGPD sobre a tomada de decisão individual automatizada é a seguinte: O titular dos dados tem o direito de não ficar sujeito a uma decisão baseada exclusivamente no tratamento automatizado, incluindo a definição de perfis, que produza efeitos jurídicos que lhe digam respeito ou que afete significativamente de forma semelhante. Resumindo: as pessoas podem não estar sujeitas a certas decisões automatizadas com efeitos de longo alcance. O GDPR diz que as pessoas têm o ‘’direito de não estar sujeitas a’’ certas decisões. Mas é geralmente assumido que este direito implica, em princípio, uma proibição de tais decisões. Parafraseando ligeiramente Mendoza e Bygrave, quatro condições devem ser atendidas para que a disposição seja aplicada: (i) haja uma decisão, que se baseie (ii) exclusivamente (iii) no processamento automatizado de dados; (iv) a decisão tiver efeitos legais ou igualmente significativos para a pessoa’’. Um exemplo de decisões com ‘’efeito jurídico’’ seria uma decisão judicial, ou uma decisão relativa a um benefício social concedido por lei, como o pagamento de pensões. Um exemplo de uma decisão com efeitos ‘’similarmente significativos’’ seria um banco que nega crédito automaticamente e as autoridades de Proteção de Dados dizem que a diferenciação de preços on-line poderia ‘’afetar de forma semelhante e significativa’’ alguém, se levar a ‘’preços proibitivamente altos [que] efetivamente impedem alguém de certos bens ou serviços’’(Zuiderveen Borgesius, 2018, p. 22-24).

totalmente esclarecida na prática em nenhum país que tenha implantado a Convenção (Zuiderveen Borgesius*,* 2018).

Os modelos automatizados em sua totalidade são a vitrine quando o assunto é aplicação de IA e toda a questão que envolve a explicabilidade de suas decisões. Assim, embora o modelo automatizado seja o foco principal, analisar-se-á o modelo híbrido no contexto de decisões de casos jurídicos. No tocante, a casos jurídicos, decisões completamente automatizadas só podem ser aplicadas a formas simples de tratamento de casos jurídicos. O autor do texto defende que a explicabilidade no procedimento da administração pública deve ter o mesmo nível de exigência imposto pela legislação nos casos de decisões totalmente humanas. Portanto, o sistema que se concebe é baseado na ideia de que o suporte da Inteligência Artificial (IA) assume a forma de um sistema algorítmico que oferece esboços de decisões aos assistentes sociais humanos, destacando os três requisitos mencionados anteriormente (Olsen *et al.*, 2019).

Mantendo o exemplo das decisões jurídicas, exigir transparência algorítmica apoiadas pela IA, nesse caso, pode não abordar o requisito de explicabilidade de forma adequada. A decisão apoiada pela IA não deve ter um nível de explicabilidade mais granular do que a tomada de decisão humana antes de ser usada na administração pública, pois a tomada de decisões apoiada pela IA tem muito valor potencial, e a introdução desta tecnologia não deve ser evitada elevando o nível de explicabilidade acima do existente para decisões humanas (Olsen *et al.*, 2019).

Neste mesmo sentido, Olsen *et al.,* (2019, p.17) ressaltam que o princípio da explicabilidade, como é aplicado, constitui uma nova faceta da transparência algorítmica, suficiente para mitigar as preocupações relacionadas à Decisão Automatizada por Máquina (ADM). A França, diferentemente da Alemanha e dos países escandinavos, não possui um requisito geral de explicabilidade para decisões administrativas. Conforme o *Conseil Constitutionnel* determinou em 2004, o direito constitucional francês não impõe, por si só, um dever geral aos órgãos administrativos de justificar suas decisões. No entanto, além das sanções de caráter punitivo, as decisões administrativas devem ser fundamentadas de acordo com um estatuto de 1979 e o *Code des Relationships entre le Public et l'Administration* (CRPA) de 2016, que exige uma

explicação por escrito das considerações legais e factuais subjacentes à decisão.

Embora a França tenha incluído tardiamente o requisito de explicabilidade, foi pioneira na regulamentação do uso de tomadas de decisões automatizadas, com o art. 10 da *Loi Informatique et Libertés* de 1978 proibindo decisões que produzissem efeitos jurídicos para indivíduos baseadas exclusivamente em decisões automatizadas. O art. 22 do GDPR adota uma formulação semelhante, permitindo decisões totalmente automatizadas desde que certas salvaguardas sejam implementadas, embora estas tenham sido criticadas como insuficientes, especialmente em procedimentos que não envolvem "dados sensíveis".

Desse modo, entende-se que a justificativa por trás do requisito de explicabilidade é fortalecer a transparência e a confiança na administração, além de permitir sua revisão e contestação em tribunal. Essas explicações geralmente são exigidas apenas para decisões negativas e, mesmo assim, a lei permite que autoridades públicas invoquem a proteção de segredos de Estado ou outros interesses para evitar completamente uma explicação (Olsen *et al.,* 2019).

No que tange a explicabilidade no processo de tomada de decisões o Art. 41 da Carta dos Direitos Fundamentais da União Europeia (CFR) de 2000 prevê o direito a uma boa administração, que inclui no parágrafo 2 a “obrigação da administração de fundamentar as suas decisões”, adotada com sucesso na sequência de propostas do membro escandinavo afirma. Sua inclusão é uma concretização do art. 296 (2) Tratado sobre o Funcionamento da União Europeia (TFUE), segundo o qual “[o]s actos jurídicos devem indicar as razões em que se baseiam”, o que se aplica também às decisões administrativas. Art. O 41 do CFR vincula principalmente as instituições da UE, mas a mesma regra se aplica igualmente aos estados membros que implementam a legislação da UE. Geralmente, todos os atos unilaterais que geram consequências jurídicas – e se qualificam para revisão judicial nos termos do Art. 263 TFUE – exige uma explicação. Deve “conter as considerações de facto e de direito que determinaram a decisão”. Existe uma ligação clara entre o alcance do poder discricionário disponível e o âmbito do dever de fundamentar, ou seja, as decisões precisam de ser “mais profundamente fundamentadas quanto maior

for o poder discricionário". O requisito de explicabilidade foi ainda mais concretizado pelo Código Europeu de Boas Práticas Comportamento Administrativo, um documento de direito não vinculativo de 2002 destinado ao pessoal da Comissão Europeia, bem como pela jurisprudência da UE (Olsen *et al.*, 2019).

Neste mesmo sentido, a diferença entre cenários automatizados e não automatizados no contexto do Art. 22 do Regulamento Geral de Proteção de Dados (GDPR)[11], que se aplica à "Tomada automatizada de decisões individuais, incluindo criação de perfis". Segundo os autores, o Art. 22 estipula que o titular dos dados tem o direito de não ser sujeito a uma decisão baseada exclusivamente no tratamento automatizado, que produza efeitos jurídicos significativos, a menos que proibido por lei com "salvaguardas suficientes" ou por "consentimento direto". Essas salvaguardas incluem transparência na fase de entrada (informar e obter consentimento) e explicabilidade dos resultados (revisão da própria decisão). O GDPR prevê auditoria externa através das Autoridades de Proteção de Dados (APD) e estipula o direito de contestar a decisão e obter intervenção humana. Também proíbe o uso de "categorias especiais" de dados pessoais, exceto em circunstâncias específicas do Art. 9(2). Além disso, quanto ao requisito da explicabilidade de tal sistema, o GDPR exige que os titulares dos dados sejam informados sobre a "existência de tomada de decisão automatizada" e recebam "informações significativas sobre a lógica envolvida, bem como o significado e as consequências previstas de tal processamento para os dados (Olsen *et al.,* 2019).

Ao promover a transparência e a rastreabilidade, a explicabilidade ajuda a assegurar que os sistemas de IA operem de maneira ética e justa, respeitando a autonomia humana e prevenindo danos. A transparência refere-se à abertura com que as capacidades, finalidades e funcionamento dos sistemas de IA são comunicados, permitindo que os usuários e reguladores entendam como esses sistemas operam. A explicabilidade, por outro lado, foca em fornecer informações detalhadas sobre a lógica e os critérios usados nas decisões automatizadas. Juntas, essas práticas promovem a confiança, a responsabilidade e a justiça nos sistemas de IA, assegurando que as decisões

[11] General Data Protection Regulation.

sejam justificáveis e que os impactos sejam claramente comunicados aos afetados.

### 3.2 Supervisão humana sobre as decisões automatizadas

Atualmente, muitos defensores da ética na inteligência artificial argumentam que a supervisão humana é um princípio central para o desenvolvimento e implementação responsáveis da IA (Hickok, 2021). Por exemplo, a Comissão Europeia, em sua Comunicação de 2019, destacou a agência humana e a supervisão como o primeiro de sete requisitos principais para a confiabilidade das aplicações de IA.

Os riscos e desafios que a supervisão humana visa abordar incluem ameaças à autonomia humana, falta de transparência em modelos algorítmicos, preocupações com privacidade e proteção de dados, e o problema da discriminação (Green, 2022).

Além disso, as diretrizes éticas elaboradas pela Comissão Europeia do Conselho da Europa para a eficiência da justiça (CEPEJ) enfatizam a importância do controle dos usuários sobre a IA e a necessidade de garantir que a implementação da IA não prejudique o acesso à justiça, indo até mesmo ao ponto de sugerir o direito a um julgamento justo (Koulu, 2020).

O autor Koulu (2020, p 725) discute a supervisão humana de decisões automatizadas, ressaltando que, embora necessária, a clareza nem sempre é evidente. A supervisão humana pode ser crucial para garantir a conformidade legal e a proteção dos direitos fundamentais. No contexto da tomada de decisões jurídicas, essas funções são tradicionalmente atribuídas a juízes, assistentes sociais e burocratas na administração pública, parlamentares e outros decisores humanos que exercem discricionariedade. As instituições jurídicas são conhecidas pelas ações humanas e a tomada de decisão é exercida através do poder discricionário concedido pelo Estado. A importância dessa discricionariedade é destacada pela dificuldade de automatizar decisões legais complexas. A algoritmização oferece novas formas de entender a necessidade do elemento humano e as possíveis perdas ao substituir decisores humanos por supervisores[12].

[12] A Recomendação da UNESCO sobre a Ética da Inteligência Artificial número 63 ressalta que os Estados membros devem aumentar a capacidade do Poder Judiciário de tomar

Neste mesmo sentido, Koulu (2020, p. 728) coloca o conceito de “White paper”, numa lista quatro principais requisitos para aplicações de IA de alto risco, todos envolvendo diferentes formas de supervisão humana. No entanto, a ênfase na avaliação ex ante pode ser problemática, pois não aborda totalmente a necessidade de mecanismos *ex post* para proteção jurídica eficiente. O “White paper” explica que os requisitos legais obrigatórios para a IA precisam ser decididos como parte da concepção do quadro regulamentar. Em suma, a tomada de decisões jurídicas deve ser considerada um domínio de alto risco no qual pode ser necessária regulamentação jurídica rígida sobre a IA. Aqui, a supervisão humana torna-se uma ferramenta regulatória potencialmente importante.[13]

A Supervisão pode ocorrer através da supervisão humana individual e através da supervisão pública inclusiva, conforme preceitua a Recomendação sobre a ética na IA, documento produzido pela Unesco[14] (UNESCO, 2021).

---

decisões relacionadas a sistemas de IA, de acordo com o Estado de direito e em sintonia com o direito e as normas internacionais, inclusive no uso de sistemas de IA em suas deliberações, ao mesmo tempo assegurando que seja mantido o princípio da supervisão humana. No caso de o Judiciário utilizar sistemas de IA, são necessárias salvaguardas suficientes para garantir, entre outros, a proteção dos direitos humanos fundamentais, do Estado de direito, da independência judicial, bem como o princípio da supervisão humana, assim como o princípio da supervisão humana, assim como para garantir que o desenvolvimento e o uso de sistemas de IA no próprio Judiciário sejam confiáveis, orientados ao interesse público e centrado no ser humano (UNESCO, 2021).

[13] Os principais requisitos para aplicações de IA de alto risco listados no topo do “White paper” são as recomendações do HLEG, que incluem a supervisão humana.O exemplo de restrições baseadas no design é encontrado em carros sem motorista que transferem o controle para humanos em determinadas situações. A política emergente de IA da UE parece centrar-se na avaliação *ex ante*, em detrimento da marginalização da elaboração de mecanismos *ex post*. Esta abordagem pode colidir com os mecanismos *ex post*, com base nos quais normalmente é produzida a proteção jurídica, embora o tema mereça uma análise aprofundada. Ainda não está claro como a ênfase *ex ante* é capaz de fornecer soluções para a reconhecida falta de reparação eficiente. Os documentos políticos propõem a supervisão humana nas suas diferentes formas como o principal mecanismo *ex post* para garantir a proteção legal. No entanto, o mecanismo por si só tem pouca substância, deixando em aberto o que precisa de acontecer entre o sistema de IA e o superintendente humano para construir uma proteção significativa. Presume-se que a supervisão humana proporcione proteção, seguindo em parte o enquadramento do RGPD que lhe confere valor intrínsico como mecanismo processual que justifica o processamento automatizado. Koulu (2020, p. 728).

[14] Os Estados-membros devem garantir que seja sempre possível atribuir responsabilidade ética e legal em qualquer estágio do ciclo de vida dos sistemas de IA, assim como em casos de recursos judiciais relacionados a esses sistemas, a pessoas físicas ou a entidades existentes. A supervisão humana se refere, portanto, não apenas à supervisão humana individual, mas também à supervisão pública inclusiva, como for apropriado. É possível que, às vezes, as pessoas decidam confiar em sistemas de IA por motivos de eficácia, mas a decisão de ceder o controle em contextos limitados continua sendo de seres humanos, pois estes podem recorrer àqueles sistemas para tomar decisões e agir, mas um sistema de IA jamais poderá substituir a responsabilidade e a prestação de contas finais humanas. Como regra, decisões de vida e morte não devem ser transferidas a sistemas de IA (UNESCO, 2021, p. 22).

Algumas referências tratam das perspectivas jurídicas na tomada de decisão como híbrida entre humanos e algoritmos (Enarsson, Enqvist e Naarttijärvi, 2022). Nas palavras de Firlej e Taeihagh (2021, p. 6) destacam a importância do controle humano sobre as Armas Autônomas, com medidas adotadas pela administração militar dos EUA que demonstram o conceito de controle humano. Existem duas principais formulações deste conceito: controle humano desde o projeto e controle humano como um "dedo no botão". A primeira formulação envolve o projeto de sistemas que permitem aos comandantes e operadores exercerem julgamento humano apropriado no uso da força. A segunda formulação implica que o sistema deve permitir o controle humano direto, interrompendo ações ou buscando informações adicionais do operador humano quando necessário. As implicações operacionais do controle humano desde o projeto incluem garantir um nível significativo de confiabilidade e previsibilidade em relação à interface humana. Antes de serem implementadas, as armas devem passar por uma revisão legal e por rigorosos testes de verificação e validação (V&V) de *hardware* e *software*, além de testes operacionais e realistas (T&E). A revisão jurídica e os procedimentos de avaliação técnica são cruciais para obter a confiança necessária no nível de concepção das armas autônomas (Firlej e Taeihagh, 2021).

Neste mesmo estudo, a consequência operacional do controle humano, supervisão humana, é preceituada como "dedo no botão" é que devem existir diretrizes específicas para que um humano possa desativar o acionamento de uma arma. Segundo a diretiva, um operador humano pode anular a arma se o sistema não completar os combates de acordo com as intenções do operador. Entretanto, não há diretrizes claras sobre como essa função deve ser executada. Pode-se argumentar que o operador humano é um agente altamente experiente, portanto seria capaz de compreender as limitações do sistema. A pesquisa sugere que os humanos podem superestimar ou subestimar as capacidades de uma máquina e sofrem de "viés de automação", aceitando decisões da máquina mesmo quando incorretas. A delegação de tarefas às máquinas pode degradar as habilidades humanas, especialmente em aplicações de segurança. Portanto, a crescente dependência de capacidades autônomas exigirá mudanças nos conceitos de comando e controle para garantir maior confiança (Firlej e Taeihagh, 2021).

Os autores Sterz *et al.,* (2024, p. 2) sugerem que a supervisão humana eficaz não deve ser um fim em si mesma, mas sim melhorar a tomada de decisões em contextos de alto risco envolvendo IA. Eles destacam que humanos podem incorporar considerações éticas e normas sociais melhor do que sistemas baseados em IA, especialmente em casos únicos ou incomuns em contextos que afetam seres humanos. Além disso, associar o termo “eficaz” à supervisão humana mostra a esperança de que um ser humano possa contribuir efetivamente para uma situação conjunta juntamente com um sistema baseado em IA, permitindo assim decisões melhores do que as do ser humano ou do sistema sozinho. A supervisão humana eficaz pode contribuir para decisões mais seguras, precisas, confiáveis e justas, mitigando riscos para os direitos fundamentais. Além disso, ela pode evitar resultados ética e socialmente indesejáveis, melhorando a confiança na tecnologia.

Existem críticas no tocante a supervisão humana e o autor Sterz *et al.* preceitua que:

> Os críticos (em particular) apontam para a falta de evidências empíricas que mostrem que, a partir de agora, podemos integrar de forma confiável as habilidades humanas e do sistema de uma forma que leve a melhores decisões em comparação com as decisões que cada entidade produziria em isolamento. Por exemplo, parece haver desafios substanciais no que diz respeito a alcançar confiança adequada em sistemas baseados em IA. Os humanos podem confiar excessivamente nos resultados produzidos por sistemas automatizados, levando a situações em que não detectam resultados errados ou injustos. Noutras situações, os humanos podem ter pouca confiança nos sistemas baseados em IA ou uma confiança demasiado elevada nas suas próprias capacidades, o que leva as pessoas a ignorar os resultados do sistema realmente precisos ou justos (Sterz *et al.*, 2024, p. 2).

Assim, para construir uma supervisão humana eficaz em sistemas de IA de alto risco, a proposta do artigo 14, n. 1 da Lei da IA estabelece que esses sistemas devem ser projetados para permitir uma supervisão eficiente durante o uso. O objetivo principal, conforme definido no art. 14, n. 2, é prevenir ou minimizar riscos à saúde, segurança ou direitos fundamentais. Portanto, a eficácia da supervisão humana é essencial para mitigar esses riscos e garantir a confiabilidade geral dos sistemas. Este princípio atua como um limiar mínimo que deve ser alcançado para garantir a eficácia das medidas de supervisão implementadas. A Lei de IA, no artigo 14, n. 3, obriga o fornecedor a permitir

uma supervisão eficaz por parte do implementador, identificando "medidas apropriadas". O artigo 14, n. 4 especifica aspectos indicativos da eficácia dessas medidas, como: a) compreensão das capacidades e limitações do sistema de IA; b) consciência do viés de automação; c) interpretação correta dos resultados; d) capacidade de ignorar, anular ou reverter decisões do sistema; e) possibilidade de interromper o funcionamento do sistema. Essas exigências refletem as condições de eficácia na supervisão dos sistemas de IA. (Sters *et al.*, 2024).

A supervisão humana nos sistemas de IA é crucial para garantir a responsabilidade ética e a segurança. Ela assegura que as decisões automatizadas estejam alinhadas com valores humanos e normas sociais, permitindo a intervenção em casos de anomalias ou decisões problemáticas. Isso mitiga riscos para os direitos fundamentais e evita resultados prejudiciais. Isto proporciona uma camada adicional de confiabilidade, garantindo que os sistemas de IA operem de maneira justa, transparente e segura, reforçando a confiança pública e a legitimidade das tecnologias de IA.

### 3.3 Necessidade de uma tecnorregulação e as primeiras regulamentações

A análise da necessidade e existência de tecnorregulação nas tecnologias é crucial. Segundo o filósofo e eticista italiano Luciano Floridi, estamos entrando na "Era do *Design*" e devemos nos esforçar para que seja a "Era do bom *design*". Da mesma forma, Lawrence Lessig, professor de Harvard, afirmou em sua obra "Code 2.0" que "Code is Law". Essas declarações destacam como somos regulados pela arquitetura das plataformas digitais, tanto quanto pelo Direito, normas sociais e economia. É essencial entender a interação entre humanos e artefatos técnicos, que possuem funções e usos determinados na fase de *design* e desenvolvimento, imprimindo-lhes uma moralidade intrínseca que influencia a sociedade. Esses elementos têm um impacto crescente no âmbito social e político (Magrani, 2019).

Segundo Magrani (2019, p. 214) estamos vivendo a era da tecnorregulação sem que haja um norteador ético-jurídico para a preservação dos direitos constitucionais.

> Existe, no entanto, uma diferença crucial entre o debate jurídico sobre a automatização da década de 1990 e a discussão actual sobre o processamento automatizado. O salto tecnológico diz respeito à lógica envolvida nesse processamento automatizado. Este último considera cada vez mais uma classe particular de algoritmos que aumentam ou substituem a análise e a tomada de decisão por seres humanos, como ocorre com a disciplina de aprendizagem automática, ou seja, algoritmos capazes de definir ou modificar regras de tomada de decisão de forma autônoma. O segundo passo da nossa fenomenologia tem, portanto, a ver com o campo da IA e, mais particularmente, com a mudança crucial da automação para a autonomia artificial.

Magrani observa que a tecnorregulação já se estabeleceu, sendo utilizada principalmente para fins comerciais, frequentemente desconsiderando direitos constitucionais e regulações específicas, como o Marco Civil da Internet no Brasil, que enfatiza a liberdade de expressão no ciberespaço. Ele destaca que a ordem jurídica tem a função de regulamentar o comportamento humano usando técnicas, controlando assim outras técnicas que definem comportamentos humanos e processos tecnológicos. Dessa forma, Magrani concebe a lei como uma metatecnologia, regulamentando a evolução tecnológica e as interações humanas dentro deste contexto (Magrani, 2019).

Sobre a tecnorregulação Lawrence Lessig argumenta que a arquitetura da internet, composta por *hardwares* e *softwares*, com estrutura técnica e códigos que regem o seu funcionamento regula o comportamento humano de maneira eficaz, às vezes até mais do que o direito, a economia e as normas sociais. Ele destaca que os algoritmos que governam a arquitetura dos sites nos torna reféns dos algoritmos, regulando tanto o nosso comportamento quanto a lei e podem restringir o acesso à informação, a autonomia individual, a privacidade e a liberdade de expressão, criando obstáculos significativos para os usuários (Magrani, 2019).

O autor Pagallo *et al.,* (2015, p. 21) preceitua que os contextos de regulação pela tecnologia ou tecnorregulação podem concentrar-se em produtos/serviços, locais ou pessoas e podem ter implicações teóricas da regulação pela tecnologia. Conforme expressado:

> Hoje é comum vivenciarmos aplicações tecnorregulatórias em produtos e serviços (limitadores de velocidade em automóveis, filtragem de internet, serviços de informação personalizados etc.). A regulação pela tecnologia no domínio espacial pode ser referida como “inteligência ambiental”, onde o monitoramento de velocidade, câmeras CCTV, edifícios inteligentes, *software* de reconhecimento facial juntamente

> com tecnologias de computação vestíveis são os exemplos pioneiros. A implantação de ferramentas tecno-regulatórias visando pessoas é um cenário de futuro próximo, onde o curso de conduta desejado será ligado aos seres humanos, seja por meio de manipulação genética, administração de drogas ou por outros meios que possam ser usados para alterar o funcionamento do cérebro (Burk, 2002). Essa nova modalidade engendrada por meio de capacidades regulatórias de algoritmos, análises preditivas e seus efeitos recombinantes resulta na erradicação da base moral, do empreendimento normativo e da cadeia de causalidade conforme entendida nos sistemas jurídicos convencionais (Custers, 2013).

Assim, a questão não é se o direito pode ser reduzido à lógica formal e, portanto, totalmente incorporado em programas de computador. Ao contrário, máquinas computacionais têm uma normatividade intrínseca que pode eliminar deliberadamente certas escolhas de ação, ditando indiretamente o comportamento desejado. Em um cenário tecnorregulatório, o direito opera em um nível mais alto como "metatecnologia", surgindo normas jurídicas que não comandam diretamente a conduta humana, mas regulam o design de sistemas que limitam e governam a sociedade (Pagallo *et al.*, 2015).

Em um ambiente tecnorregulatório, o processo legal pode ser visto em três fases: direção (criação de regras), detecção e correção, que se tornam parte de um sistema interno opaco e incorporado. Quando a tecnologia é usada para guiar a conduta humana a fim de assegurar o cumprimento das normas, o caráter normativo do regime regulatório se perde, tornando impossível o descumprimento por meio de escolhas de design. Nesse contexto, a lógica do direito muda de "deveria/não deveria" para "pode/não pode", restringindo a capacidade dos indivíduos de raciocinar com as regras e tomar decisões. (Pagallo *et al.*, 2015).

A tecnorregulação, conforme destacado pelo autor, é criticada por reduzir o sistema jurídico a um conjunto de processos informacionais complexos que operam principalmente sobre correlações em vez de fatos. Isso desafia as reivindicações do direito à racionalidade, objetividade, neutralidade, autonomia e universalidade em diversos níveis e contextos. A legitimidade em uma democracia depende da capacidade dos indivíduos escolherem entre cursos de ação alternativos, ao invés de apenas decidirem como restringir sua liberdade (Nemitz, 2018; Pagallo *et al*., 2015)

Neste mesmo sentido, Magrani (2019, p. 218) acrescenta uma crítica para evitar um cenário de tecnorregulação completa que se sobreponha às regulamentações jurídicas vigentes — ou insurgentes —, bem como ao norteamento ético da esfera pública, para que a tecnologia não se regule por si só, mas busque uma regulação mais efetiva, a partir de uma pespectiva metatecnológica do Direito. Cita-se *ipsis litteris*:

> As maneiras diferentes em que podemos entender os propósitos normativos da lei como uma metatecnologia nos levam a expandir nossa visão jurídica tradicional. Por exemplo, uma abordagem metarregulatória no campo da automação legal deve nos permitir determinar se, e até que ponto, os legisladores não devem (ou não podem) delegar decisões a sistemas automatizados. Além disso, o enfoque deve ser sobre o impacto da tecnologia no Estado de Direito, no próprio papel da lei e em como a tecnologia compete com outros sistemas regulatórios. Devemos também prestar atenção aos princípios e valores que estão em jogo ao delegarmos a tomada de decisões a sistemas automatizados, nomeadamente com questões de interpretação e deliberação (Magrani, 2019).

Sob aspecto sincrônico, para que o direito funcione eficazmente como uma metatecnologia, é imprescindível basear-se em diretrizes éticas apropriadas à era da hiperconectividade, orientando o progresso tecnológico por uma perspectiva centrada no ser humano, ao mesmo tempo reconhecendo a influência potencial dos agentes não humanos. O intuito é estabelecer uma regulação mais eficaz para tecnologias autônomas, garantindo a proteção dos direitos fundamentais e a continuidade da espécie humana (Magrani, 2019).

Sobre a tecnorregulação, Bayamlioglu e Leenes (2018, p. 298) preceituam que a tecnorregulação se trata da influência intencional no comportamento dos indivíduos, ao incorporar normas em sistemas e dispositivos tecnológicos. Dependendo do contexto, esses modelos podem ser referidos como: ‘regulação pela tecnologia’, ‘normatividade tecnológica’, ‘software regulatório’, ‘direito como design’, ‘regulação baseada em design’ ou ‘regulação algorítmica’. As configurações tecnorregulatórias podem focar produtos/serviços, lugares ou pessoas, abrangendo uma infinidade de práticas e designs complexos. Atualmente, isso é comumente experimentado em controles de direção de carros, filtragem na internet, sistemas de gerenciamento de direitos digitais, lombadas, personalização de serviços de informação, etc. Crescentemente, a tecnorregulação também se manifesta em sistemas que tomam decisões sobre indivíduos e criam efeitos.

Neste contexto, Bayamlıoğlu e Leenes discutem três possíveis prejuízos da tecnorregulação baseada em dados, que podem comprometer o Estado de direito. Primeiramente, existe o risco de colapso do empreendimento normativo, quando algoritmos substituem regras estabelecidas. Em segundo lugar, a substituição de bases causais por cálculos correlativos compromete a justiça e a previsibilidade. Por fim, ocorre a erosão do empreendimento moral, onde a tecnorregulação pode resultar em decisões injustas ou inadequadas, exacerbando desigualdades e distorcendo o processo jurídico. Daí surge um significativo questionamento: poderá o “estado de direito” ser trocado pelo “estado da tecnologia”? Se ocorrer, certamente será acompanhado por discursos kafkianos, huxleyanos e orwellianos de distopia (Bayamlioglu e Leenes, 2018).

Assim, o mesmo autor ressalta que quando a tecnologia é usada para orientar a conduta humana com finalidade de implementação de certas normas, o direito sofre no que tange ao seu caráter normativo e também em relação à autonomia humana. Modelos baseados em dados que implementam regras ou quadros jurídicos podem minar a base moral do sistema jurídico, prejudicando o Estado de direito. Além disso, a natureza fluida dos sistemas baseados em dados, com normas sujeitas a mudanças contínuas, reduz a ancoragem moral. Embora garantam conformidade e eficiência, esses sistemas sacrificam a autonomia individual e a moralidade, privando os indivíduos da capacidade de raciocinar com as regras[15] (Bayamlioglu e Leenes, 2018).

Sobre a regulamentação da IA, a introdução de princípios éticos precedeu a criação de regras legais. A ética não prevê regras jurídicas

[15] Os modelos baseados em dados que implementam regras ou quadros jurídicos prejudicam o Estado de direito ao minar a base moral do sistema jurídico em muitas frentes. Em primeiro lugar, os argumentos neste contexto referem-se principalmente às noções de autonomia e dignidade humanas como os princípios superiores da ordem jurídica e política europeia desde o Iluminismo. Quando a tecnologia é usada para orientar a conduta humana com vista a garantir o cumprimento ou a implementação de certas normas, não só o carácter normativo do direito sofre erosão, mas também a autonomia humana e os fundamentos morais sobre os quais as próprias normas se baseiam. Especialmente quando é adoptada uma abordagem regulamentar ex-ante – não deixando margem para violação ou escolha quanto à forma de cumprimento – a nossa concepção do direito afasta-se do “deveria/não deveria” para o “pode/não pode”, o que significa que o que não é o legal também não pode ser feito.Esta natureza maleável e “fluida” dos sistemas baseados em dados torna-os particularmente atraentes como ferramenta regulatória, mas muito pouco atrativos do ponto de vista a perspectiva da moralidade do agente – eliminando as oportunidades de agir de uma forma moralmente por vontade própria e, assim, minando as condições exigidas para uma comunidade moral florescente (Bayamlioglu e Leenes, 2018, p. 309).

vinculativas, mas tem valor social e moral, especialmente em relação às novas tecnologias por duas razões principais: o impacto da IA nos seres humanos e a incapacidade inicial do legislador europeu de regular efetivamente essas tecnologias. Nos últimos anos, a União Europeia começou a traduzir esses princípios éticos em regras legais, que atualmente são propostas regulamentares e se tornarão vinculativas no futuro. As novas tecnologias, especialmente as equipadas com IA, devem ser regulamentadas para servir aos seres humanos e proteger aqueles que estão estruturalmente vulneráveis a elas (Gatt *et al.*, 2023).[16]

Neste contexto, em todos os níveis de regulação, amadureceu a consciência da necessidade de uma estratégia e regulamentação concreta da IA. Em maio de 2019, o Conselho da Europa adotou recomendações para maximizar o potencial dos sistemas de IA e prevenir impactos negativos nos direitos humanos. O seu principal interesse é verificar o impacto dos sistemas de IA nos direitos humanos nos sectores público e privado, impondo aos Estados-Membros a garantia dos direitos humanos através de informações precisas, transparência e supervisão independente e eficaz da conformidade tecnológica, mas sem criar obstáculos à identificação de responsabilidades e soluções em caso de violação destes direitos humanos. Na mesma direção, o Conselho da Europa nomeou um comité *ad hoc* para a Inteligência Artificial (CAHAI), em setembro de 2019, que visa avaliar o impacto da IA no indivíduo e na sociedade, bem como na *soft law* e na *hard law* existentes instrumentos que lidam com IA. Essas recomendações visam garantir direitos humanos com transparência e supervisão independente sem criar obstáculos para a

[16] A ética não prevê regras jurídicas vinculativas, mas tem valor social e moral e assume particular importância em relação às novas tecnologias por duas razões: (i) em primeiro lugar, pelo impacto da IA nos seres humanos; (ii) em segundo lugar, porque inicialmente (antes de 2021) o legislador europeu não estava preparado para regular o fenómeno através de uma regulamentação vinculativa capaz de realmente compreender as novas tecnologias e ser capaz de acompanhar o seu desenvolvimento. Durante os últimos anos (2021 e 2022), em vez disso, a União Europeia começou a traduzir estes princípios éticos em regras jurídicas, que agora constituem propostas regulamentares, mas que se tornarão regras vinculativas num futuro próximo. (...) Mais detalhadamente, as instituições internacionais, a União Europeia e os seus Estados-Membros sempre se basearam nos valores da dignidade humana, e nesta igualdade, liberdades e respeito pelos direitos humanos, solidariedade, abordagem está sendo aplicada também em relação à robótica e à inteligência artificial. (...) Na verdade, as tecnologias de IA não são neutras e, além de acarretarem uma série de vantagens, também implicam sérios riscos para os indivíduos que apresentam uma condição estrutural de vulnerabilidade em relação a elas. Portanto, as novas tecnologias, especialmente quando equipadas com IA, devem ser devidamente regulamentadas, para colocar estas novas tecnologias ao serviço dos seres humanos (Gatt *et al.*, 2023, p. 6).

identificação de responsabilidades. Em setembro de 2019, o Conselho da Europa formou o comitê CAHAI para avaliar o impacto da IA e, em dezembro de 2020, publicou um estudo de viabilidade para um quadro jurídico baseado nos direitos humanos, democracia e Estado de direito. Iniciativas da UE também foram desenvolvidas, destacando a resolução do Parlamento Europeu sobre regras de direito civil relativas à robótica e a proposta de Lei de IA em 2021 (Gatt *et al.,* 2023, p. 7).[17]

A tecnorregulação da IA envolve a integração de normas diretamente nos sistemas e dispositivos tecnológicos, moldando comportamentos humanos através do design e da arquitetura. Este modelo regulatório pode ser mais eficaz que as formas tradicionais de regulação, mas levanta questões éticas, especialmente sobre autonomia e dignidade humana. As primeiras regulamentações da IA na União Europeia, como o Conselho da Europa e o CAHAI, buscam equilibrar o desenvolvimento tecnológico com a proteção dos direitos humanos, garantindo transparência, responsabilidade e supervisão independente na implementação da IA. No próximo capítulo, estudar-se-á como a utilização dos princípios da ética podem ser usados para o desenvolvimento da IA responsável no estrangeiro e de que forma as contribuições estrangeiras se ajustam à realidade do desenvolvimento ético da IA no Brasil.

## 3.4 Principais Iniciativas Multilaterais

No contexto do desenvolvimento e aplicação da Inteligência Artificial, a necessidade de estabelecer diretrizes éticas e legais que orientem o uso responsável dessa tecnologia é um desafio global. Diversas organizações internacionais têm se mobilizado para criar marcos regulatórios que assegurem que a IA seja utilizada de maneira a promover o bem comum, respeitando os direitos humanos e a dignidade das pessoas.

[17] Os protagonistas do debate são o Parlamento Europeu e a Comissão Europeia. Embora o primeiro parecesse desde o início certo da necessidade de intervir para fornecer regras específicas para a automação e a IA, a Comissão Europeia inicialmente considerou completamente eficaz a regulamentação existente também aplicada às novas tecnologias e depois, desde 2021, avaliou a possibilidade de intervir para regular esse fenômeno (Proposta de Lei IA, Proposta IALD, Proposta de revisão do PLD 16).

Neste capítulo, abordaremos duas das principais iniciativas internacionais nesse sentido: os princípios desenvolvidos pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e as recomendações promovidas pela Organização das Nações Unidas para Educação, a Ciência e a Cultura (UNESCO).

### 3.4.1 Indicação e análise dos princípios da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)

Os princípios de Inteligência Artificial da OCDE (Organização para a Cooperação e Desenvolvimento Econômico) foram primeiramente adotados em 2019 e passaram por atualização em maio de 2024. Os princípios fornecem diretrizes para os envolvidos na área de inteligência artificial ao buscarem o desenvolvimento de uma IA confiável, e também oferecem aos formuladores de políticas sugestões para a implementação de políticas eficazes relacionadas à inteligência artificial. Os países usam os princípios de IA da OCDE para estruturar políticas e criar um arcabouço de estruturas de risco para a IA. Assim, os cinco princípios estruturais da IA segundo a OCDE são: Crescimento inclusivo, desenvolvimento sustentável e bem-estar; direitos humanos e valores democráticos, incluindo justiça e privacidade; transparência e explicabilidade; robustez, segurança e proteção; e responsabilidade[18] (Hartmann Peixoto, 2020a; Hartmann Peixoto, 2020b; OECD, 2024b).

Aprofundando no tema princípios, de acordo com a OCDE, o princípio do crescimento inclusivo, desenvolvimento sustentável e bem estar tem como intuito promover uma IA responsável e confiável em busca de resultados benéficos para as pessoas e para o planeta. A finalidade deve ser promover a inclusão de população sub-representadas, reduzir desigualdades econômicas, sociais de gênero dentre outras e proteger o meio ambiente promovendo o desenvolvimento sustentável deste. O princípio de direitos humanos e valores democráticos na IA possuem a função de fazer com que os participantes no

[18] A Recomendação da OCDE sobre a IA é o primeiro padrão intergovernamental sobre a IA. Hoje, há 47 adeptos aos princípios. Sendo que os governos que se comprometem com os princípios da IA e são membros da OCDE são: Austrália, Alemanha, México, Reino Unido, Estados Unidos, Itália, Canadá, Japão, Turquia e a União Europeia. Os que não são membros da OCDE são: Argentina e Brasil. No entanto, os membros do G20 comprometeram-se com os Princípios de IA do G20 extraídos dos princípios de IA da OCDE. Os membros do G20 não mencionados acima são: China, Índia, Indonésia, Rússia, Arábia Saudita, África do Sul (OCDE 2019)

desenvolvimento da IA atuem de acordo com o Estado de Direito, direitos humanos e valores democráticos de forma que sejam centrados no ser humano ao longo de todo o ciclo de vida do sistema de IA. Estes valores incluem a não discriminação e a igualdade, liberdade, dignidade, autonomia dos individuos, privacidade e a proteção dos dados, a diversidade, equidade, justiça social e os direitos internacionalmente reconhecidos. Além disso, deve-se combater a desinformação amplificada pela IA com a participação humana na supervisão da IA quando for necessário (Hartmann Peixoto, 2020a; OECD, 2024b).

No que tange a transparência e explicabilidade, os intervenientes na IA devem ser transparentes e responsáveis em relação aos sistemas de IA. Eles devem proporcionar informações adequadas e compreensíveis sobre as capacidades, limitações e processos dos sistemas de IA, bem como permitir que as partes interessadas contestem os resultados. Por fim, os dois outros princípios elencados pela OCDE; princípio da robustez, segurança e proteção; e princípio da responsabilidade prevêem que os sistemas de IA devem funcionar de forma robusta e segura durante todo o seu ciclo de vida, assim deve ser implementado mecanismos para garantir que se os sistemas de IA apresentarem comportamentos indesejados, possa ser substituído, reparado ou desativado com segurança e os atores que operam a IA devem ser

responsabilizados pelo adequado funcionamento dos sistemas de IA[19] (Hartmann Peixoto, 2020a; OECD, 2024b).

Estes cinco princípios elencados anteriormente estão em conformidade com a atualização feita pela OCDE em maio de 2024 em virtude dos rápidos avanços nas tecnologias de IA, particularmente em IA de uso geral e generativa, os princípios atualizados fornecem e abordam desafios com um foco em segurança, privacidade e direitos de propriedade intelectual e integridade das informações. Com o rápido avanço tecnológico da IA em todos os setores, de 2019 até aqui (2024), fez com que os governos interessados em IA pudessem discutir o assunto para conseguirem entender os novos desafios éticos sobre o assunto. Assim, os princípios servem como um norte para o desenvolvimento responsável da IA e mesmo que os princípios da IA da OCDE não sejam vinculativos, entende-se que os benefícios de sua implementação de forma gradual em legislações e regulamentações de IA de países será uma forma de beneficiar e moldar a governança da IA a nível global (OECD, 2024b; OECD, 2024c).

Nestes termos, pretende-se justificar que desenvolver padrões éticos e de segurança dependem de uma cooperação internacional. Portanto, uma harmonização das legislações e regulamentações a nível global poderá permitir

---

[19] O princípio da transparência e explicabilidade: os intervenientes na IA devem comprometer-se com a transparência e a divulgação responsável em relação aos sistemas de IA. Para o efeito, devem fornecer informações significativas, adequadas ao contexto e consistentes com o estado da arte: promover uma compreensão geral dos sistemas de IA, incluindo as suas capacidades e limitações,para conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho,sempre que possível e útil, fornecer informações simples e fáceis de entender sobre as fontes de dados/entradas, fatores, processos e/ou lógica que levaram à previsão, conteúdo, recomendação ou decisão, para permitir que aqueles afetados por um sistema de IA entendam a saída e,fornecer informações que permitam que aqueles afetados negativamente por um sistema de IA contestem seus resultados. No que tange ao princípio da responsabilidade, os atores da IA devem ser responsabilizados pelo funcionamento adequado dos sistemas de IA e pelo respeito aos princípios acima, com base em suas funções, no contexto e de forma consistente com o estado da arte. Para o efeito, os intervenientes na IA devem garantir a rastreabilidade, nomeadamente em relação aos conjuntos de dados, processos e decisões tomadas durante o ciclo de vida do sistema de IA, para permitir a análise dos resultados do sistema de IA e das respostas ao inquérito, adequadas ao contexto e consistentes com o estado do arte.Os atores da IA devem, com base em suas funções, no contexto e em sua capacidade de agir, aplicar uma abordagem sistemática de gerenciamento de risco a cada fase do ciclo de vida do sistema de IA de forma contínua e adotar conduta empresarial responsável para abordar riscos relacionados aos sistemas de IA, incluindo, conforme apropriado, por meio da cooperação entre diferentes atores da IA, fornecedores de conhecimento de IA e recursos de IA, usuários do sistema de IA e outras partes interessadas. Os riscos incluem aqueles relacionados a preconceitos prejudiciais, direitos humanos, incluindo segurança, proteção e privacidade, bem como direitos trabalhistas e de propriedade intelectual (OECD, 2024b).

que as tecnologias sejam gerenciadas através de fronteiras com regulamentações diferentes, mas que não sejam conflitantes (OECD, 2024b).

Outra consideração muito relevante sobre os princípios da IA previstos pela OCDE é em relação aos riscos reais de IA e as considerações éticas sobre o assunto. Neste quesito, percebe-se que os termos escolhidos para os Princípios de IA demonstram um compromisso da OCDE em relação à definição de padrões claros e alcançáveis. Ao invés de se ater a considerações éticas abstratas, a OCDE busca estabelecer diretrizes mais concretas que possam ser implementadas e seguidas de forma prática. Essa abordagem visa promover a adoção responsável da IA e garantir que seu desenvolvimento e uso sejam conduzidos de maneira ética e transparente. Dessa forma, ao adotar uma linguagem mais específica e objetiva, a OCDE demonstra seu compromisso em fornecer diretrizes concretas e factíveis para a utilização da IA, levando em consideração os princípios de valores humanos, justiça, transparência e explicabilidade. Essas diretrizes têm como objetivo guiar governos, empresas e demais instituições no desenvolvimento e uso responsável da IA, visando o benefício da sociedade como um todo[20] (OECD, 2024b; OECD, 2024c).

A OCDE desenvolveu um importante documento para a classificação de risco dos sistemas de IA conhecido como "*Framework for the Classification of AI systems*", este preceitua a necessidade de classificar os sistemas de IA em diferentes categorias de risco, variando de baixo a alto risco, com base em

[20] Os Princípios por si só não podem moldar uma IA confiável, mesmo que sejam elaborados para corresponder à ampla amplitude da IA. O Observatório de Políticas OCDE.AI dispõe de uma riqueza de ferramentas e conhecimentos especializados da OCDE, dos seus membros, parceiros e partes interessadas para definir políticas para uma IA fiável. Guiado pelos conhecimentos da Rede de Peritos em IA da OCDE — composta por especialistas em IA de diversos setores — o que a OCDE.AI tem para oferecer está na vanguarda da definição da governação global da IA. Vale a pena conhecer algumas das ferramentas e conteúdos mais relevantes do Observatório. Os painéis de políticas nacionais da OCDE.AI fornecem o maior repositório de iniciativas e políticas de IA de mais de setenta governos em todo o mundo. Dados em tempo real sobre temas cruciais dos nossos parceiros ajudam os decisores políticos a desenvolver, implementar e aperfeiçoar políticas de IA. Eles oferecem uma ampla gama de visualizações de dados relacionados à IA para análise comparativa, monitoramento e desenvolvimento de melhores práticas.O AI Incidents Monitor (AIM) é uma fonte de incidentes e riscos de IA conforme aparecem na imprensa global. Essas informações fornecem insights sobre riscos e auxiliam no estabelecimento de um entendimento coletivo de danos comuns de IA e IA confiável.O portal de recursos de IA generativa cobre os benefícios, riscos e aspectos em evolução da IA generativa, oferecendo ferramentas e recursos para uma compreensão abrangente de uma perspectiva política. O Catálogo de Ferramentas e Métricas para IA da OCDE é um recurso que centraliza diversas ferramentas, mecanismos e práticas para ajudar todos os intervenientes a garantir a fiabilidade da IA (OECD, 2024b).

critérios como o potencial de impacto sobre os direitos fundamentais, a segurança, a privacidade e a dignidade humana. Por exemplo, sistemas de IA usados em contextos críticos, como saúde, transporte ou segurança pública, são geralmente classificados como de alto risco e, portanto, estão sujeitos a normas e supervisão mais rigorosas. Além disso, o *framework* sugere que as abordagens regulatórias devem ser proporcionais ao nível de risco identificado, assegurando que sistemas de baixo risco não sejam sobrecarregados com regulamentações excessivas, enquanto sistemas de alto risco recebam a devida atenção para mitigar possíveis danos. Esse tipo de estrutura é essencial para guiar legisladores e reguladores na criação de políticas eficazes e equilibradas para o desenvolvimento e a implementação segura e ética de tecnologias de inteligência artificial (OECD, 2022a).

E, por fim, e não menos importante, as principais recomendações da OCDE para a construção de uma IA inovadora e confiável está intimamente ligada aos princípios descritos, portanto, as recomendações da OCDE são: o investimento em pesquisas privadas e públicas para o desenvolvimento de IA; promoção de um ecossistema (dados, tecnologias de IA e infraestruturas computacionais e de conectividade e mecanismos para a partilha de conhecimentos de IA) inclusivo; moldar um ambiente político e de governança interoperável favorável para a IA (assegurando obter uma IA fiável); desenvolvimento de capacidade humana a fim de preparar o ser humano para a transição do mercado e  a cooperação internacional para uma IA confiável [21] (OECD, 2024b; Hartmann Peixoto, 2020a).

Portanto, a atuação da OCDE no desenvolvimento da IA tem um impacto significativo ao promover diretrizes e padrões internacionais para uma IA responsável. A OCDE estabelece princípios para a IA que visam assegurar

[21] Esta recomendação da OCDE é sobre moldar um ambiente político e de governança interoperável favorável para a IA. Os governos devem promover um ambiente político ágil que apoie a transição da fase de investigação e desenvolvimento para a fase de implantação e operação de sistemas de IA fiáveis. Para o efeito, devem considerar a utilização da experimentação para proporcionar um ambiente controlado no qual os sistemas de IA possam ser testados e ampliados, conforme apropriado. Deverão também adoptar abordagens baseadas em resultados que proporcionem flexibilidade na consecução dos objectivos de governação e cooperar dentro e entre jurisdições para promover uma governação e ambientes políticos interoperáveis, conforme apropriado.Os governos devem rever e adaptar, conforme apropriado, os seus quadros políticos e regulamentares e mecanismos de avaliação à medida que se aplicam aos sistemas de IA para incentivar a inovação e a concorrência por uma IA fiável (OCDE, 2019).

sistemas seguros, transparentes, e éticos. Esses princípios ajudam a orientar políticas públicas e práticas empresariais, promovendo a confiança pública e a cooperação internacional. A OCDE também incentiva a pesquisa e o desenvolvimento de tecnologias de IA que respeitem os direitos humanos e promovam o bem-estar econômico e social, além de fornecer documento base para a avaliação e classificação de sistemas de IA, levando em conta riscos, impactos e a complexidade dos sistemas envolvidos.

### 3.4.2 Promoção e avanços das recomendações da Organização das Nações Unidas para Educação, a Ciência e a Cultura (UNESCO)

A Recomendação da Organização das Nações Unidas para Educação e a Cultura (UNESCO) sobre ética da Inteligência Artificial, aprovada em 23 de novembro de 2021, enfatiza tanto a utilidade das tecnologias de IA para a humanidade e todos os países quanto às questões éticas que levantam. Entre elas, está a preocupação com as distorções que podem gerar e agravar, resultando em discriminação, desigualdade, exclusão digital, exclusão em geral e ameaça à diversidade cultural, social e biológica. Também há a necessidade de transparência e compreensibilidade do funcionamento dos algoritmos e dos dados com que eles foram alimentados, bem como seu potencial impacto sobre a dignidade humana, os direitos humanos e as liberdades fundamentais (UNESCO, 2021).

Destaca-se que é importante ressaltar que os perigos e avanços éticos não devem ser obstáculos para o progresso da inteligência artificial, porém é crucial que as pesquisas sejam conduzidas de maneira ética e que as inovações fundamentem as tecnologias de IA nos direitos humanos, nas liberdades fundamentais, nos valores e princípios, assim como na reflexão ética e moral. Além disso, é necessário destacar que um marco normativo para as tecnologias de IA e suas implicações sociais deve ter como base os marcos jurídicos internacionais e nacionais, os direitos humanos, a ética, a necessidade de acesso a dados, informações e conhecimento, a liberdade de pesquisa e inovação, o bem-estar humano, ambiental e ecológico, e deve vincular valores e princípios éticos aos desafios e oportunidades relacionadas às tecnologias de IA, com base no entendimento comum e nos objetivos

compartilhados. Os valores e princípios éticos podem ajudar no desenvolvimento e implementação de medidas políticas e normas legais baseadas em direitos, fornecendo orientação diante do rápido desenvolvimento tecnológico (UNESCO, 2021).

Os objetivos desta Recomendação são fornecer um conjunto universal de valores, princípios e ações para guiar os Estados na criação de legislações, políticas e outros instrumentos relacionados à IA, em conformidade com o direito internacional. Além disso, a Recomendação visa orientar as ações de indivíduos, grupos, comunidades, instituições e empresas privadas para garantir a inclusão da ética em todas as fases do ciclo de vida dos sistemas de IA. Entre os objetivos também estão a proteção, promoção e respeito aos direitos humanos, liberdades fundamentais, dignidade e igualdade humana, incluindo a igualdade de gênero; a salvaguarda dos interesses das gerações presentes e futuras; a preservação do meio ambiente, biodiversidade e ecossistemas; e o respeito à diversidade cultural. Ademais, a Recomendação busca promover o diálogo multissetorial e pluralista entre diversas partes interessadas, construindo consenso sobre questões éticas relacionadas aos sistemas de IA. Por fim, promove o acesso equitativo aos avanços e conhecimentos em IA e ao compartilhamento dos benefícios, com atenção especial às necessidades e contribuições dos países de baixa renda, incluindo os menos desenvolvidos[22] (UNESCO, 2021).

No artigo "Promovendo a Ética na Inteligência Artificial: Perspectiva da UNESCO" (Ramos *et al*., 2024), é ressaltado que muitos países estão atualmente em processo de implementação da Recomendação sobre Ética da Inteligência Artificial da UNESCO, com o objetivo de adotar, em nível nacional e institucional, estruturas regulatórias que busquem promover o uso ético e o desenvolvimento da IA. A IA tem o potencial de proporcionar um futuro mais justo, equitativo e sustentável; isso implica inclusive a discussão das questões

[22] As alterações às leis nacionais existentes ou a elaboração de nova legislação que aborde sistemas de IA devem cumprir com as obrigações legais dos Estados-membros relativas a direitos humanos e promover esses direitos e as liberdades fundamentais em todo o ciclo de vida daqueles sistemas. A promoção deles também deve assumir a forma de iniciativas de governança, bons exemplos de práticas colaborativas em relação a sistemas de IA, e diretrizes técnicas e metodológicas, nacionais e internacionais, à medida que as tecnologias de IA avancem. Os diversos setores, incluindo o setor privado, em suas práticas relativas aos sistemas de IA, devem respeitar, proteger e promover os direitos humanos e as liberdades fundamentais, usando instrumentos novos e existentes em combinação com esta Recomendação. (UNESCO, 2021)

de gênero. Garantir a inclusão dessa questão na IA, destaca a necessidade de coleta de dados desagregados por gênero e descrevendo abordagens concretas para assegurar que os sistemas de IA não reproduzam disparidades de gênero. Além disso, a recomendação também promove o empreendedorismo, a participação e o envolvimento feminino (UNESCO, 2021).

Por fim, é crucial uma boa regulamentação para promover ambientes robustos e favoráveis à IA responsável. A suposta dicotomia entre regulamentação e inovação é falha porque quadros regulamentares eficazes proporcionam certeza, beneficiando tanto os inovadores quanto os usuários. A questão não é se deve haver regulamentação, mas sim qual é a melhor forma de fazê-la. O objetivo não deve ser a regulamentação em si, mas a construção de uma ambiente propício à IA que contribua para o bem público (Ramos *et al*., 2024).

Neste âmbito, tanto vozes privadas como públicas precisam de uma revisão na maneira como as tecnologias de IA são desenvolvidas e implementadas, sinalizando a necessidade de governança e supervisão mais rigorosas. Desde a adoção da Recomendação sobre Ética da IA em 2021, a UNESCO tem trabalhado na construção de ferramentas e sistemas de apoio para sua implementação. Em ausência de quadros regulamentares abrangentes, os países correm o risco de não conseguir avaliar e abordar responsabilidades e obrigações: quadros de IA claros e responsáveis são fundamentais para garantir o Estado de direito no mundo digital[23] (Ramos *et al*., 2024).

As iniciativas multilaterais e colaborativas representam uma oportunidade para a cooperação entre países na promoção de diretrizes éticas para o desenvolvimento da IA responsável. Isso inclui parcerias, acordos e fóruns internacionais que visam a harmonização de padrões éticos e a criação de diretrizes comuns. A análise dessas iniciativas permite avaliar o nível de colaboração internacional e o impacto na governança global da IA. A UNESCO

[23] Vários países estão a trabalhar no desenvolvimento de políticas e quadros jurídicos de IA, incluindo regulamentos. Só em 2022, leis que mencionam o termo "IA"foram debatidas nos órgãos legislativos de 127 países, e 37 delas foram aprovadas (num total de 123 leis), de acordo com IA Index report 2023 da Universidade de Stanford (Ramos *et al*., 2024)

representa uma iniciativa multilateral e colaborativa eficaz e importante para a construção de Recomendações Éticas para os Estados-membros.

# 4 Contribuições estrangeiras para o desenvolvimento normativo da IA

Este capítulo é fundamental para compreender como alguns países têm enfrentado os desafios éticos e legais que emergem com o avanço da IA. Aqui examina-se as iniciativas da União Europeia, Estados Unidos e Canadá; com suas legislações e diretrizes e como têm contribuído para moldar um desenvolvimento da IA que seja ao mesmo tempo inovador, responsável e um exemplo para o Brasil.

A análise restrita a esses três pontos se justifica na medida em que representam, individualmente, no âmbito da IA, um seu peso econômico e político, um exemplo no foco em direitos humanos e valores democráticos . A abrangência é tal que servem de influência na construção de normas internacionais, apresentam uma abordagem equilibrada entre inovação e regulação. Além disso, são, em seu conjunto, inquestionavelmente lideranças globais na criação de normas em IA.

Depois de entender como se dá esses posicionamentos exemplares, nos capítulos seguintes veremos como eles também contribuem e se distinguem para a construção normativa brasileira em IA.

## 4.1. Legislação e Regulação da IA na União Europeia

Neste tópico será apresentado o que há de mais imprescindível no bloco da União Europeia em termos de regulação e legislação de IA. Os textos foram escolhidos para o destaque são:

Diretrizes da UE sobre ética em inteligência artificial; *“White Paper”* sobre Inteligência Artificial; Relatório sobre implicações de segurança e responsabilidade da IA, IoT e da Robótica; Diretiva Geral de Segurança dos Produtos e legislação sobre segurança dos produtos; Diretiva de Máquinas/Regulamento (UE) 2023/1230 do Parlamento Europeu; Regulamento de IA (AI Act); Regulamento Geral sobre a Proteção de Dados (GDPR).

### 4.1.1 Diretrizes da UE sobre ética em inteligência artificial (*"Ethics Guidelines for Trustworthy AI"*)

Em abril de 2019, a UE publicou as Diretrizes sobre a Ética na IA, pela qual pode-se inferir que a UE é considerada pioneira no que tange ao estabelecimento de um quadro de normas éticas para a IA. Deste modo, fazendo uma cronologia sobre o debate da IA no Parlamento Europeu, a nível da UE, a Comissão Europeia começou a avaliar o impacto da IA fazendo recomendações abrangentes sobre regras de direito civil em robótica em janeiro de 2017. O Parlamento elaborou um código de ética para engenheiros robóticos e solicitou à Comissão que considerasse a criação de uma agência europeia para a robótica e a IA, encarregada de fornecer um arcabouço de conhecimentos técnicos, éticos e regulamentares necessários para coordenar este ambiente de IA. Nesta mesma cronologia, em 2018, a Comissão adotou uma comunicação para promover o desenvolvimento da IA na Europa e em 2019 publicou um plano de ação coordenado sobre IA, endossado pelo Conselho da União Europeia que tinha como finalidade coordenar as estratégias nacionais de IA dos Estados Membros da UE (Mandiega, 2019).

Neste tocante, em abril de 2019, a Comissão publicou um conjunto de diretrizes éticas que não possuem característica vinculativa para se obter uma IA fiável. Este documento foi preparado pelo Grupo de Perito de Alto Nível (HLEG) da Comissão em IA, grupo composto por 52 especialistas independentes, este documento visa oferecer orientações de como promover um desenvolvimento ético destes sistemas. Este documento traz como principal foco a IA centrada no ser humano, respeitando os direitos humanos e os Tratados da União Europeia e a Carta dos Direitos Fundamentais da União Europeia. Para que a UE consiga entregar um sistema de IA confiável, segundo o documento contendo as diretrizes éticas para que a IA respeite as leis e regulamentos aplicáveis aos sistemas de IA terá de ser robusta tanto do ponto de vista técnico como social e ser centrado no ser humano[24] (Mandiega, 2019; Hamon, Junklewitz e Sanchez, 2020).

[24] Vários projetos à escala da UE em cooperação ou sob orientação da Comissão Europeia: o projeto AI4EU2 e o observatório AI Wacht3, e a aliança europeia para a IA4.

Neste mesmo sentido, existem 7 (sete) requisitos éticos que orientam a UE no desenvolvimento de sistemas de IA, o destaque vai para agência humana e supervisão; robustez e segurança; transparência e responsabilidade[25]. Como dito anteriormente, a supervisão humana deverá ser realizada para avaliar o impacto dos sistemas de IA na aplicação dos direitos fundamentais, esta supervisão deverá ser feita previamente e posteriormente a sua implementação. Outra questão importante elencada é que o direito dos utilizadores não podem ser baseados exclusivamente no tratamento automatizado, uma vez que isto pode implicar em questões que afetem negativamente as partes. A Robustez técnica e segurança exige que os algoritmos sejam confiáveis, seguros e robustos para lidar com todas as possíveis falhas ao longo do processo da implantação de um sistema de IA, a UE propõe que para alcançar este patamar deve-se promover a cooperação entre a comunidade da IA e a comunidade de segurança e, consequentemente analisar a possibilidade de modificar o quadro jurídico que rege responsabilidade para um quadro de responsabilidade baseado na conduta humana para um quadro de responsabilidade mais baseado em máquinas (Mandiega, 2019; Hamon, Junklewitz e Sanchez, 2020).

A transparência é fundamental para garantir que a IA não seja parcial e preconceituosa. O documento de diretrizes da IA preocupa-se em resguardar medidas de transparência na indústria da IA para garantir a transparência, como a documentação e rastreabilidade dos conjuntos de dados e processos usados na construção de sistemas de IA. Além disso, sistemas de IA devem ser identificáveis, e os usuários precisam saber que estão interagindo com IA. O princípio da explicabilidade exige que as decisões e sistemas de IA sejam compreensíveis e rastreáveis pelos humanos. Segundo o documento de diretrizes ética da IA no que tange a responsabilidade deve ser criado mecanismos para garantir a responsabilidade e a responsabilização pelos sistemas de IA e pelos resultados, com aplicação de ferramentas de avaliação

[25] Os sete requisitos da UE para alcançar uma IA fiável são: agência humana e supervisão; robustez e segurança; privacidade e governaça de dados; transparência; diversidade, não discriminação e justiça social e ambiental e bem-estar e responsabilidade (Mandiega, 2019). De acordo com Hamom (2020, p. 13) os requisitos de agência humana e supervisão, robustez técnica e segurança; privacidade e governação de dados e transparência enfatizam elementos que se ligam diretamente aos campos da robustez da IA, explicabilidade, considerações legais de proteção de dados e segurança cibernética.

de impacto bem como aplicação de relatórios e auditorias (Mandiega, 2019; Hamon, Junklewitz e Sanchez, 2020).

As diretrizes éticas não possuindo características vinculativas abrem precedentes de discussão para preocupações de caráter de que a própria indústria tecnológica pode moldar o debate ético sobre algoritmos e sistemas automatizados de tomada de decisão, assim é necessário que haja uma supervisão regulamentar para monitorar a aplicação das diretrizes éticas da UE. Neste quesito de coordenação regulatória vários estados membros da UE começaram a trabalhar no seu quadro regulamentar e de ética (Mandiega, 2019).

Seguindo a cronologia no tocante ao Plano Coordenado sobre a IA 2021 Revisão, este se movimenta de mãos dadas com a proposta de Regulamentos sobre a IA, estabelecendo estratégias para acelerar os investimentos em IA em tecnologias de IA; atuando em estratégias e programas de IA de forma a garantir que a UE se beneficie dos primeiros a adotar e alinhar a política da IA para enfrentar os desafios globais. O Plano Coordenado sobre IA baseia-se na Colaboração estabelecida entre a Comissão e os Estados-Membros durante o Plano Coordenado de 2018, este Plano tem a função de entender as mudanças políticas e o investimento necessário a nível dos Estados-Membros para reforçar a posição de liderança da Europa no desenvolvimento da IA centrada no ser humano, sustentável, segura, inclusiva e fiável (European Commission, 2021).

4.1.2. Publicação do *"White Paper"* sobre Inteligência Artificial: uma abordagem europeia à excelência e à confiança e o Relatório sobre implicações de segurança e responsabilidade da IA, IoT e da Robótica.

O White paper dispõe que a IA é um conjunto de tecnologias que combinam dados, algoritmos e poder computacional. Assim, a Europa tem como objetivo empregar sistemas de IA e ter um quadro regulamentar baseado nos seus valores fundamentais para se posicionar no mercado de tal forma que se torne líder mundial em desenvolvimento de tecnologia baseada em dados e em suas aplicações. Assim, o documento *White Paper on Artificial Intelligence-A European approach to excellence and trust* apresenta opções políticas para o desenvolvimento ético, seguro e responsável da IA, de forma a alinhar

estratégias entre setores públicos e privados em toda a Europa e de elaborar um quadro regulamentar robusto de forma a respeitar os direitos fundamentais e os direitos dos consumidores, dando enfoque aos sistemas de IA que representam um risco elevado. Este é considerado um dos principais guias para a estratégia da IA na Europa e apoia a ideia do desenvolvimento de uma IA centrada no ser humano (European Commission, 2020a).

Este documento enfatiza a necessidade de equilibrar a promoção da inovação com a proteção dos direitos fundamentais, propondo um quadro regulatório baseado em uma abordagem de risco. Assim, sistemas de IA que representam riscos elevados para a segurança e os direitos dos cidadãos, como aqueles utilizados em setores críticos, devem ser submetidos a regulamentações mais rigorosas. Essa abordagem visa garantir que a IA contribua de forma positiva para a sociedade, minimizando possíveis impactos negativos (European Commission, 2020a).

Ademais, o documento sublinha a importância do investimento em pesquisa e desenvolvimento para manter a liderança da Europa no campo da IA. Para tanto, incentiva a criação de uma infraestrutura tecnológica robusta, o apoio a startups, e a formação de redes de centros de inovação digital. Esse compromisso com a excelência e a inovação é visto como fundamental para assegurar que a Europa não apenas acompanhe, mas também lidere o avanço tecnológico global. Além disso, o White Paper destaca a necessidade de uma governança harmonizada entre os Estados-Membros da UE, promovendo a cooperação internacional na definição de normas globais para a IA. Essa harmonização é essencial para criar um mercado único digital, onde as tecnologias de IA possam ser desenvolvidas e implementadas de forma consistente e segura (European Commission, 2020a).

Por fim, o documento reitera o compromisso da União Europeia com a proteção de dados e privacidade, alinhando o desenvolvimento da IA com o Regulamento Geral sobre a Proteção de Dados (GDPR). A Comissão Europeia defende que as tecnologias de IA devem ser transparentes, explicáveis e auditáveis, garantindo que os benefícios sejam amplamente distribuídos e que os operadores sejam responsabilizados pelos impactos de suas tecnologias. Essa abordagem inclusiva e responsável reflete o objetivo da UE de promover

uma IA ética que respeite a diversidade e os valores fundamentais da sociedade europeia (European Commission, 2020a).

O relatório que acompanha o *"White Paper"* sobre as implicações de segurança e responsabilidade da IA, Internet das Coisas e robótica concluiu que a legislação atual de segurança de produtos tem lacunas, especialmente na Diretiva Máquinas, que precisam ser abordadas. A legislação de segurança de produtos da UE, que inclui a Diretiva Geral de Segurança dos Produtos, foi desenvolvida em um contexto onde os produtos eram predominantemente físicos e tangíveis, o que limita sua eficácia na abordagem dos desafios apresentados por tecnologias digitais emergentes como a IA. A legislação atual não aborda suficientemente a capacidade dos sistemas de IA de aprender e evoluir após a comercialização, o que pode criar novos riscos não previstos inicialmente. Além disso, a complexidade e a opacidade desses sistemas dificultam a atribuição de responsabilidade em caso de falhas. A legislação também não considera adequadamente as interações dinâmicas e a autonomia dos sistemas de IA, nem a necessidade de supervisão contínua ao longo de seu ciclo de vida (European Commission, 2020b).

Neste tocante, O Relatório sobre as implicações de segurança e Responsabilidade da IA, da Internet das Coisas e da Robótica ressalta que é imprescindível a Europa estar empenhada em fazer investimentos nestas áreas para poder criar oportunidades e benefícios a sociedade, mas para que isto aconteça é preciso um quadro jurídico claro e previsível que aborde os desafios tecnológicos. Assim, um quadro jurídico claro no tocante a segurança e responsabilidade no que diz respeito ao funcionamento destes sistemas de IA irão produzir um ambiente de proteção dos consumidores e a segurança jurídica para as empresas envolvidas (European Commission, 2020b).

### 4.1.3 Diretiva Geral de Segurança dos Produtos e legislação sobre segurança dos produtos (Diretiva 2001/95/EC)

A legislação da União Europeia sobre segurança de produtos visa garantir que os produtos no mercado atendam a altos requisitos de saúde, segurança e meio ambiente e que possam circular livremente pela União. A

Diretiva Geral de Segurança de Produtos[26] complementa a legislação setorial, assegurando a segurança de todos os produtos de consumo. A vigilância do mercado é normatizada pelo Regulamento de Vigilância do Mercado. No transporte, há regras adicionais para veículos, aeronaves e navios, além de responsabilidades para operadores e autoridades. Assim, a padronização europeia é essencial, especialmente frente à digitalização e tecnologias emergentes, sendo a cooperação internacional crucial para a competitividade da indústria europeia (European Commission, 2020b).

O Relatório sobre as implicações de segurança e Responsabilidade da IA, da Internet das Coisas e da Robótica expõe que a Legislação sobre segurança de produtos da União foi escrita antes do surgimento de tecnologias digitais como IA, IoT ou robótica, portanto não possui previsão jurídica para todos os desafios e riscos encontrados nesta área atualmente. No entanto, como a estrutura que protege a segurança de produtos é neutra na questão da tecnologia, isso significa que ela pode ser aplicada às novas tecnologias de produtos que incorporem estas tecnologias. Além disso, existem alguns atos legislativos posteriores que fazem parte dessa estrutura como nos setores de dispositivos médicos ou automóveis, já consideraram alguns aspectos do surgimento de tecnologias digitais, como por exemplo, decisões automatizadas, software como um produto separado e conectividade (European Commission, 2020b).

A regulamentação da União em relação à segurança dos produtos geralmente não inclui requisitos obrigatórios específicos contra ameaças cibernéticas que impactam a segurança[27] dos usuários. No entanto, existem disposições relacionadas à segurança no Regulamento sobre Dispositivos Médicos[28], na Diretiva sobre instrumentos de medição[29], na Diretiva sobre

---

[26] Diretiva 2001/95/CE do Parlamento Europeu e do Conselho, de 3 de dezembro de 2001, relativa à segurança geral dos produtos (JO L 11 de 15.1.2002, p. 4–17) .

[27] O conceito de segurança na atual legislação de segurança de produtos da União está em linha com um conceito estendido de segurança para proteger consumidores e usuários. Assim, o conceito de segurança de produtos abrange proteção contra todos os tipos de riscos decorrentes do produto, incluindo não apenas riscos mecânicos, químicos e elétricos, mas também riscos cibernéticos e riscos relacionados à perda de conectividade de dispositivos.Poderiam ser consideradas disposições explícitas a este respeito no âmbito das peças legislativas relevantes da União, a fim de proporcionar uma melhor proteção dos utilizadores e mais segurança jurídica. (European Commission, 2020b).

[28] Regulamento (UE) 2017/745 sobre dispositivos médicos (European Commission, 2020b).

[29] Diretiva 2014/32/UE relativa à disponibilização no mercado de instrumentos de medição (European Commission, 2020b).

Equipamentos de Rádio[30] ou na legislação de aprovação de tipo de veículo. A Lei de Segurança Cibernética[31] estabelece um quadro de certificação voluntária de segurança cibernética para produtos, serviços e processos de TIC, enquanto a regulamentação relevante da União Europeia sobre segurança de produtos estabelece requisitos obrigatórios. Os exemplos dados pelo relatório foram um relógio inteligente desenvolvido para crianças, que não causa danos diretos à criança em si, no entanto, pode ser usados como ferramentas para se ter acesso a localização dos menores desde que a tecnologia não utilize requisitos mínimos de segurança e outro exemplo foi um carro de passeio utilizado na Alemanha, descobriu-se que o sistema de rádio pode ter uma vulnerabilidade para que terceiros possam ter acesso ao sistema de controle do veículo (European Commission, 2020b).

No quadro atual da União Europeia, quando os produtores descobrem que um produto apresenta riscos de segurança durante seu ciclo de vida, eles são obrigados a informar imediatamente às autoridades competentes e tomar medidas para prevenir esses riscos para os usuários. Isso inclui a retirada do produto do mercado, quando necessário, e a comunicação clara aos consumidores sobre os riscos identificados e as ações corretivas adotadas. Essa obrigação visa garantir a proteção contínua da saúde e segurança dos consumidores. Isto acontece na legislação de transporte, no setor ferroviário. Quando um veículo ferroviário é modificado após a certificação, o autor da modificação deve seguir um procedimento específico. Critérios claros determinam se é necessário envolver a autoridade competente, garantindo que as modificações não comprometam a segurança e que todas as medidas necessárias sejam adotadas para manter a conformidade com os padrões de segurança (European Commission, 2020b).

Na legislação de segurança de produtos da União Europeia, as atualizações de *software* podem ser tratadas como operações de manutenção por motivos de segurança, desde que não modifiquem significativamente o produto e não introduzam novos riscos não previstos na avaliação inicial. No entanto, se a atualização de *software* alterar substancialmente o produto, ele pode ser considerado um novo produto, e a conformidade com a legislação de

---

30 Diretiva de Equipamentos de Rádio 2014/53/UE (European Commission, 2020b).

31 Regulamento (UE) 2019/881 (European Commission, 2020b).

segurança de produtos deve ser reavaliada no momento da modificação (European Commission, 2020b).

A legislação da União Europeia sobre segurança de produtos considera a complexidade das cadeias de valor, impondo obrigações a vários operadores econômicos com base no princípio da "responsabilidade partilhada". Embora a responsabilidade do produtor pelo produto final seja adequada para cadeias de valor complexas, a cooperação explícita entre operadores econômicos e usuários poderia fornecer maior segurança jurídica. Cada ator na cadeia de valor, incluindo produtores de software e usuários que modificam produtos, deve assumir a responsabilidade e fornecer as informações e medidas necessárias ao próximo ator na cadeia. Na prática, isso quer dizer que fabricantes, importadores e, em certas circunstâncias, distribuidores e fornecedores podem ser responsabilizados objetivamente, ou seja, sem necessidade de provar culpa. O objetivo é garantir que os consumidores possam buscar reparação de forma eficaz, independentemente de quem na cadeia de fornecimento é responsável pelo defeito que causou o dano. Dessa forma, todos os operadores têm um incentivo para garantir a segurança dos produtos que colocam no mercado (European Commission, 2020b).

No que tange a **responsabilidade** pelos produtos a nível da União, tem o objetivo de construir um mercado único e forte para os produtos, que minimize os danos aos utilizadores destes produtos e que preveja indenização em razão dos danos promovidos por bens defeituosos. A nível nacional, os quadros de responsabilidade civil não harmonizados complementam as regras da União Europeia, garantindo indenização por danos resultantes de várias causas, como produtos e serviços, e abrangendo diferentes responsáveis, como proprietários, operadores ou prestadores de serviços. Estes quadros legais nacionais asseguram que, independentemente do âmbito da legislação da UE, as vítimas de danos possam ser compensadas de maneira adequada e justa, reforçando a proteção do consumidor e a segurança dos produtos e serviços no mercado (European Commission, 2020b).

Os regimes de responsabilidade na União Europeia são eficazes, baseando-se na Diretiva de Responsabilidade do Produto[32] (Diretiva 85/374/EEC) e em sistemas nacionais não harmonizados. A diretiva

[32] Product Liability Directive (Directiva 85/374/EEC) (European Parliament, 2024).

harmonizou a responsabilidade do fabricante por produtos defeituosos, estabelecendo um sistema de responsabilidade estrita ou objetiva. Em casos de danos físicos ou materiais, a parte prejudicada pode obter indenização se comprovar o dano, o defeito no produto (isto é, que ele não forneceu a segurança esperada) e a relação de causalidade entre o produto defeituoso e o dano. Conclui-se que a diretiva de Responsabilidade do Produto adota uma abordagem baseada na responsabilidade objetiva, a ideia é proteger os consumidores e garantir que aqueles que colocam produtos no mercado sejam incentivados a garantir sua segurança (European Commission, 2020b).

### 4.1.4 Diretiva de Máquinas (Diretiva 2006/42/CE) foi substituída pelo regulamento (UE) 2023/1230 do Parlamento Europeu

A Diretiva 2006/42/CE do Parlamento Europeu e do Conselho foi substituída pelo regulamento (UE) 2023/1230 do Parlamento Europeu e do Conselho de 14 de junho de 2023. A substituição por um regulamento, em vez de uma diretiva, é significativa porque os regulamentos da UE têm aplicação direta em todos os Estados-Membros, sem a necessidade de transposição para as legislações nacionais. Isso assegura uma maior harmonização e uniformidade na aplicação das normas em todo o mercado único europeu. A experiência com a aplicação da Diretiva 2006/42/EC revelou inadequações e inconsistências na cobertura de produtos e nos procedimentos de avaliação da conformidade. Portanto, é necessário aprimorar, simplificar e adaptar as disposições desta diretiva para atender melhor às necessidades do mercado. Além disso, é fundamental fornecer regras claras sobre a estrutura dentro da qual os produtos abrangidos por este regulamento podem ser disponibilizados no mercado, garantindo segurança e conformidade adequadas (European Commission, 2023).

O presente regulamento, (UE) 2023/1230, aplica-se a produtos novos no mercado da União Europeia, sejam eles fabricados por um fabricante estabelecido na UE ou importados de um país terceiro, tanto novos quanto de segunda mão. Isso visa garantir que todos os produtos disponíveis no mercado da UE atendam aos requisitos de segurança, conformidade e qualidade estabelecidos, independentemente de sua origem ou condição. Este regulamento é aplicável quando exista a possibilidade da máquina ser usada

por um consumidor. A Diretiva 2006/42/CE sobre segurança dos produtos possui várias lacunas que este regulamento veio resguardar[33] (European Commission, 2023).

A transição da diretiva 2006/42/CE para o Regulamento (UE) 2023/1230 reflete a necessidade de uma estrutura legal que esteja mais alinhada com o atual ambiente tecnológico e que possa ser aplicada de forma consistente em toda a União Europeia, garantindo a segurança e a eficiência no uso de máquinas modernas. Assim, o novo Regulamento também introduz regras mais rigorosas e atualizadas para assegurar a segurança e a conformidade das máquinas no contexto das novas tecnologias, abordando questões emergentes como a cibersegurança e adaptando as responsabilidades dos fabricantes e operadores aos cenários tecnológicos modernos (European Commission, 2023).

As máquinas parcialmente concluídas[34] são incluídas por este regulamento. Outra questão importante são as questões de produtos que apresentem riscos de saúde e segurança, mas que são abrangidos por cobertura total ou parcial da legislação de harmonização da União mais específica que este regulamento, deverá ser enquadrado na legislação que prevê sobre os riscos de maneira mais específica. Os fabricantes devem ser responsáveis por garantir que os produtos desenvolvidos por estes estejam em conformidade com este regulamento, e se o desenvolvimento de máquinas expor uma característica mais crítica no tocante ao risco maior para os consumidores deverá passar por procedimentos mais rigorosos que conte com

---

[33] Os aparelhos domésticos destinados a uso doméstico que não sejam mobiliário operado eletricamente, equipamento de áudio e vídeo, equipamento de tecnologia da informação, máquinas de escritório, aparelhagem de comutação e comando de baixa tensão e motores elétricos são abrangidos pelo âmbito de aplicação da Diretiva 2014/35/UE do Parlamento Europeu e do Conselho e, por conseguinte, deverão ser excluídos do âmbito de aplicação do presente regulamento. Alguns desses produtos, por exemplo, máquinas de lavar roupa, estão progressivamente a incorporar funções Wi-Fi e, por conseguinte, são abrangidos pela Diretiva 2014/53/UE do Parlamento Europeu e do Conselho como equipamento de rádio. Esses produtos deverão também ser excluídos do âmbito de aplicação do presente regulamento. (...) No que tange ao software, a este respeito, máquinas que não tenham apenas o upload de software destinado à aplicação específica prevista pelo fabricante, e que seja objeto do procedimento de avaliação de conformidade da máquina, devem se enquadrar na definição de máquinas e não nas definições de produtos relacionados ou máquinas parcialmente concluídas (European Commission, 2023).

[34] Máquinas parcialmente concluídas são produtos dentro do escopo deste Regulamento que precisam passar por construção adicional para poderem desempenhar sua aplicação específica, ou seja, as operações bem definidas para as quais o produto foi projetado (European Commission, 2023).

uma participação de um organismo com a finalidade de fazer a avaliação de conformidade[35] (European Commission, 2023).

O objetivo do presente regulamento é garantir que os produtos abrangidos cumpram requisitos que assegurem um elevado nível de proteção da saúde e segurança das pessoas, animais domésticos, bens e, quando aplicável, do ambiente, enquanto mantém o funcionamento do mercado interno. Esse objetivo, devido à necessidade de harmonização, é melhor alcançado a nível da União Europeia, em conformidade com o princípio da subsidiariedade do artigo 5.º do Tratado da União Europeia. O regulamento não excede o necessário para atingir esse objetivo, seguindo o princípio da proporcionalidade (European Commission, 2023).

O Regulamento 2023/1230 da União Europeia, que se concentra na segurança e conformidade das máquinas, continua em vigor e é complementar ao Regulamento de IA (AI Act). Enquanto o Regulamento 2023/1230 aborda aspectos gerais de segurança das máquinas, incluindo aquelas que incorporam tecnologias de inteligência artificial, o AI Act regula especificamente o uso da IA, com ênfase em transparência, governança e gestão de riscos. A coexistência desses regulamentos assegura uma abordagem integrada e abrangente para a segurança e inovação tecnológica na União Europeia (European Commission, 2023).

### 4.1.5 Regulamento de IA (*AI Act*)

A proposta para o Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas sobre Inteligência Artificial (Lei sobre Inteligência Artificial) e altera certos atos legislativos da União estabelecendo regras harmonizadas sobre IA, foi aprovada, e o Parlamento Europeu adotou o *AI Act* em 13 de março de 2024, e o Conselho da União Europeia deu a aprovação em 21 de maio de 2024 (European Parliament, 2024).

Neste tocante, o regulamento sobre a Inteligência Artificial dispõe que o conceito da IA deverá ser feito em colaboração com as organizações internacionais ativas no desenvolvimento da IA, de forma a assegurar a

[35] Artigo 26: Os Estados-Membros devem notificar a Comissão e os outros Estados-Membros dos organismos autorizados a realizar tarefas de avaliação da conformidade por terceiros, em conformidade com o presente regulamento (European Commission, 2023).

segurança jurídica. A lei de IA (*AI Act*) tem o intuito de regular o desenvolvimento, colocação no mercado e uso de IA na UE, com foco em abordagens baseadas em risco. Assim, os principais preceitos são: a abordagem no risco (inaceitável, alto risco, risco limitado e risco mínimo); transparência e responsabilização; proibição de práticas de IA aceitáveis; governança e supervisão; fomento à inovação e proteção dos direitos fundamentais. Estes preceitos visam assegurar o desenvolvimento da IA na UE de forma segura, ética e alinhados com os valores e direitos fundamentais europeus (Renda *et al*, 2021; European Parliament, 2024).

A *AI Act* deverá ser aplicável a prestadores e a responsáveis pela implantação de sistemas de IA que estejam estabelecidos em um país terceiro, na medida em que esteja prevista a utilização na União Europeia dos resultados produzidos por esses sistemas. Essa medida garante que a proteção dos direitos fundamentais e a segurança dos cidadãos europeus sejam mantidas, independentemente da origem do sistema de IA, promovendo um ecossistema de confiança e transparência no uso de tecnologias (European Parliament, 2024).

O *AI Act* do Parlamento Europeu estabelece uma estrutura de classificação de risco para sistemas de IA, dividindo-os em categorias conforme o impacto potencial sobre direitos fundamentais, segurança e dignidade humana. Conforme descrito no Artigo 5, as práticas de IA consideradas inaceitáveis são proibidas, pois representam riscos extremos. Esses incluem, por exemplo, sistemas que manipulam comportamentos humanos por meio de técnicas subliminares, sistemas de pontuação social realizados por governos e sistemas de identificação biométrica em tempo real em espaços públicos para fins de aplicação da lei, salvo em situações muito restritas (European Parliament, 2024).

Além disso, o documento identifica sistemas de IA de alto risco, que, segundo o Artigo 6, são aqueles que podem causar danos significativos à saúde, segurança e direitos fundamentais. Esses sistemas são amplamente utilizados em setores críticos, como infraestrutura, educação, emprego, serviços essenciais e aplicação da lei. Para garantir a segurança e a ética, esses sistemas devem cumprir requisitos rigorosos antes de serem disponibilizados no mercado ou colocados em operação (European Parliament,

2024).

Por outro lado, os sistemas de IA de risco limitado e mínimo são mencionados em artigos subsequentes. Os sistemas de risco limitado, embora não apresentem um alto risco, ainda exigem medidas de transparência, como informar os usuários quando interagem com uma IA. Já os sistemas de risco mínimo, que incluem tecnologias como filtros de spam ou sistemas de recomendação, não estão sujeitos a regulamentações específicas. Esse esquema de classificação visa equilibrar a inovação tecnológica com a proteção dos direitos fundamentais, assegurando o desenvolvimento e uso responsável da IA na União Europeia (European Parliament, 2024).

No que tange à gestão de riscos em sistemas de IA de alto risco, é imprescindível que se estabeleça, implemente, documente e mantenha um sistema de gestão de riscos robusto. Esse sistema deve ser compreendido como um processo contínuo e iterativo, sendo essencial que seja planejado e executado ao longo de todo o ciclo de vida do sistema de IA, de modo que sejam realizadas revisões e atualizações regulares e sistemáticas. Ademais, esse processo contínuo deve englobar várias etapas inter-relacionadas. Inicialmente, é fundamental a identificação e análise dos riscos conhecidos e previsíveis que o sistema de IA pode representar à saúde, à segurança ou aos direitos fundamentais, quando utilizado de acordo com seu propósito pretendido (European Parliament, 2024).

Em seguida, deve-se proceder à estimativa e avaliação desses riscos, inclusive considerando os cenários de uso indevido que sejam razoavelmente previsíveis. É igualmente necessário avaliar outros riscos que possam surgir a partir da análise de dados obtidos por meio do sistema de monitoramento pós-mercado, conforme estipulado no Artigo 72. Concomitantemente, devem ser adotadas medidas de gestão de riscos que sejam apropriadas e direcionadas para mitigar os riscos identificados. O Artigo 72 exige a criação de um sistema de monitoramento pós-mercado para sistemas de IA de alto risco, que deve coletar e analisar dados continuamente para assegurar a conformidade regulatória ao longo do ciclo de vida do sistema. Esse monitoramento deve considerar também interações com outros sistemas de IA, quando aplicável, e ser orientado por um plano documentado, conforme especificações da Comissão Europeia (European Parliament, 2024).

São considerados sistemas de Alto Risco  aqueles sistemas de IA que gere imprecisões técnicas na identificação biométrica[36] à distância de pessoas específicas, estes resultados podem gerar conclusões enviesadas e consequentemente a discriminações no que tange a idade, etnia, raça sexo ou deficiência. Esta tecnologia deverá ser utilizada em casos específicos e por motivos de interesse público, como podemos citar a busca de pessoas determinadas vítimas de crime, pessoas desaparecidas, ameaças de ataques terroristas e em caso de busca de infratores nos quais a pena não seja inferior a 4 (quatro) anos. No caso de ser utilizado esta tecnologia de identificação em espaços públicos, em tempo real, deverá ser precedida de autorização expressa e específica de uma autoridade judiciária ou de uma autoridade administrativa independente de um Estado-Membro[37] (Renda *et al.*, 2021; European Parliament, 2024).

Também são considerados sistemas de alto risco sistemas de IA que visam identificar ou inferir emoções de pessoas singulares podendo gerar resultados discriminatórios, uma vez que a expressão de emoção pode mudar de acordo com culturas, e até mesmo em razão de especificidades de indivíduos específicos. Isto pode gerar um contexto prejudicial aos direitos e liberdades das pessoas envolvidas. Portanto, este tipo de sistema está proibido em situações diversas inclusive em contextos de trabalho e de educação (Renda *et al.*, 2021; European Parliament, 2024).

A classificação como de risco elevado ou não em relação a um produto que possua IA, não tem relação com o fato de o produto possuir um sistema de IA embutido, mas sim segundo critérios específicos relatados nos Regulamentos (UE) 2017/745 e (UE) 2017/746, em que é prevista uma avaliação de conformidade por terceiros de acordo com o risco. A classificação de risco elevado ou de alto risco é usada em setores que afetam diretamente os direitos fundamentais[38] e a segurança dos cidadãos, como por exemplo,

---

[36] A identificação biométrica refere-se ao uso de tecnologias de IA para identificar indivíduos com base em características biológicas únicas, como impressões digitais, reconhecimento facial, íris, voz, entre outros (European Parliament, 2024).

[37] Se não houver possibilidade de pedir essa autorização em razão da gravidade da situação,  a autoridade competente deverá fazê-lo no mais tardar em 24 horas, apresentando os motivos. Se de fato, a autorização for recusada deverá ser apagados todos os dados relacionados a essa autorização (European Parliament, 2024).

[38] Impacto sobre os Direitos Fundamentais: Sistemas de IA que podem afetar significativamente os direitos fundamentais dos cidadãos da UE são considerados de alto risco. Isso

setores como saúde (diagnóstico médico), transporte (veículos autônomos), energia (redes elétricas automatizadas), justiça (sistema de decisão judicial), educação (pontuação em exames) e emprego (recrutamento) e mesmo dentro de setores menos críticos, certas funções desempenhadas pela IA são automaticamente consideradas de alto risco como por exemplo a identificação biométrica em espaços públicos (uso de IA para reconhecimento facial ou outras formas de identificação biométrica em espaços públicos), sistemas de IA que avaliam ou influenciam o comportamento de pessoas (recrutamento de pessoas) (Renda *et al.*, 2021; European Parliament, 2024)

É importante salientar que os riscos abordados neste contexto referem-se apenas àqueles passíveis de mitigação ou eliminação por meio do desenvolvimento ou design do sistema de IA, ou ainda pela provisão de informações técnicas adequadas. Assim, as medidas de gestão de riscos devem considerar os efeitos e as possíveis interações decorrentes da aplicação combinada dos requisitos estabelecidos, com o intuito de minimizar os riscos de forma eficaz, buscando um equilíbrio adequado na implementação dessas medidas (European Parliament, 2024).

Além disso, as medidas de gestão de riscos precisam assegurar que o risco residual associado a cada perigo, bem como o risco residual geral do sistema de IA, seja considerado aceitável. Para tanto, os sistemas de IA de alto risco devem ser submetidos a testes rigorosos, cujo objetivo é identificar as medidas de gestão de riscos mais adequadas. Tais testes devem garantir que os sistemas de IA desempenhem consistentemente seu propósito pretendido, ao mesmo tempo em que permanecem em conformidade com os requisitos regulamentares (European Parliament, 2024).

Outrossim, os procedimentos de teste podem incluir a aplicação em condições reais, conforme descrito no Artigo 60. Esses testes devem ser realizados ao longo do processo de desenvolvimento e, necessariamente, antes que o sistema de IA seja colocado no mercado ou em operação. Por fim, é crucial que os testes sejam conduzidos com base em métricas e limiares probabilísticos previamente definidos, ajustados ao propósito pretendido do sistema (European Parliament, 2024).

---

inclui:privacidade e proteção de dados pessoais, não discriminação e igualdade; liberdade de expressão e proteção contra prejuízos físicos e psicológicos (European Parliament, 2024)

Ademais, os testes podem ser realizados em ambientes reais, conforme o estabelecido no Artigo 60. Esses testes devem ocorrer durante o desenvolvimento e obrigatoriamente antes de o sistema de IA ser comercializado ou utilizado. É essencial que os testes utilizem métricas e parâmetros probabilísticos previamente estabelecidos que considerem fatores éticos, alinhados ao objetivo do sistema. Durante a implementação do sistema de gestão de riscos, é importante avaliar se o sistema de IA de alto risco pode causar impactos negativos em menores de 18 anos e outros grupos vulneráveis, adotando as medidas apropriadas conforme necessário. Ao longo da implementação do sistema de gestão de riscos, deve-se considerar se, dado seu propósito, o sistema de IA de alto risco pode ter impacto adverso em indivíduos com menos de 18 anos, bem como em outros grupos vulneráveis, adotando medidas adequadas conforme necessário (European Parliament, 2024).

Sistemas de IA de alto risco estão sujeitos a requisitos rigorosos de conformidade (antes de ser colocado no mercado, o sistema deve passar por uma avaliação para garantir que cumpre os requisitos do regulamento), transparência e explicabilidade (deve estar claro para os usuários que estão interagindo com uma IA, e eles devem ser informados sobre como as decisões são tomadas); gestão de riscos (os desenvolvedores e operadores desses sistemas devem implementar medidas robustas para identificar e mitigar riscos associados ao uso da IA); supervisão humana e monitoramento contínuo de sistemas de IA de alto risco e além disso, incluindo avaliações de impacto sobre direitos fundamentais (Renda *et al.*, 2021; European Parliament, 2024).

Portanto, no que diz a gestão dos riscos, a regulamentação foca em assegurar que os sistemas de IA de maior risco sejam rigorosamente controlados, enquanto que para sistemas de risco menor, as medidas são proporcionais ao risco que representam, não exigindo um sistema de gestão de risco robusto, mas sim um nível adequado de transparência e controle (European Parliament, 2024).

Antes de utilizar sistemas de IA classificados como de alto risco, é obrigatório que os implementadores realizem uma avaliação de impacto sobre os direitos fundamentais, conforme estipulado no Artigo 27. Essa avaliação tem como objetivo identificar riscos específicos que o sistema possa representar

para os direitos das pessoas afetadas e, a partir disso, adotar as medidas necessárias para mitigar tais riscos. A avaliação deve incluir uma descrição detalhada do processo no qual o sistema será empregado, além de identificar as categorias de pessoas potencialmente afetadas e os riscos específicos relacionados aos direitos fundamentais. Ademais, é necessário atualizar essa avaliação sempre que houver alterações significativas na forma como o sistema é utilizado ou em sua conformidade com os requisitos regulatórios, reportando essas mudanças às autoridades competentes (European Parliament, 2024).

No que se refere à notificação de organismos de avaliação da conformidade, o Artigo 29 estabelece que esses organismos devem submeter um pedido de notificação à autoridade competente do Estado-Membro onde estão estabelecidos. Esse pedido deve ser acompanhado de uma descrição das atividades de avaliação de conformidade que o organismo realizará, bem como dos tipos de sistemas de IA para os quais o organismo declara ter competência. Caso o organismo não possua um certificado de acreditação, ele deve fornecer toda a documentação necessária para comprovar sua conformidade com os requisitos estabelecidos, e essa documentação deve ser atualizada sempre que houver mudanças relevantes (European Parliament, 2024).

Além disso, para garantir que os sistemas de IA de alto risco sejam adequadamente avaliados em condições práticas, o Artigo 60 permite que os provedores ou futuros provedores realizem testes desses sistemas em condições reais, fora dos ambientes controlados conhecidos como "*AI regulatory sandboxes*". Esses testes devem ser conduzidos com base em um plano previamente aprovado pela autoridade de vigilância de mercado do Estado-Membro onde o teste ocorrerá. É essencial que tais testes garantam a conformidade com as revisões éticas exigidas pela legislação da União ou nacional e que qualquer dado coletado durante esses testes só seja transferido para países terceiros se houver garantias adequadas de conformidade com a legislação da União Europeia (European Parliament, 2024).

O Artigo 29 do AI Act é o principal fundamento para a exigência de uma avaliação de impacto especificamente para sistemas de IA de alto risco. Este artigo estabelece que, antes da implementação de tais sistemas, é necessário

realizar uma avaliação do impacto nos direitos fundamentais para identificar e mitigar possíveis riscos. Esta obrigação não se estende de forma explícita a sistemas de IA classificados como de risco limitado ou mínimo, que são tratados de forma diferente no regulamento, com ênfase em medidas como transparência e controles mais leves. Portanto, a avaliação de impacto no AI Act não é um evento único antes da implementação. Ela é um processo contínuo que deve ser revisitado e atualizado ao longo do ciclo de vida do sistema de IA, especialmente em resposta a mudanças significativas ou novos riscos identificados durante o uso. Isso assegura que a gestão de riscos e a proteção dos direitos fundamentais sejam mantidas de maneira dinâmica e adaptável (European Parliament, 2024).

A supervisão humana nos sistemas de Inteligência Artificial (IA) de alto risco é uma exigência crucial prevista nos Artigos 9 e 14 do "Artificial Intelligence Act" (AI Act). Esses artigos estabelecem que tais sistemas devem ser projetados e desenvolvidos de maneira que possibilitem a supervisão eficaz por operadores humanos, garantindo que a tecnologia opere de forma segura e dentro dos limites dos direitos fundamentais. A importância dessa supervisão se dá pelo fato de que, mesmo com a aplicação de medidas técnicas e automáticas de mitigação de riscos, a intervenção humana continua sendo essencial para prevenir ou minimizar potenciais danos à saúde, segurança ou direitos das pessoas (European Parliament, 2024).

Ademais, a supervisão humana deve ser proporcional ao nível de risco, autonomia e contexto de uso do sistema de IA. Isso significa que, quanto maior o risco ou a autonomia do sistema, mais rigorosas devem ser as medidas de supervisão implementadas. As medidas de supervisão podem incluir ferramentas integradas ao sistema, que permitam ao operador humano monitorar, corrigir ou desativar o sistema em situações de risco. Essa abordagem assegura que a supervisão não seja apenas uma formalidade, mas uma parte integral e ativa do funcionamento seguro do sistema (European Parliament, 2024).

Os Artigos 9 e 14 também preveem que as medidas de supervisão sejam definidas pelo fornecedor do sistema antes que ele seja colocado no mercado ou em operação. Isso implica uma responsabilidade significativa para os desenvolvedores, que devem garantir que os usuários finais tenham as

ferramentas e os conhecimentos necessários para exercer a supervisão eficaz. Essa exigência busca assegurar que, independentemente das capacidades técnicas do sistema de IA, o controle humano sobre suas operações seja mantido, proporcionando uma camada adicional de segurança e conformidade legal (European Parliament, 2024).

Por fim, a integração da supervisão humana conforme descrita nos Artigos 9 e 14 reflete uma abordagem equilibrada entre inovação tecnológica e responsabilidade ética. Ao exigir que os sistemas de IA de alto risco incluam mecanismos de supervisão humana, o AI Act promove o desenvolvimento de tecnologias avançadas que respeitam os direitos fundamentais e priorizam a segurança dos indivíduos, estabelecendo um padrão de proteção que combina a autonomia das máquinas com a imprescindibilidade do julgamento humano (European Parliament, 2024).

Os artigos 95, 96, 64 e 65 do "*Artificial Intelligence Act*" estabelecem diretrizes importantes sobre boas práticas e governança na aplicação de sistemas de Inteligência Artificial (IA) na União Europeia. O Artigo 95 promove a criação de códigos de conduta que encorajam a aplicação voluntária dos requisitos estabelecidos pelo regulamento, exceto para sistemas de IA de alto risco que estão sujeitos a regras rigorosas e obrigatórias dentro IA Act. Esses códigos de conduta devem ser desenvolvidos com base nas melhores práticas da indústria e nas soluções técnicas disponíveis, buscando orientar as organizações na implementação de sistemas de IA de forma ética e responsável (European Parliament, 2024).

O Artigo 96 determina que a Comissão Europeia deve fornecer diretrizes claras para a implementação prática do regulamento. Essas diretrizes incluem a aplicação dos requisitos obrigatórios, a identificação de práticas proibidas, a gestão de modificações substanciais em sistemas de IA e a garantia de transparência. O objetivo é assegurar que as organizações compreendam e cumpram os regulamentos de maneira eficaz. Além disso, o Artigo 64 estabelece o Escritório de IA, que é responsável por desenvolver a expertise e as capacidades da União em Inteligência Artificial. Este escritório desempenha um papel central na promoção da governança eficaz da IA na Europa (European Parliament, 2024).

Complementando essa estrutura, o Artigo 65 cria o "*European Artificial Intelligence Board*" (Conselho Europeu de Inteligência Artificial), que coordena e supervisiona a aplicação das normas do AI Act, garantindo uma abordagem harmonizada em todos os Estados-Membros. Esses artigos, em conjunto, fornecem uma estrutura robusta para a governança e boas práticas na implementação de sistemas de IA, promovendo um ambiente regulatório que favorece a inovação responsável e a proteção dos direitos fundamentais (European Parliament, 2024).

Há ocorrência de diversos artigos que abordam a responsabilidade civil dos provedores e operadores de sistemas de IA, assegurando a proteção dos direitos dos consumidores e a conformidade com as normas de segurança. O **Artigo 61** estabelece que os provedores de sistemas de IA são responsáveis por quaisquer danos causados durante os testes em condições reais, fora dos "AI regulatory sandboxes". Nesse contexto, os provedores têm a obrigação de adotar medidas imediatas de mitigação ou, em caso de incidentes graves, suspender os testes, garantindo que sejam responsabilizados de acordo com as leis aplicáveis da União Europeia e nacionais (European Parliament, 2024).

Adicionalmente, o **Artigo 26** define que, caso um fornecedor realize modificações substanciais em um sistema de IA de alto risco, ele será considerado responsável pelo sistema como se fosse o fornecedor original, assumindo todas as obrigações legais e de conformidade relacionadas à responsabilidade por danos causados. Já o **Artigo 31** estabelece que os organismos notificados, responsáveis por avaliar a conformidade dos sistemas de IA, devem ter seguro de responsabilidade adequado para cobrir suas atividades de avaliação, ou o Estado-Membro deve assumir essa responsabilidade. Isso assegura que quaisquer danos resultantes de falhas na avaliação sejam devidamente cobertos (European Parliament, 2024).

O **Artigo 100**, por sua vez, prevê a imposição de multas administrativas a instituições, corpos e agências da União que não cumprirem as práticas de IA proibidas ou outros requisitos. Este artigo considera a natureza, gravidade e duração da infração, além de outros fatores mitigantes ou agravantes, garantindo que as infrações sejam adequadamente penalizadas para reforçar a conformidade com as normas (European Parliament, 2024).

Complementando essas disposições, o Artigo 69 do AI Act trata especificamente da proteção dos direitos dos consumidores em relação à responsabilidade civil, estabelecendo que todas as disposições do regulamento devem ser aplicadas sem prejuízo dos direitos e remédios previstos em outras leis da União, incluindo a Diretiva do Conselho 85/374/EEC sobre responsabilidade por produtos defeituosos. Importante ressaltar que esta Diretiva aplica o princípio da responsabilidade objetiva, o que significa que o produtor é responsabilizado por quaisquer danos causados por defeitos em seus produtos, independentemente de culpa ou negligência. Isso garante que, mesmo no contexto de sistemas de IA, os direitos dos consumidores de buscar compensação por danos permanecem intactos, e as obrigações dos provedores de IA em relação à responsabilidade civil continuam aplicáveis conforme as legislações existentes na União Europeia (European Parliament, 2024).

Esses artigos, em conjunto, garantem que tanto os desenvolvedores quanto os operadores de sistemas de IA sejam devidamente responsabilizados por qualquer dano ou violação decorrente do uso dessas tecnologias, reforçando a conformidade com as normas estabelecidas para proteger os direitos fundamentais e a segurança dos consumidores, especialmente considerando a aplicação da responsabilidade objetiva pela Diretiva 85/374/EEC (European Parliament, 2024).

No contexto do AI Act da União Europeia, as autoridades de vigilância do mercado nos respectivos Estados-Membros desempenham um papel crucial na supervisão dos sistemas de Inteligência Artificial (IA), garantindo sua conformidade com as normas regulamentares estabelecidas. Conforme o Artigo 63 do AI Act, essas autoridades são responsáveis por verificar se os sistemas de IA em operação são seguros e respeitam os direitos fundamentais, além de investigar incidentes que possam surgir durante o uso desses sistemas. Quando um incidente é relatado, essas autoridades têm o dever de avaliar os riscos envolvidos, aplicar as medidas corretivas necessárias e, se necessário, impor sanções aos provedores que não estejam em conformidade com o AI Act (European Parliament, 2024).

A comunicação de incidentes relacionados a sistemas de IA de alto risco deve ser feita imediatamente à autoridade de vigilância do mercado competente no Estado-Membro onde o incidente ocorreu ou onde o sistema está em operação. Conforme as diretrizes do AI Act, essa notificação é fundamental para permitir uma resposta rápida e eficaz, mitigando possíveis danos à saúde, segurança ou aos direitos fundamentais das pessoas. A estrutura de supervisão estabelecida pelo AI Act é essencial para garantir que os sistemas de IA sejam avaliados continuamente e que quaisquer riscos sejam gerenciados de maneira proativa (European Parliament, 2024).

Além disso, as autoridades de vigilância do mercado cooperam com outras entidades nacionais e europeias, como o "*European Artificial Intelligence Board*", para coordenar as ações de supervisão e assegurar uma aplicação harmonizada do regulamento em toda a União Europeia. Essa cooperação, conforme delineado no Artigo 63, é vital para manter um padrão elevado de proteção e segurança na utilização de sistemas de IA, garantindo que as normas sejam aplicadas de forma consistente e que os consumidores e a sociedade estejam protegidos contra os potenciais riscos associados ao uso de tecnologias de IA (European Parliament, 2024).

Os artigos 99, 101, 71 e 100 desempenham um papel crucial ao estabelecer um quadro claro para monitoramento, cumprimento e aplicação de penalidades, assegurando que as disposições do regulamento sejam respeitadas e que os riscos associados ao uso de IA sejam minimizados. O Artigo 99 estabelece as bases para a imposição de penalidades financeiras significativas para aqueles que violam as disposições do AI Act. Ele define que os Estados Membros da União Europeia devem estabelecer regras claras sobre as penalidades, que podem incluir desde advertências até multas administrativas severas. As multas são particularmente rigorosas para infrações graves, como a não conformidade com práticas de IA proibidas descritas no Artigo 5 do regulamento. Nesses casos, as multas podem chegar a até 35 milhões de euros ou até 7% do faturamento anual global da empresa, dependendo de qual valor for maior, quando as práticas forem enquadradas com práticas de IA proibidas. Para outras violações menos graves, as multas podem atingir 15 milhões de euros ou 3% do faturamento anual, enquanto a provisão de informações falsas ou enganosas às autoridades pode resultar em

multas de até 7,5 milhões de euros ou 1% do faturamento anual global (European Parliament, 2024).

O Artigo 101 complementa o Artigo 99 ao focar especificamente nos fornecedores de modelos de IA de propósito geral. Este artigo permite à Comissão Europeia impor multas de até 15 milhões de euros ou 3% do faturamento global anual da empresa por infrações relacionadas ao não cumprimento das obrigações estipuladas pelo regulamento. Isso inclui, por exemplo, a falha em fornecer informações corretas ou a falta de cooperação em medidas solicitadas pela Comissão para mitigar riscos sistêmicos associados ao uso de IA. A ênfase do Artigo 101 está em garantir que os fornecedores de modelos de IA mantenham um alto padrão de conformidade, dada a importância e o impacto potencial desses modelos em diversas aplicações (European Parliament, 2024).

O Artigo 71 aborda o monitoramento pós-comercialização, exigindo que os fornecedores de sistemas de IA considerados de alto risco implementem e mantenham um sistema contínuo de monitoramento. Este sistema é necessário para coletar e analisar dados sobre o desempenho dos sistemas de IA durante seu uso real, com o objetivo de identificar e mitigar riscos que possam surgir após a introdução do produto no mercado. A falha em estabelecer ou manter tal sistema pode levar à imposição de penalidades, conforme delineado no Artigo 99. Este artigo destaca a responsabilidade contínua dos fornecedores em garantir que seus sistemas de IA funcionem de maneira segura e eficiente ao longo do tempo (European Parliament, 2024).

O Artigo 100 complementa o quadro regulatório estabelecendo a necessidade de cooperação e intercâmbio de informações entre as autoridades competentes dos Estados-Membros e a Comissão Europeia. Este artigo garante que as penalidades aplicadas e as medidas corretivas implementadas sejam comunicadas de forma eficiente entre as entidades relevantes, promovendo uma aplicação uniforme e coordenada do regulamento em toda a União Europeia. Essa cooperação é essencial para evitar a fragmentação regulatória e assegurar que as práticas inadequadas sejam tratadas de maneira consistente em todos os Estados-Membros (European Parliament, 2024).

A correlação entre os Artigos 99, 101, 71 e 100 cria um sistema robusto de fiscalização e aplicação de penalidades, assegurando que todos os operadores de IA, desde desenvolvedores até fornecedores de modelos de IA de propósito geral, estejam sujeitos a um regime rigoroso de conformidade. Os valores significativos das multas destacam a seriedade com que a União Europeia aborda as violações das suas regulamentações de IA, enfatizando a importância da transparência, da segurança e da proteção dos direitos fundamentais. As penalidades não só funcionam como um meio de punição, mas também como um forte dissuasor contra práticas negligentes ou maliciosas no desenvolvimento e implementação de IA. Os artigos mencionados trabalham em conjunto para assegurar que o uso de IA na União Europeia esteja em conformidade com os mais altos padrões de ética, segurança e respeito aos direitos fundamentais, estabelecendo penalidades rigorosas para aqueles que não cumprirem com essas obrigações (European Parliament, 2024).

### 4.1.6 Regulamento Geral sobre a Proteção de Dados (GDPR)

**O Regulamento (UE) 2016/679** (Regulamento Geral sobre a Proteção de Dados-GDPR) estabelece regras para a proteção de dados pessoais e garante a privacidade dos indivíduos, conferindo-lhes direitos sobre seus dados pessoais e impondo obrigações às entidades que processam esses dados. Esse ato jurídico da União Europeia constitui a base para um tratamento de dados sustentável e responsável, especialmente quando os conjuntos de dados incluem uma combinação de dados pessoais e não pessoais. O presente regulamento de IA não pretende afetar a aplicação dessas leis da União já em vigor que regulam o tratamento de dados pessoais, incluindo as funções e competências das autoridades de supervisão independentes responsáveis pelo cumprimento desses instrumentos. Essas medidas garantem que, mesmo com o desenvolvimento de novas tecnologias de IA, a proteção de dados pessoais continue a ser uma prioridade, assegurando a privacidade e os direitos dos indivíduos (European Parliament, 2024).

O texto de Hamon, Junklewitz e Sanchez, (2020, p.8), no quadro abaixo explicita quais os artigos da GDPR que tem ligação com a Inteligência Artificial e seu respectivo processamento algorítmico de dados.

**Quadro 1:** Direitos de proteção de dados introduzidos no RGPD vinculados ao processamento algorítmico de dados.

| Direito | Referências do GDPR |
|---|---|
| Direito de ser notificado sobre tomadas de decisão exclusivamente automatizadas | Art. 13, art. 14 |
| Direito de notificação e acesso à informação sobre lógicas envolvidas no tratamento automatizado | Art. 13, art. 14, art. 15 |
| Direito a informação de significância e efeitos potenciais da tomada de decisão exclusivamente automatizada | Art. 13, art. 14. art. 15 |
| Direito de não estar sujeito a decisões exclusivamente automatizadas | Art. 22 |
| Direito de contestar uma decisão em processos de tomada de decisão exclusivamente automatizados | Art. 22 |
| Direito de obter intervenção humana | Considerando 71, art. 22 |
| Direito de obter uma explicação | Considerando 71 |

Fonte: https://ai-watch.ec.europa.eu/system/files/2022-01/dpad_report.pdf. p. 12. Adaptado. Tradução livre. Acesso em: 8 ago. 2024.

A GDPR entrou em vigor em 2018 na Europa sendo de extrema importância para a IA, uma vez que como vimos no quadro acima toca em pontos sensíveis no que tange ao uso de IA em sistemas de tomada de decisão que usam dados pessoais. Assim, o considerando 71 do regulamento destaca a importância de proteger os indivíduos contra decisões baseadas unicamente em processamento automatizado, incluindo a definição de perfis, que possam produzir efeitos jurídicos ou afetá-los significativamente de forma similar. O artigo 22 da lei da GDPR complementa o considerando 71 estabelecendo que os indivíduos têm o direito de não ser sujeitos a decisões baseadas unicamente em processamento automatizado, incluindo a definição de perfis, a menos que certas condições sejam atendidas, como consentimento

explícito ou a necessidade para a execução de um contrato. Estipula que, mesmo quando tais decisões são permitidas, devem existir medidas adequadas para proteger os direitos, liberdades e interesses legítimos dos indivíduos (Hamon, Junklewitz e Sanchez, 2020).

Os artigos 13 e 14 do GDPR estabelecem que os titulares dos dados devem ser informados sobre a existência de processos de tomada de decisão automatizada, incluindo a definição de perfis. Além disso, os responsáveis pelo tratamento de dados devem fornecer informações sobre a lógica subjacente a essas decisões automatizadas, bem como a importância e as possíveis consequências desse processamento. O artigo 15.º concede aos titulares dos dados o direito de acessar seus dados pessoais e obter informações sobre o tratamento realizado, incluindo detalhes sobre qualquer decisão automatizada (Hamon, Junklewitz e Sanchez, 2020).

Assim, a governança de dados surgiu para colocar em prática um conjunto de boas práticas, procedimentos, normas e regras para garantir que os dados sejam gerenciados de forma responsável e de forma lícita para garantir que os direitos fundamentais sejam respeitados na União Europeia em razão da utilização da IA (Hamon, Junklewitz e Sanchez, 2020).

### 4.2. Legislação e Regulação da IA nos Estados Unidos

A National AI Initiative Act de 2020 estabelece o plano estratégico e a estrutura legal para o avanço da IA nos Estados Unidos, enquanto a Executive Order on Safe, Secure, and Trustworthy AI fornece diretrizes específicas para garantir que esse avanço seja seguro, confiável e ético. Ambos os documentos trabalham em conjunto para posicionar os EUA como líder global em IA, ao mesmo tempo em que abordam as implicações de segurança e ética associadas a essa tecnologia.

A regulamentação da inteligência artificial nos Estados Unidos é caracterizada por uma abordagem setorial e fragmentada, sem uma lei federal única que regule todos os aspectos da IA. A regulação ocorre principalmente através de agências federais específicas e legislações estaduais, com várias iniciativas legislativas em andamento para abordar as complexidades e desafios emergentes da IA.

### 4.2.1 National AI Initiative Act

Os Estados Unidos não têm uma legislação federal específica e abrangente para regulamentar a IA[39]. Para empresas que operam nos Estados Unidos, o cenário da regulamentação da IA continua menos claro, pois não há uma legislação federal abrangente semelhante ao UE *AI Act*, nem leis estaduais substanciais especificamente para IA. Contudo, algumas leis estaduais de privacidade podem se aplicar a sistemas de IA que processam dados pessoais (The White House, 2019a; Schreck, Schreck e Charkoudian, 2023).

No entanto, várias iniciativas e ordens executivas estabeleceram diretrizes e normas em diferentes aspectos da IA. A *National AI Initiative Act*, por meio da Ordem Executiva 13859 de fevereiro de 2019, identificou cinco linhas de esforço: 1)Investimento em pesquisa de IA; 2) Liberação de recursos federais de computação e dados de IA; 3) Definição de padrões técnicos de IA; 4) Construção da força de trabalho de IA americana; 5) Envolvimento com aliados internacionais. Essas ações foram consolidadas na lei *National AI Initiative Act* de 2020, que inclui um plano regulatório e diretrizes para o uso federal de IA e estabelece padrões técnicos para a tecnologia. Assim, o *National AI Iniciative Act de 2020* trata de aspectos específicos do desenvolvimento e implementação da IA com foco maior em promover a inovação, pesquisa e desenvolvimento ético da IA do que em estabelecer uma regulação abrangente. O documento "*American Artificial Initiative: Year One Annual Report*" reforça estes setores como foco de investimento e destaca o compromisso dos Estados Unidos em liderar o desenvolvimento responsável e ético da inteligência artificial, enquanto se prepara para os desafios e oportunidades da era digital. O foco é tanto na manutenção da liderança tecnológica quanto na promoção de práticas que garantam a confiança pública

[39] O termo “inteligência artificial” ou “IA” tem o significado estabelecido em 15 U.S.C. 9401(3): um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas de inteligência artificial utilizam dados baseados em máquinas e humanos para perceber ambientes reais e virtuais; abstrair tais percepções em modelos por meio de análise de forma automatizada; e usar a inferência de modelos para formular opções de informação ou ação (The White House, 2023).

e o respeito aos valores fundamentais do país (The White House, 2019a; The White House, 2020b; Schreck, Schreck e Charkoudian, 2023).

Assim, como visto acima, as 5 (cinco) linhas de esforços significam que os investimentos em pesquisa tem o objetivo de fortalecer e desenvolver um ambiente vibrante da indústria, academia, e governo dos EUA e prioriza os gastos federais com ideias de alto nível que beneficiarão o povo americano. A iniciativa também priorizará o acesso a recursos federais de IA para promover a inovação e a segurança, enquanto mantém padrões rigorosos de proteção e privacidade. Outra parte importante da iniciativa é o papel das agências federais de promover a confiança pública através do estabelecimento de orientações para o desenvolvimento e uso da IA, neste tocante, contam com o Instituto Nacional de Padrões e Tecnologia (NIST)[40] para desenvolver padrões técnicos apropriados e por fim, e não menos importante, a iniciativa se preocupa com o futuro dos empregos nos EUA e pede para que as agências priorizem programas de bolsas de qualificação para ajudar o trabalhador americano a desenvolver habilidades relevantes para a IA (The White House, 2019b).

O documento "*American Artificial Initiative: Year One Annual Report*" (2020) ressalta que além de investimento em pesquisa e desenvolvimento em IA é necessário a liberação de recursos de IA- a iniciativa defende o aumento do acesso a dados federais de alta qualidade, modelos e recursos computacionais. O governo federal está comprometido em tornar esses recursos mais acessíveis para pesquisadores e desenvolvedores de IA, enquanto assegura que as questões de segurança, privacidade e confidencialidade sejam mantidas. O uso da computação em nuvem também é incentivado para acelerar a inovação em IA. O documento aborda a necessidade de revisar e ajustar as regulamentações existentes para evitar obstáculos desnecessários à inovação em IA. Além disso, destaca o papel do governo na promoção do desenvolvimento de padrões técnicos que garantam a segurança, confiabilidade e interoperabilidade das tecnologias de IA, ajudando

[40] NIST AI Risk Management Framework (2023):O Instituto Nacional de Padrões e Tecnologia (NIST) publicou um quadro de gerenciamento de riscos da IA, que serve como um guia voluntário para empresas que desenvolvem, implantam ou utilizam sistemas de IA. Este quadro visa melhorar a confiabilidade e segurança dos sistemas de IA através de diretrizes sobre práticas seguras, resilientes, explicáveis e justas (Schreck, Schreck e Charkoudian, 2023)

a facilitar sua adoção em várias indústrias (The White House, 2020b).

Além disso, destaca o papel do governo na promoção do desenvolvimento de padrões técnicos que garantam a segurança, confiabilidade e interoperabilidade das tecnologias de IA, ajudando a facilitar sua adoção em várias indústrias. Por conseguinte, reconhece que a IA está transformando a economia e o mercado de trabalho, o relatório enfatiza a importância de capacitar os trabalhadores americanos com as habilidades necessárias para prosperar na era da IA. Isso inclui a promoção de educação em ciência, tecnologia, engenharia e matemática (STEM), além de programas de requalificação e aprendizado contínuo que sejam inclusivos e acessíveis para todos (The White House, 2020b).

O governo dos EUA está engajado em promover colaborações internacionais que apoiem a pesquisa e desenvolvimento em IA e abram mercados globais para as indústrias americanas de IA. Os Estados Unidos buscam liderar o desenvolvimento de princípios globais para a gestão confiável da IA, promovendo a inovação de maneira que respeite os direitos humanos e os valores democráticos e está focado em incorporar tecnologias de IA em operações governamentais para melhorar a eficiência e a prestação de serviços públicos. O governo está comprometido em garantir que o uso de IA em suas operações seja transparente, seguro e alinhado com os valores fundamentais da nação, incluindo a proteção de direitos civis e liberdades (The White House, 2020b).

A *National AI Initiative Act* de 2020 estabelece uma estrutura abrangente para o desenvolvimento e uso de IA nos Estados Unidos, mas não aborda diretamente temas como gestão de riscos, avaliação de impacto algorítmico, supervisão humana, responsabilidade civil e penalidades. A lei foca principalmente na promoção da IA de forma ética, segura e responsável, delegando a responsabilidade pelo desenvolvimento de padrões técnicos e diretrizes éticas a agências federais como o NIST. Esses padrões são essenciais para garantir a segurança e confiabilidade das tecnologias de IA, mas a lei não fornece detalhes específicos sobre a implementação de uma estrutura formal de gestão de riscos ou mecanismos específicos de supervisão (The White House, 2020a; The White House, 2020b).

Embora a lei não trate diretamente de responsabilidade civil ou direito do consumidor, ela sugere que o desenvolvimento de IA deve ser alinhado com princípios éticos que protejam os consumidores de impactos negativos, como decisões automatizadas injustas e violações de privacidade. Da mesma forma, a comunicação de incidentes relacionados a IA, como falhas ou violações de segurança, não é mencionada explicitamente, mas é implícita na ênfase em transparência e confiabilidade. Essas questões, no entanto, são geralmente reguladas por outras legislações específicas, como regulamentos de segurança cibernética e proteção de dados, que complementam a estrutura estabelecida pela National AI Initiative Act (The White House, 2020a; The White House, 2020b).

Em suma, a **National AI Initiative Act** oferece uma base sólida para o avanço da IA nos Estados Unidos, promovendo o desenvolvimento seguro e ético das tecnologias. No entanto, deixa muitos detalhes práticos, como gestão de riscos, proteção ao consumidor e comunicação de incidentes, para serem abordados por outras legislações e regulamentações específicas, com a expectativa de que as agências federais desenvolvam padrões técnicos e diretrizes que abordem essas questões de forma mais detalhada (The White House, 2020a; The White House, 2020b).

De acordo com esta perspectiva, a Casa Branca destaca que os Estados Unidos não precisam escolher entre liberdade e tecnologia. A convicção é de que é possível aliar valores como liberdade, direitos humanos e respeito pela dignidade humana ao avanço da IA. Como parte estratégica desse desenvolvimento, princípios regulatórios foram estabelecidos para orientar o desenvolvimento da IA no setor privado, integrando a Iniciativa Americana de regulação da IA. O governo acredita que, com esse marco regulatório, será mais capaz de abordar os desafios éticos que surgirem ao longo do processo de desenvolvimento da IA. Assim, o *National AI Initiative Act* estabelece uma estrutura geral e legal para o avanço da IA nos EUA (The White House, 2020a; The White House, 2020b).

### 4.2.2 Executive Order on Safe, Secure, and Trustworthy AI

A *Executive Order on Safe, Secure and Trustworthy AI* visa estabelecer diretrizes para o desenvolvimento de tecnologias de IA que protejam a

segurança nacional, privacidade, direitos civis e outras considerações éticas. Ela pode incluir regras para o uso de IA em diversas áreas, como defesa, economia, saúde, e pode exigir a cooperação entre agências governamentais, setor privado e a comunidade de pesquisa para alcançar esses objetivos. A ordem executiva são emitidas diretamente pelo Presidente dos EUA e não passam pelo Congresso Nacional, são direcionadas às agências do governo federal, regulando como essas agências devem operar dentro do âmbito da lei existente (The White House, 2023).

Os princípios e políticas para o desenvolvimento da IA são destacados na ordem Executiva (*Executive Order on Safe, Secure, and Trustworthy AI*)[41]: 1) desenvolvimento de IA segura e responsável; 2) promoção da inovação, da concorrência e da colaboração responsáveis; 3) desenvolvimento e utilização responsáveis da IA; 4) políticas de IA devem ser consistentes dedicadas ao avanço da equidade e direitos civis; 5) interesses dos americanos que usam ou compram IA devem ser resguardados; 6) privacidade[42] e liberdade civis dos americanos devem ser protegidos à medida que a IA avança; 7) a importância de gerir riscos decorrentes utilização da IA pelo próprio governo federal; 8) Governo federal deve liderar o caminho para o progresso social, econômico e tecnológico. Esses princípios e políticas foram formalizados por várias ações governamentais, como a Ordem Executiva sobre Desenvolvimento Seguro, Seguro e Confiável de IA e a adesão dos EUA aos princípios globais da OCDE para a IA confiável. Essas diretrizes visam promover a inovação enquanto protegem os direitos e liberdades civis dos indivíduos (The White House, 2023).

Assim, a ***National AI Initiative Act de 2020*** e a ***Executive Order on Safe, Secure, and Trustworthy AI*** são complementares na estratégia dos Estados Unidos para liderar o desenvolvimento de IA. A *National AI Initiative Act* é uma lei federal que estabelece a base legal e estratégica para promover a pesquisa, o desenvolvimento e a aplicação ética da IA em diversas áreas, como defesa, saúde e educação. Por outro lado, a *Executive Order on Safe,*

[41] Esta ordem executiva destaca a importância de proteger a privacidade e as liberdades civis enquanto a IA avança. Ela exige que as agências federais utilizem ferramentas políticas e técnicas para proteger a privacidade e mitigar riscos associados ao uso inadequado de dados pessoais pela IA (The White House, 2023)

[42] As agências usarão as ferramentas políticas e técnicas disponíveis, incluindo tecnologias de melhoria da privacidade (PETs), quando apropriado, para proteger a privacidade e combater os riscos legais e sociais mais amplos - incluindo a restrição dos direitos da Primeira Emenda - que resultam da coleta e uso indevidos de dados (The White House, 2023)

*Secure, and Trustworthy AI* é uma diretiva presidencial que fornece diretrizes específicas para garantir que as tecnologias de IA sejam implementadas de maneira segura, confiável e ética, nas agências federais, em conformidade com os objetivos da National AI Initiative. A ordem executiva detalha as práticas que as agências federais devem adotar para garantir que a IA utilizada pelo governo proteja a segurança nacional, a privacidade e os direitos civis. Enquanto a *National AI Initiative Act* estabelece o framework geral e legal para o avanço da IA nos EUA, a ordem executiva orienta a implementação prática desses princípios, focando na segurança e na ética. Juntas, elas garantem que o desenvolvimento da IA seja feito de maneira responsável e competitiva (The White House, 2023; Schreck, Schreck e Charkoudian, 2023).

### 4.2.3 Legislação nos estados dos EUA

Segundo Mandiega (2019, p. 10) nos EUA, várias empresas de tecnologia já desenvolveram códigos de conduta para IA e ética, mas há um crescente pedido por uma regulamentação mais robusta liderada pelo governo. Grupos colaborativos, como a Partnership on AI, que inclui Microsoft, Amazon, Facebook e Apple, se comprometeram a desenvolver e compartilhar as melhores práticas, especialmente no campo da ética. A *Association for Computing Machinery* (ACM) também publicou um Código de Ética e Conduta Profissional em 2018 para orientar os profissionais de informática. Empresas como Microsoft e Google têm suas próprias diretrizes éticas. Por exemplo, a Microsoft possui um conselho consultivo de IA, enquanto o Google divulgou seus princípios de IA, uma carta de ética para guiar o desenvolvimento e uso responsável da IA em suas pesquisas e produtos.

Apesar desses esforços de autorregulação, há uma preocupação crescente de que isso não seja suficiente para enfrentar os desafios éticos da IA. Em 2018, o Instituto *AI Now* emitiu um relatório concluindo que as estruturas de governança interna na maioria das empresas de tecnologia não garantem a responsabilidade pelos sistemas de IA. O relatório argumenta que as agências governamentais precisam ter mais poder para supervisionar, auditar e monitorar as tecnologias de IA, especialmente aquelas que envolvem reconhecimento facial. Essas medidas adicionais de supervisão são essenciais para garantir que a IA seja desenvolvida e utilizada de forma ética e

responsável, protegendo os direitos dos consumidores e prevenindo abusos tecnológicos que ferem os direitos fundamentais (Mandiega, 2019).

Os princípios orientadores para os estados que compõem os EUA na governança do design, desenvolvimento e uso da IA devem seguir as seguintes diretrizes: 1) diálogo colaborativo (diálogo com partes interessadas de diversas disciplinas); 2) proteção contra impactos não intencionais (proteger indivíduos de impactos ou usos não intencionais, mas previsíveis, de um sistema de IA inseguro ou ineficaz ; 3) Autonomia e proteção de dados (proteger individuos de práticas abusivas de dados e garantir que eles tenham autonomia sobre como um sistema de IA coleta e usa seus dados); 4) transparência e opção de escolha (assegurar que os individuos saibam como o sistema de IA está sendo utilizado e conceder a oportunidade do usuário optar por não usar um sistema de IA em favor de uma alternativa humana) ; 5) proteção contra discriminação (proteger os indivíduos da discriminação e garantir que os sistemas de IA sejam projetados de forma equitativa); 6) Responsabilização (garantir que aqueles que desenvolvem e implantam sistema de IA cumpram regras e padrões que regem estes sistemas e sejam responsabilizados caso não os cumpram) (The Council Of State Governments, 2023).

Segundo o autor Parinandi *et al.* (2024, p. 4), embora o governo federal tenha sido bastante passivo em relação à regulamentação da IA, os estados dos EUA começaram a tomar iniciativas próprias para lidar com a inovação da IA e a proteção do consumidor. Em 2011, Nevada foi pioneiro ao se tornar o primeiro estado a permitir o uso de veículos autônomos. Seguindo esse exemplo, outros estados, como o Tennessee, criaram leis que impedem governos locais de proibir veículos que utilizam IA. Com o tempo, e após avaliarem os impactos dos veículos autônomos, os estados passaram a propor leis focadas na segurança, proteção de dados e privacidade dos consumidores. Essa regulamentação mista permite que a inovação continue, enquanto protege os cidadãos dos riscos associados. Recentemente, as legislaturas estaduais têm ampliado seu foco para abordar outras implicações da IA, indo além dos veículos autônomos e abordando possíveis ameaças à segurança dos consumidores em diferentes áreas.

Embora a IA traga muitos benefícios econômicos, seus riscos para mercados e consumidores exigem regulamentação urgente. A IA beneficia empresas e consumidores, mas seu uso crescente necessita de uma legislação equilibrada que promova a inovação e proteja os consumidores. Há consenso sobre a necessidade de regulamentação, mas divergências sobre o grau de restrição. A proteção do consumidor é a principal razão para regulamentar a IA, devido aos potenciais riscos econômicos e individuais. Os regimes regulatórios variam de altamente restritivos a mais permissivos. Alguns defendem que as leis existentes de proteção ao consumidor, como na regulamentação bancária, podem ser adaptadas para a IA, enquanto outros acreditam na necessidade de novas legislações específicas. Este debate reflete a complexidade de equilibrar a inovação tecnológica com a proteção dos direitos dos consumidores (Parinandi *et al.*, 2024).

Existem alguns exemplos de legislação feita pelos estados nos EUA em proteção aos consumidores e que abordam questões éticas no que tange ao desenvolvimento e aplicação da IA. Em 2019, Delaware aprovou uma lei para que suas agências estaduais planejem e minimizem o impacto da IA na perda de empregos. Alabama e Illinois criaram políticas para formar comitês e programas focados na promoção da inovação em IA e seu efeito no desenvolvimento econômico. Nova Jersey também aprovou uma legislação que encarrega o Comissário do Trabalho e Desenvolvimento da Força de Trabalho de estudar como a IA afeta o crescimento econômico. Essas iniciativas mostram como diferentes estados estão respondendo aos impactos econômicos da IA de maneira proativa e estratégica (Parinandi *et al.*, 2024).

O Estado do Illinois agora proíbe empregadores e credores de usar IA que considere características raciais ao analisar elegibilidade para emprego ou crédito (Illinois HB 2557)(The Council of State Governments -) (U.S. Chamber of Commerce). No Colorado, as seguradoras não podem usar algoritmos que discriminem com base em raça, sexo, gênero e outras características (Colorado SB 21-169)(The Council of State Governments -)(Brennan Center for Justice). Em Idaho, é proibido o uso de vieses algorítmicos na determinação de sentenças e fianças para réus. Esses exemplos mostram que os estados estão conscientes das questões éticas envolvidas no desenvolvimento e aplicação dos riscos da IA para os consumidores e estão criando leis para proteger seus

direitos de maneira segura e responsável (The Council Of State Governments, 2023; Parinandi *et al.*, 2024).

Nota-se que quatro estados - Illinois, Nova York, Texas e Vermont - aprovaram leis para garantir que o design, desenvolvimento e uso da IA sejam baseados em um diálogo colaborativo com especialistas de várias áreas. Pode-se explicitar com as seguintes legislações **Illinois** (HB 3563, 2023); **Nova York** (AB A4969, 2023 e SB S3971 B, 2019); **Texas** (HB 2060, 2023) e **Vermont** (HB 378, 2018). Esses estados criaram forças-tarefa, grupos de trabalho ou comitês para estudar os impactos da IA nos consumidores e no setor público, além de identificar desafios de segurança cibernética. Esses grupos têm a missão de recomendar novas leis ou regulamentos para proteger os consumidores, promovendo um uso seguro e eficaz da IA (The Council Of State Governments, 2023).

Quatro estados — Califórnia (AB 302, 2023), Connecticut (SB 1103, 2023), Louisiana (SCR 49, 2023) e Vermont (HB 410, 2022) — aprovaram leis para proteger indivíduos contra impactos previsíveis e não intencionais de sistemas de IA inseguros ou ineficazes. Essas leis exigem que agências estaduais revisem os sistemas de IA em uso, relatem ao governador os riscos potenciais e proponham políticas ou códigos de ética para abordar essas preocupações.No estado de Nova York a NYC 2021/144 exige que os empregadores divulguem quando e como os sistemas de IA são usados, especialmente em processos de contratação, e que obtenham consentimento dos empregados para a coleta de dados (The Council Of State Governments, 2023; Parinandi *et al.*, 2024).

Onze estados — Califórnia, Colorado, Connecticut, Delaware, Indiana, Iowa, Montana, Oregon, Tennessee, Texas e Virgínia — promulgaram leis para proteger indivíduos de práticas abusivas de dados e garantir que eles tenham controle sobre como seus dados são coletados e usados por sistemas de IA. Essas leis permitem que os consumidores optem por não participar da "criação de perfil" quando isso impacta processos automatizados de tomada de decisão que produzem efeitos legais ou significativos, como tratamento injusto, impactos negativos na saúde ou finanças, e serviços financeiros, de empréstimo, moradia, seguro ou educação. No Texas, foi criado um Conselho Consultivo de IA para estudar o impacto dos sistemas de IA usados por

agências estaduais e avaliar a necessidade de um código de ética estadual (Brennan Center for Justice). Essas leis mostram que os estados estão se preparando para enfrentar questões éticas da IA, protegendo consumidores contra discriminação e promovendo transparência e responsabilidade (The Council Of State Governments, 2023; Parinandi *et al.*, 2024)

Três estados — Califórnia (SB 1001, 2023), Illinois (HB 2557, 2019) e Maryland (HB 1202, 2020) — e a cidade de Nova York (2021/144, 2021) aprovaram leis exigindo que empregadores e empresas informem quando e como um sistema de IA está sendo usado. Em alguns casos, é necessário obter o consentimento dos funcionários para o uso de IA que coleta seus dados. Três estados — Califórnia (SB 36, 2019), Colorado (SB 21-169, 2021) e Illinois (HB 0053, 2021) — aprovaram leis para proteger indivíduos contra discriminação e garantir que os sistemas de IA sejam projetados de forma justa. Essas leis visam prevenir a discriminação algorítmica, onde a IA pode tratar pessoas de maneira injusta com base em raça, cor, etnia, sexo, religião ou deficiência, entre outros fatores (The Council Of State Governments, 2023).

E por fim, Doze estados — Califórnia, Colorado, Connecticut, Delaware, Indiana, Iowa, Montana, Oregon, Tennessee, Texas, Virgínia e Washington — aprovaram leis para garantir que os desenvolvedores e implementadores de sistemas de IA estejam em conformidade com as regras e padrões de IA e sejam responsabilizados caso não cumpram. Entretanto, Washington ainda precisa aprovar leis de privacidade de dados que explicitamente governem os sistemas de IA. O estado estabeleceu um grupo de trabalho para estudar como os sistemas automatizados de tomada de decisão podem ser revisados e auditados periodicamente para garantir que sejam justos, transparentes e responsáveis. Essas legislações visam assegurar que a IA seja usada de maneira ética e segura, protegendo os indivíduos de possíveis abusos e garantindo transparência e responsabilidade no uso dessas tecnologias (The Council Of State Governments, 2023).

Pelo fato dos EUA ter uma economia baseada no liberalismo econômico, os estados têm uma maior liberdade para legislar. No tocante a aprovação de legislação de IA pelos estados é fato que a política (partido Republicano ou

Democrata)[43] do estado irá contribuir ou não para a aprovação de uma legislação. No geral, as condições econômicas influenciam a adoção de políticas de IA. Estados com maior desemprego e inflação tendem a evitar políticas de IA, provavelmente porque a tecnologia é vista como um risco para os empregos. Por outro lado, estados com economias maiores são mais propensos a adotar essas políticas, possivelmente porque têm mais recursos para gerenciar o setor emergente da IA. Isso indica que a saúde econômica e o tamanho da economia de um estado são fatores importantes na decisão de implementar políticas de IA (Parinandi *et al.*, 2024).

Existem dois tipos de legislação de IA nos estados que compõem os EUA, uma centrada na proteção dos consumidores e a outra centrada nos negócios e na economia. O que se percebe é que a grande maioria das legislações está sendo construída para a proteção dos consumidores. Isto deve-se a uma explicação óbvia como uma forma de proteger o mercado interno de grandes impactos econômicos. Apesar de os estados que possuem maior renda per capita e que têm menores problemas com taxas de desemprego terem uma abordagem mais voltada para a construção de legislação no sentido de alavancar o potencial que a aplicação da IA gera no desenvolvimento econômico nos estados, o contrário se percebe nos estados com menor renda per capita que possuem uma legislação mais consumerista e protetiva (Parinandi *et al.*, 2024).

A grande questão problemática será com o passar do tempo, no qual os estados que possuem um governo centrado em sua grande maioria no partido democratas terá uma política legislativa voltada à proteção dos consumidores maior do que uma política legislativa centrada nos negócios e na economia. Desta forma, a pergunta que se faz é como estes estados irão sopesar as duas realidades para obter maiores  e melhores avanços tecnológicos, unindo a proteção ao consumidor. O contrário também é preocupante já que os estados com uma política de governo centrada no partido republicano terá um maior avanço na legislação de negócios e na economia em detrimento a uma legislação focada na proteção dos consumidores. A esperança é que as legislações e regulamentações às quais as empresas estão sujeitas mudem de

[43] A existência de efeitos partidários sugere que a proteção do consumidor na IA tem maior probabilidade de ser apoiada pelos Democratas.

acordo com as questões políticas que envolvem a IA como o crescimento econômico, proteção dos direitos fundamentais e conforme a necessidade de segurança e inovação de cada ente federativo (Parinandi *et al.*, 2024).

4.2.4 Documentos que abordam sobre a gestão de risco e avaliação de impacto dos sistemas de Inteligência Artificial.

4.2.4.1 *AI Risk Management Framework (AI RMF)-Framework NIST*

O *AI Risk Management Framework (AI RMF)-Framework NIST* é um documento de gerenciamento de riscos de IA que pode impulsionar o uso de práticas responsáveis e confiáveis no ecossistema de desenvolvimento de sistemas de IA. Este framework é fundamental para auxiliar organizações a navegar pelas complexidades inerentes ao desenvolvimento e à implementação de IA, garantindo que os sistemas sejam não apenas tecnicamente competentes, mas também alinhados com valores éticos e legais. A estrutura proposta pelo AI RMF permite que as empresas identifiquem, avaliem e priorizem os riscos de forma sistemática, possibilitando a implementação de controles adequados e a adoção de estratégias de mitigação que sejam proporcionais aos riscos identificados (NIST, 2023).

De acordo com a Lei da Iniciativa Nacional de Inteligência Artificial de 2020, o AI RMF tem como objetivo servir como um recurso para organizações que criam, desenvolvem, implantam ou utilizam sistemas de IA, auxiliando-as na gestão dos diversos riscos associados a essa tecnologia e incentivando práticas de desenvolvimento e uso que sejam confiáveis e responsáveis. Essa estrutura foi concebida para ser opcional, respeitadora de direitos, não vinculada a um setor específico e aplicável a diferentes cenários de uso, oferecendo flexibilidade para que organizações de todos os portes, setores e segmentos da sociedade possam implementar as abordagens sugeridas. Para a AI RMF o termo riscos negativos são consequências prejudiciais e risco positivo refere-se a riscos positivos (NIST, 2023).

A gestão de riscos em sistemas de IA, conforme descrita no AI RMF, enfrenta diversos desafios significativos que precisam ser abordados para garantir a confiabilidade dessas tecnologias. Um dos principais desafios é a **medição de risco**, que se torna particularmente complicada quando os riscos

ou falhas não são bem definidos ou compreendidos, o que dificulta a quantificação precisa desses riscos. Mesmo que a medição adequada não seja possível, isso não necessariamente indica que o risco seja alto ou baixo. Além disso, os riscos relacionados a software, hardware e dados de terceiros complicam ainda mais essa medição, especialmente quando as metodologias de avaliação entre as partes envolvidas não estão alinhadas ou quando falta transparência. A forma como os clientes integram esses componentes pode introduzir riscos adicionais. Da mesma forma, o acompanhamento de riscos emergentes é essencial, embora identificar e monitorar novos riscos e desenvolver técnicas para medi-los seja uma tarefa desafiadora. Criar métricas precisas pode ser problemático, pois elas podem ser simplificadas demais, manipuladas ou não considerar adequadamente as diferenças entre os grupos afetados (NIST, 2023).

Adicionalmente, a medição de risco em diferentes fases do ciclo de vida da IA revela que riscos não aparentes em estágios iniciais podem surgir e aumentar à medida que o sistema evolui, o que, juntamente com as diferentes perspectivas das partes envolvidas, torna a avaliação de riscos mais complexa.O risco em ambientes do mundo real é diferente daquele risco medido em um laboratório. Outro fator complicador é a inescrutabilidade dos sistemas de IA, que dificulta a medição de riscos devido à falta de clareza e transparência, especialmente quando a explicabilidade ou a documentação são limitadas. Finalmente, a linha de base humana representa um desafio ao comparar o desempenho da IA com a atividade humana, já que os sistemas de IA operam de maneira diferente dos humanos, tornando difícil estabelecer métricas de comparação padronizadas. Esses desafios mostram a complexidade inerente à gestão de riscos em IA, exigindo que as organizações adotem abordagens robustas e adaptáveis para lidar com essas questões (NIST, 2023).

Embora o AI RMF ajude a priorizar riscos, ele não define diretrizes específicas sobre tolerância ao risco, que é a disposição de uma organização ou agente de IA para aceitar riscos visando atingir objetivos. Essa tolerância é influenciada por exigências legais, políticas e normas, e varia conforme o contexto, a aplicação e as prioridades organizacionais. As tolerâncias ao risco

podem mudar com o tempo, à medida que sistemas e normas evoluem, e diferem entre organizações devido a suas prioridades e recursos específicos.

Ao aplicar o RMF de IA, os riscos mais elevados identificados em um contexto específico devem ser priorizados e gerenciados com maior rigor. Se um sistema de IA apresentar riscos inaceitáveis, como impactos iminentes, danos graves ou catastróficos, o desenvolvimento e a implantação devem ser interrompidos até que esses riscos sejam adequadamente controlados. Sistemas de baixo risco podem ter uma priorização menor, mas a avaliação regular dos riscos é essencial, especialmente para aqueles que interagem diretamente com humanos ou lidam com dados sensíveis. Sistemas que interagem apenas com outros sistemas computacionais podem exigir menor atenção inicial, mas ainda precisam ser monitorados. O risco residual, que permanece após a mitigação, deve ser documentado e comunicado aos usuários finais para alertá-los sobre possíveis impactos negativos (NIST, 2023).

Para reduzir os riscos negativos de IA e que a IA seja considerada confiável é necessário que os sistemas tenham as seguintes características- válidos e confiáveis, seguros, protegidos e resiliente, responsáveis e transparentes, explicáveis e interpretáveis, com maior privacidade, justo (com preconceito prejudicial gerenciado). De acordo com o AI Risk Management Framework (AI RMF), os sistemas de IA devem ser projetados para serem válidos e confiáveis, assegurando resultados precisos e consistentes, e também devem ser seguros, protegidos contra ameaças cibernéticas e resilientes, capazes de operar eficazmente mesmo em condições adversas. Além disso, é essencial que esses sistemas sejam responsáveis e transparentes, permitindo a monitorização e o controle humano, e que sejam explicáveis e interpretáveis, de modo que as decisões possam ser compreendidas e auditadas (NIST, 2023).

A proteção da privacidade é outra prioridade, com práticas que garantam o uso responsável dos dados pessoais. Também é crucial que os sistemas de IA sejam justos, com preconceitos prejudiciais ativamente gerenciados para evitar discriminação e garantir equidade. Essas características são fundamentais para construir sistemas de IA confiáveis, éticos e aceitos pela sociedade, minimizando riscos enquanto se maximiza o benefício social. A robustez significa “capacidade de um sistema de manter seu nível de

desempenho sob uma variedade de circunstâncias” (NIST, 2023. p. 14).

No contexto do AI Risk Management Framework (AI RMF), as funções de Governar (Governança)**,** Mapear (Mapeamento), Medir (medição) e Gerenciar (gestão) constituem os elementos centrais do framework, cada uma desempenhando um papel crucial na gestão eficaz dos riscos associados aos sistemas de inteligência artificial. A função de Governar envolve a criação de estruturas e processos de governança que garantem a supervisão adequada dos sistemas de IA. Essa função abrange o estabelecimento de políticas, normas e diretrizes que orientam o desenvolvimento, a implementação e o uso da IA, assegurando que as decisões estejam alinhadas com os valores éticos, regulatórios e estratégicos da organização. Além disso, Governar implica na definição clara de responsabilidades e na criação de mecanismos de prestação de contas, que são essenciais para monitorar o cumprimento dessas diretrizes (NIST, 2023).

A função de Mapear refere-se à identificação e compreensão dos riscos inerentes aos sistemas de IA em diferentes contextos e fases do seu ciclo de vida. Isso inclui a análise do ambiente operacional, a identificação dos possíveis riscos e a avaliação do impacto que esses riscos podem ter sobre a organização e seus stakeholders. Ao mapear os riscos, as organizações podem obter uma visão mais clara do cenário de ameaças e, assim, elaborar estratégias de mitigação mais eficazes. A função de Medir envolve a avaliação e quantificação dos riscos identificados, utilizando métricas, testes e outras metodologias que ajudam a determinar a gravidade e a probabilidade dos riscos associados aos sistemas de IA. Essa função também inclui o monitoramento contínuo do desempenho dos sistemas de IA e da eficácia das estratégias de mitigação implementadas, permitindo, assim, ajustes e melhorias conforme necessário (NIST, 2023).

Finalmente, a função de Gerenciar concentra-se na implementação de estratégias destinadas a mitigar, transferir, aceitar ou evitar os riscos identificados. Isso envolve a aplicação de controles, a realização de auditorias e a adaptação das práticas de desenvolvimento e operação dos sistemas de IA com base nas avaliações de risco. Nesta função, a supervisão humana é essencial para o gerenciamento dos riscos. Além disso, Gerenciar inclui a comunicação dos riscos residuais aos stakeholders e a adoção de uma

abordagem proativa para lidar com novos riscos que possam surgir ao longo do tempo. Essas funções são interdependentes e, em conjunto, constituem a base para uma gestão robusta e eficaz dos riscos de IA, permitindo que as organizações enfrentem os desafios éticos, técnicos e sociais associados ao uso da inteligência artificial de maneira integrada e adaptativa (NIST, 2023).

Assim, o documento traz a supervisão humana em momentos específicos que tratem de garantir que as operações do sistema de IA estejam em conformidade com normas éticas e regulatórias, onde a presença de uma supervisão humana adequada é essencial. Assim, destaca-se a importância de manter um equilíbrio entre a automação e a intervenção humana, especialmente em cenários onde as decisões de IA têm impactos significativos sobre indivíduos, grupos ou a sociedade em geral (NIST, 2023).

O AI RMF utiliza perfis de casos de uso para aplicar suas funções, categorias e subcategorias a cenários ou aplicações específicas, levando em conta os requisitos, a tolerância ao risco e os recursos do usuário. Esses perfis ajudam a gerenciar riscos em diferentes estágios do ciclo de vida da IA e setores, alinhando a gestão de riscos aos objetivos organizacionais e às exigências legais. Além disso, os perfis temporais do AI RMF descrevem o estado atual e o estado desejado das práticas de gerenciamento de risco de IA, permitindo que as organizações identifiquem e abordem lacunas entre o presente e o futuro ideal. Essa abordagem facilita a priorização de ações de mitigação e a comparação de estratégias com outras, otimizando o uso de recursos para atingir os objetivos de gestão de riscos de IA. Os perfis intersetoriais tem função proporcionar diretrizes e melhores práticas aplicáveis a diferentes setores ou indústria (NIST, 2023).

O apêndice A serve como um guia para entender as responsabilidades de cada ator em relação à gestão de riscos e à garantia de que os sistemas de IA sejam implementados e operados de forma a minimizar riscos e maximizar benefícios. Ele oferece uma visão clara de como cada participante contribui para o sucesso do gerenciamento de riscos em IA, facilitando a coordenação entre diferentes funções e garantindo que todos os aspectos críticos sejam adequadamente abordados. As tarefas de avaliação de impacto de IA são essenciais para garantir que os sistemas de inteligência artificial sejam desenvolvidos e implementados de forma responsável e ética. Essas tarefas

incluem a avaliação dos requisitos de responsabilização do sistema, a mitigação de preconceitos prejudiciais, a análise dos impactos em segurança, responsabilidade e proteção dos produtos. Profissionais como assessores e avaliadores de impacto, que possuem conhecimentos técnicos, humanos, socioculturais e jurídicos, desempenham papéis cruciais nesse processo (NIST, 2023).

Por outro lado, as tarefas de governança e supervisão são realizadas por atores de IA com autoridade gerencial, fiduciária e legal, que são responsáveis por garantir que o desenvolvimento e a implementação dos sistemas de IA estejam alinhados com os objetivos e a sustentabilidade da organização. Os principais responsáveis pela governança da IA incluem a gestão organizacional, a liderança sênior e o Conselho de Administração, que se preocupam com o impacto dos sistemas de IA na organização como um todo e garantem a conformidade com as normas éticas e legais. Em resumo, os quadros do AI RMF são uma ferramenta prática e abrangente que orienta as organizações sobre como gerenciar os riscos de IA de maneira estruturada e eficaz, desde a identificação de riscos até a implementação de medidas de mitigação e o monitoramento contínuo (NIST, 2023).

#### 4.2.4.2 NIST Special Publication 1270: "A Proposal for Identifying and Managing Bias in Artificial Intelligence"

Os dois documentos da NIST, "A Proposal for Identifying and Managing Bias in Artificial Intelligence" (NIST SP 1270) e "Artificial Intelligence Risk Management Framework (AI RMF 1.0)" (NIST AI 100-1), são complementares na missão de promover o uso responsável e confiável de sistemas de IA. Enquanto o primeiro documento foca especificamente na identificação e gestão de vieses dentro dos sistemas de IA, o segundo expande essa abordagem ao criar um quadro abrangente de gestão de riscos que engloba não apenas o viés, mas também outros aspectos críticos como segurança, privacidade e responsabilidade (NIST, 2021; NIST, 2023).

O NIST SP 1270 introduz a necessidade de abordar os vieses como parte fundamental do desenvolvimento de IA confiável, destacando como essas tecnologias podem perpetuar desigualdades e como a gestão eficaz do viés é

crucial para evitar impactos negativos na sociedade. Por outro lado, o AI RMF 1.0 formaliza essa e outras preocupações dentro de um framework mais amplo, que orienta as organizações em todo o ciclo de vida da IA, desde o planejamento até a implementação, com foco em mitigar riscos e maximizar benefícios de maneira sistemática e integrada (NIST, 2021; NIST, 2023). Ambos os documentos reforçam a importância de uma abordagem colaborativa e interdisciplinar, envolvendo múltiplas partes interessadas para garantir que os sistemas de IA sejam desenvolvidos e utilizados de maneira que respeite os direitos humanos, promova a equidade e mantenha a confiança do público. Essa correlação entre a gestão de viés e a gestão de riscos como um todo sublinha a visão da NIST de que a confiança na IA depende tanto da mitigação de riscos técnicos quanto da consideração cuidadosa dos impactos sociais e éticos dessas tecnologias (NIST, 2021; NIST, 2023).

O documento NIST SP 1270 aborda o viés em sistemas de IA como um desafio crítico para o desenvolvimento de tecnologias confiáveis e responsáveis. Ele define o viés como desvios sistemáticos que podem resultar em resultados prejudiciais ou discriminatórios, muitas vezes exacerbados pela automação e pela escala em que as tecnologias de IA operam. O documento discute como esses vieses podem se manifestar em várias etapas do ciclo de vida da IA, desde a concepção e desenvolvimento até a implementação e uso final (NIST, 2021).

O NIST identifica várias formas de viés, como o viés estatístico, viés cognitivo humano, viés institucional e viés de feedback, que podem ocorrer em diferentes contextos e aplicações de IA. O documento enfatiza que, embora a erradicação total do viés não seja realista, é essencial identificar, medir e gerenciar esses vieses para mitigar seus impactos negativos. Além disso, o NIST propõe uma abordagem em três estágios para lidar com o viés, que inclui a identificação e a mitigação de vieses nas fases de pré-design, design e desenvolvimento, e implementação de sistemas de IA (NIST, 2021) .

Os vieses estatísticos, cognitivos humanos e institucionais são manifestações distintas de desvios que podem impactar significativamente a eficácia e a equidade dos sistemas de Inteligência Artificial (IA). Esses vieses, embora variados em suas origens e consequências, compartilham a característica comum de distorcer os resultados de sistemas automatizados,

comprometendo sua confiabilidade e aumentando o risco de resultados injustos ou prejudiciais (NIST, 2021).

O viés estatístico, por exemplo, refere-se a uma tendência sistemática em estimativas ou medições que resulta em valores consistentemente distorcidos, seja acima ou abaixo de seu valor verdadeiro. Este tipo de viés pode surgir sem qualquer intenção discriminatória, sendo frequentemente o resultado de erros sistemáticos na coleta ou análise de dados. Como tal, o viés estatístico exemplifica como problemas aparentemente técnicos podem ter repercussões significativas na precisão e validade das decisões automatizadas (NIST, 2021).

Por outro lado, o viés cognitivo humano envolve erros sistemáticos no pensamento humano, muitas vezes baseados em heurísticas limitadas que simplificam excessivamente a realidade. Esses vieses podem levar a julgamentos inadequados ao interpretar as saídas de sistemas de IA, especialmente quando usuários ou desenvolvedores impõem suas próprias interpretações enviesadas aos resultados gerados por algoritmos. Dessa forma, o viés cognitivo humano destaca a complexa interação entre as percepções humanas e a tecnologia, onde decisões tendenciosas podem ser amplificadas pela automação (NIST, 2021).

No contexto da Inteligência Artificial (IA), esses vieses cognitivos podem influenciar negativamente a forma como os resultados gerados por algoritmos são interpretados. Por exemplo, um desenvolvedor ou usuário de IA pode, consciente ou inconscientemente, aplicar suas próprias crenças ou preconceitos ao analisar a saída de um sistema automatizado. Isso pode resultar em decisões que não refletem corretamente os dados ou a situação real, mas sim as suposições enviesadas do ser humano. Essa situação se torna ainda mais preocupante quando a IA amplifica essas decisões tendenciosas. Como as tecnologias de IA podem operar em grande escala e tomar decisões em alta velocidade, qualquer viés humano que entre no sistema pode ser multiplicado, resultando em impactos amplos e potencialmente prejudiciais. Portanto, o viés cognitivo humano evidencia a necessidade de um cuidado especial na interação entre humanos e tecnologia, para garantir que as decisões automatizadas sejam justas e precisas (NIST, 2021).

Finalmente, o viés institucional, ou sistêmico, refere-se a práticas ou procedimentos institucionais que inconscientemente favorecem certos grupos sociais em detrimento de outros. Esse viés é particularmente insidioso, pois pode perpetuar desigualdades ao longo do tempo, refletindo normas e práticas históricas que, mesmo sem intenção explícita de discriminação, continuam a impactar negativamente certos segmentos da sociedade. Assim, o viés institucional evidencia como estruturas e políticas aparentemente neutras podem, na verdade, reforçar desigualdades existentes, tornando essencial a vigilância e a revisão contínua dessas práticas para promover uma maior equidade (NIST, 2021).

O documento também inclui um glossário abrangente de termos relacionados ao viés em IA, fornecendo definições detalhadas para ajudar pesquisadores e desenvolvedores a entender e abordar os diferentes tipos de viés. A NIST destaca a importância de desenvolver padrões e práticas comuns que possam ser aplicados em diversos setores para reduzir o viés e aumentar a confiança pública em sistemas de IA (NIST, 2021).

### 4.2.5 Responsabilidade Civil nos Estados Unidos nos casos de danos causados por sistemas de IA.

A responsabilidade civil relacionada aos sistemas de IA nos Estados Unidos é tratada atualmente sem uma legislação federal específica que aborde diretamente essa questão. Em vez disso, a responsabilidade é tratada por meio de conceitos tradicionais do direito civil, como negligência e responsabilidade por produtos defeituosos, e a avaliação dos desenvolvedores de IA é feita caso a caso, considerando se foram tomadas precauções adequadas no desenvolvimento e teste do sistema. Além disso, se o sistema de IA for considerado um produto defeituoso que não atenda aos padrões de segurança, a responsabilidade por produtos defeituosos pode ser aplicada (Kharitonova, Savina e Pagnini, 2022).

Ademais, há uma discussão sobre a "lacuna de responsabilidade", um conceito que destaca as dificuldades em atribuir responsabilidade quando sistemas de IA autônomos causam danos, especialmente quando a tomada de decisão desses sistemas não é diretamente controlada por humanos. Essa lacuna está fomentando debates sobre a necessidade de novas abordagens

regulatórias que possam abordar de forma mais eficaz os riscos específicos apresentados pelas tecnologias de IA, particularmente à medida que elas se tornam mais complexas e autônomas (Kharitonova, Savina e Pagnini, 2022).

Além disso, Kharitonova, Savina e Pagnini (2022), analisam como a responsabilidade civil é tratada em casos de danos a consumidores causados por sistemas de IA. Nesse contexto, nos casos em que sistemas de IA causam danos a consumidores nos Estados Unidos, a responsabilidade civil é geralmente tratada através de duas principais abordagens legais: a negligência e a responsabilidade por produtos defeituosos. A responsabilidade por negligência aplica-se quando o desenvolvedor ou fabricante não adota o cuidado necessário durante o desenvolvimento, teste ou implementação da IA, resultando em prejuízos ao consumidor. Para que a negligência seja comprovada, é necessário demonstrar que houve uma violação de dever e que essa violação causou diretamente o dano. Por outro lado, a responsabilidade por produtos defeituosos ocorre quando a IA, considerada um produto, apresenta defeitos que causam danos, sendo que, nesse caso, o consumidor não precisa provar negligência, apenas que o defeito existia e foi o causador do dano. A situação se complica ainda mais quando ocorrem atualizações contínuas de software, onde uma nova falha introduzida pode resultar em responsabilidade adicional para o desenvolvedor. Assim, a responsabilidade civil por danos causados por IA é tratada dentro das estruturas legais tradicionais, mas adaptada para enfrentar os desafios específicos que essas tecnologias apresentam.

### 4.2.6 Aspectos relacionados a *Algorithmic Accountability Act*

O *Algorithmic Accountability Act*[44], introduzido em 2019 e reintroduzido em 2022, é um dos principais projetos de lei focados na responsabilidade de sistemas automatizados, incluindo IA. Ele exige que as empresas realizem avaliações de impacto para identificar, mitigar e comunicar possíveis danos que seus sistemas possam causar, especialmente em relação à privacidade, segurança e justiça. Embora o projeto de lei não trate

[44] Lei de Responsabilidade Algorítmica

diretamente de penalidades, ele estabelece uma estrutura para supervisão e responsabilidade (Mokander, 2022).

Especificamente, o projeto de lei exige que as empresas realizem avaliações de impacto tanto antes quanto depois da implementação dos sistemas de IA, identificando e mitigando quaisquer riscos associados. Além disso, ele impõe requisitos de transparência, onde as empresas devem documentar e comunicar como os sistemas de IA tomam decisões, especialmente aquelas que afetam significativamente a vida dos consumidores, como decisões relacionadas ao acesso a crédito, emprego e outros serviços essenciais (Mokander, 2022).

A supervisão e a aplicação das regras deste projeto de lei seriam realizadas pela *Federal Trade Commission* (FTC), que teria o poder de impor penalidades em casos de não conformidade. Além disso, o projeto também menciona a importância de que as empresas consultem especialistas e grupos impactados durante o desenvolvimento e a implementação desses sistemas, para garantir que todos os aspectos relevantes sejam considerados. A FTC também desempenha um papel crucial na supervisão e aplicação de penalidades relacionadas ao uso de IA. A FTC pode impor penalidades, exigir a comunicação de incidentes, e supervisionar as práticas de IA para garantir que as empresas não violem os direitos dos consumidores (Mokander, 2022).

O *Algorithmic Accountability Act* reflete uma abordagem pragmática para lidar com os desafios éticos e legais associados ao uso crescente de IA e sistemas de decisão automatizados, buscando equilibrar a inovação tecnológica com a necessidade de proteger os direitos dos indivíduos e manter a confiança pública. No entanto, o projeto ainda enfrenta desafios para ser aprovado no Congresso, devido à complexidade do tema e à necessidade de apoio político significativo (Mokander, 2022).

A lei é criticada por sua aplicação limitada apenas a grandes empresas, o que pode deixar de fora um número significativo de atores menores que também utilizam essas tecnologias, criando uma lacuna regulatória. Ademais, as críticas ao ato se intensificam quando se considera sua falta de especificidade em certos aspectos cruciais. A ausência de diretrizes claras pode dificultar a aplicação da lei e permitir que empresas contornem seus requisitos, o que poderia comprometer a eficácia das proteções propostas.

Essa ambiguidade é vista como uma oportunidade perdida para criar uma estrutura regulatória mais robusta e eficaz (Mokander, 2022). Além disto, o projeto de lei aplica-se apenas às entidades sob a jurisdição da *Federal Trade Commission* (FTC), excluindo assim agências públicas, bancos, transportadoras aéreas e outras entidades importantes. Essa limitação significa que muitos sistemas de IA que poderiam ter impactos significativos na sociedade não estão sujeitos às avaliações de impacto exigidas pela lei (Gursoy, Kennedy e Kakadiaris, 2022)

Neste mesmo sentido, quanto às críticas, o *Algorithmic Accountability Act* de 2022 recebeu críticas substanciais por suas limitações em termos de alcance e eficácia regulatória. Entre as principais preocupações está a sua jurisdição restrita, que exclui várias entidades públicas e setores críticos, como bancos e agências governamentais, o que pode deixar de fora sistemas de IA com grande potencial de impacto social. Além disso, o ato não adota uma abordagem baseada em riscos, deixando de categorizar aplicações de alto risco, o que poderia exigir uma regulação mais rigorosa. A ambiguidade sobre quais entidades e sistemas são cobertos, juntamente com a prática de autoconformidade sem supervisão rigorosa da FTC, levanta dúvidas sobre a efetividade do cumprimento das normas. Por fim, a limitada disponibilidade de informações ao público impede uma verdadeira responsabilização das empresas, reduzindo a transparência e dificultando a identificação de violações. Esses pontos indicam que, apesar de ser um passo inicial importante, o projeto de lei precisa de melhorias significativas para garantir a proteção adequada dos indivíduos contra os riscos dos sistemas de decisão automatizados (Gursoy, Kennedy e Kakadiaris, 2022).

Este projeto é comparado com a AI Act da União Europeia, que adota uma abordagem mais detalhada e abrangente para a regulamentação da IA, sugerindo que os EUA poderiam aprender com alguns aspectos da legislação europeia, especialmente no que diz respeito à especificidade (delega muitas decisões detalhadas a FTC) e à aplicação em diferentes setores (grandes, médias e pequenas empresas) (Mokander, 2022; Gursoy, Kennedy e Kakadiaris, 2022).

Assim, a criação de estruturas regulatórias que equilibrem a inovação com a proteção dos direitos dos indivíduos é crucial, e a experiência europeia oferece lições valiosas que podem ser incorporadas ao modelo americano. Assim, para que o ato tenha um impacto significativo, é essencial que ele evolua, tornando-se mais específico e aplicável a uma amplitude de organizações e tecnologias (Mokander, 2022; Gursoy, Kennedy e Kakadiaris, 2022).

Por fim, existem várias regulamentações estaduais e setoriais, como a *Califórnia Consumer Privacy Act* (CCPA), concede aos residentes na Califórnia direitos sobre suas informações pessoais, impõe requisitos rigorosos para a comunicação de incidentes e a proteção de dados, que podem ser aplicáveis a sistemas de IA que manipulam dados pessoais. Esta lei também pode prever penalidades significativas para violações, exigindo que as empresas comuniquem rapidamente quaisquer incidentes de segurança e permitindo a supervisão por órgãos estaduais competentes (Pardau, 2024).

### 4.3 Legislação e Regulação da IA no Canadá

Percebe-se que o movimento de legislação e regulação da IA no Canadá iniciou-se de maneira não vinculativa através de diretrizes, códigos de ética e ferramentas de avaliação de impacto, movimento chamado de “*soft law*”, este movimento tem como objetivo de manter a evolução e desenvolvimento da IA a passos largos e não obstaculizar ou "engessar" o desenvolvimento da IA (Martin-Bariteau e Scassa, 2021).

Com o progresso natural de desenvolvimento da IA se faz necessário a migração do “*soft law*” para o “*hard law*”, leis que possam ter efeito vinculativo e que possam regulamentar de forma mais rígida e que possuam princípios jurídicos claros para enfrentar estes desafios. No entanto, reconhecendo que a regulamentação da IA é uma norma internacional emergente, há preocupações de que possa ser inflexível ou estigmatizar injustamente o campo. Isso poderia impactar negativamente as oportunidades para os canadenses e a economia do país (Martin-Bariteau e Scassa, 2021; Canada, 2024).

### 4.3.1 Pan-Canadian Artificial Intelligence Strategy

A *Pan Canadian Artificial Intelligence Strategy* tem como objetivo na segunda fase da estratégia vincular o talento de nível mundial e a capacidade de pesquisa do Canadá a programas que facilitem a comercialização e adoção dessas inovações, garantindo que as ideias e o conhecimento desenvolvidos no país sejam aplicados e transformados em produtos e serviços dentro do próprio Canadá. Desta forma, a segunda fase da estratégia centra-se em três fundamentos principais: comercialização, padrões e talento e pesquisa (Canada, 2022a).

No que tange a comercialização, o governo do Canadá apoia os Institutos Nacionais de Inteligência Artificial — Amii em Edmonton, Mila em Montreal e Instituto Vector em Toronto — para transformar pesquisas em produtos com a finalidade de gerar maior valor agregado para o país. O governo apoia a estratégia viabilizando a iniciativa com U$$ 60 milhões fornecidos no orçamento de 2021, sendo que cada instituto poderá receber até U$$ 20 milhões em financiamento ao longo de cinco anos (2021-2026) e o Cluster de inovação global do Canadá tem uma importante missão estratégica para impulsionar a economia através da inovação em setores tecnológicos de ponta. Esses clusters formam redes regionais colaborativas que reúnem empresas, universidades e centros de pesquisa, focando em áreas como tecnologia digital, inteligência artificial, manufatura avançada, indústrias de proteínas e tecnologias oceânicas (Canada, 2022a).

No segundo pilar, o governo do Canadá conta com *Standards Council of Canadá* (Conselho de Normas do Canadá) para desenvolver e promover a adoção de padrões relacionados à Inteligência Artificial. No terceiro pilar, o CIFAR (Canadian Institute for Advanced Research) está fortalecendo seus programas para atrair, reter e desenvolver talentos de pesquisa acadêmica, além de manter centros de excelência em IA. Além disto, a Digital research Alliance of Canada está fornecendo computação dedicada para pesquisadores de IA para fornecer estrutura para os pesquisadores desenvolverem seus projetos de IA com investimento econômico e de infraestrutura (Canada, 2022a).

### 4.3.2 Machine Learning for Decision Making (Diretriz sobre Tomada de Decisão Automatizada)

A **Machine Learning for Decision Making** é aplicável a departamentos (que englobam cerca de 97 instituições federais) que utilizam sistemas automatizados para automatizar, total ou parcialmente, uma decisão administrativa. Esses sistemas de decisão automatizada incluem tecnologias baseadas em IA, entre outras. A diretriz se aplica a sistemas desenvolvidos ou adquiridos após abril de 2020. No entanto, nem toda IA utilizada no serviço público federal está sujeita a essa diretriz; apenas aqueles sistemas de IA envolvidos na tomada de decisões administrativas são obrigados a cumpri-la (Mandiega, 2019).

O Canadá já implementou uma série de princípios orientadores para o uso da IA na administração e nos serviços públicos. As instituições públicas são obrigadas a incorporar princípios éticos, como privacidade e transparência, em suas aplicações de IA. A Diretriz de 2018 sobre Tomada de Decisão Automatizada (*Machine Learning for Decision-Making*) para o setor público federal estabelece as responsabilidades dessas instituições e fornece regras para ajudar a avaliar e mitigar os riscos associados à implantação de sistemas de decisão automatizada. Essa diretriz visa garantir que estes sistemas sejam usados de maneira responsável, transparente e justa, minimizando os riscos e garantindo a proteção dos direitos dos cidadãos (Mandiega, 2019; Hartmann Peixoto, 2020b).

Os princípios norteadores  para o uso da *Machine Learning for Decision Making* baseia-se de acordo com a classificação de risco e impactos da utilização destes sistemas de tomada de decisão automatizados. Existem 4 (quatro) estágios a serem considerados: leve, moderado, alto e muito alto[45].

---

[45] Avaliação de impacto Algorítmico (AIA) classifica em 4 (quatro níveis): **Leve**: **Impacto nos direitos individuais, coletivos, bem-estar, saúde, interesses econômicos ou sustentabilidade**: Os impactos são pequenos e geralmente não causam danos significativos. **Características**: Reversíveis e breves. **Moderado**:**Impacto nos direitos individuais, coletivos, bem-estar, saúde, interesses econômicos ou sustentabilidade**: Os impactos são mais significativos do que no nível leve, mas ainda gerenciáveis. **Características**: Provavelmente reversíveis e de curto prazo. **Alto**:**Impacto nos direitos individuais, coletivos, bem-estar, saúde, interesses econômicos ou sustentabilidade**: Os impactos são graves e podem causar danos substanciais.**Características**: Dificuldade de reversão e efeitos prolongados. **Muito Alto**:**Impacto nos direitos individuais, coletivos, bem-estar, saúde, interesses econômicos ou sustentabilidade**: Os impactos são extremamente graves e podem ter consequências muito sérias. **Características**: Irreversíveis e de longo prazo. O AIA

Essas classificações ajudam a avaliar e categorizar o nível de risco e impacto que diferentes ações ou tecnologias podem ter, auxiliando na tomada de decisões informadas sobre sua implementação e gestão. Deste modo, a Diretriz (Machine Learning for Decision Making) tem como intuito, dependendo do nível de impacto, colocar em prática mecanismos apropriados para testar vieses nos dados de treinamento antes da implementação, além de realizar testes e atualizações frequentes. Além disso, recomenda-se monitoramento contínuo para prevenir resultados indesejados e garantir conformidade com a legislação e os princípios diretrizes. Finalmente, o documento canadense estrutura especificações de treinamento, planos de contingência e níveis de aprovação necessários para que o sistema se torne operacional (Hartmann Peixoto, 2020b; Canada, 2023a; Canada, 2024).

A Diretriz sobre Tomada de Decisões Automatizadas no Canadá estabelece um conjunto abrangente de regras para a aplicação de sistemas automatizados na administração pública e nas avaliações relacionadas a clientes. A diretiva se aplica a todos os sistemas de decisão automatizados em produção, excluindo aqueles em fase de teste, e exige a realização de uma Análise de Impacto Algorítmico (AIA) em diferentes estágios do projeto. Inicialmente, a AIA deve ser realizada na fase de design para orientar a implementação do sistema, e, posteriormente, antes da produção para validar os resultados. As análises devem ser transparentes e publicadas no Portal do Governo Aberto, assegurando a transparência e o acesso público aos resultados (Canada, 2023a; Canada, 2024).

Além disso, a diretiva impõe a mitigação de riscos identificados durante a AIA, especialmente no que diz respeito à correção de vieses, e responsabiliza as instituições federais pela conformidade com as diretrizes estabelecidas. As instituições são incumbidas de garantir que os sistemas automatizados sejam implementados de maneira ética, transparente e responsável, minimizando riscos e protegendo os direitos dos indivíduos. Dessa forma, a diretiva busca equilibrar a inovação tecnológica com a

---

tem como objetivo identificar riscos e avaliar impactos. Além disto, a análise de impacto algorítmico deve ser concluída no início da fase de design de qualquer projeto e deve ser feito uma segunda vez antes da entrada do sistema em produção e deve ser revisado e atualizado em uma base programada após ser colocada em produção (Hartmann Peixoto, 2020; Canada, 2024).

necessidade de regulação adequada para preservar a justiça e a segurança nos processos administrativos automatizados (Canada, 2023a; Canada, 2024)

### 4.3.3 BILL C-27 (Projeto de Lei C-27)

Em 16 de junho de 2022, o Governo Canadense apresentou o projeto de Lei C-27, que, além de incluir o Consumer Privacy Protection Act (CPPA) e o Personal Information Tribunal Act (PIDPTA), introduziu o Artificial Intelligence and Data Act (AIDA). Este último representa a primeira iniciativa legislativa no Canadá destinada a regulamentar o desenvolvimento e a implementação de sistemas de IA no âmbito do setor privado (INPLP, 2023).

### 4.3.4 Artificial Intelligence and Data Act (AIDA):

Em junho de 2022, o Governo do Canadá apresentou a Lei de Inteligência Artificial e Dados (AIDA) como parte integrante do Projeto de Lei C-27, também conhecido como Lei de Implementação da Carta Digital de 2022. A AIDA é um marco significativo na implementação da Carta Digital, assegurando que os canadenses possam confiar nas tecnologias digitais que utilizam cotidianamente. Este marco regulatório estabelece que o design, desenvolvimento e uso de sistemas de IA devem ser conduzidos de maneira segura e em conformidade com os valores fundamentais dos canadenses. O objetivo desta lei não é paralisar a inovação, mas sim regular o uso desta tecnologia, que apresenta riscos de danos. O documento aborda que os sistemas de IA serão monitorados quanto aos riscos apresentados e serão resguardados pela legislação de proteção ao consumidor e direitos humanos. A AIDA é um projeto de lei que não foi promulgada (Parliament of Canada, 2022; Canada, 2023b; Canada, 2024).

A Artificial Intelligence and Data Act (AIDA) tem como contexto e propósito estabelecer uma estrutura regulatória baseada em risco, focada em sistemas de IA de "alto impacto"[46]. Este enfoque visa equilibrar o avanço tecnológico com a segurança pública e a observância de padrões éticos

[46] A versão original do AIDA continha conteúdo substantivo limitado, pois deixava a maioria dos elementos-chave do regime legal para serem definidos em uma data posterior em regulamentos (incluindo as principais obrigações de conformidade e a definição de "sistemas de alto impacto", que são o foco principal do Projeto de Lei). (WHITE & CASE, 2024; CANADÁ 2022).

rigorosos no desenvolvimento da IA. A lei foi desenvolvida por meio de consultas multidisciplinares com especialistas e aplica-se a entidades do setor privado envolvidas no *design*, desenvolvimento e implantação de IA no comércio nacional e internacional e tem como objetivo mitigar riscos de dano e viés apresentados pelo uso de sistemas de IA de "alto impacto ". Esta lei também estabelece proibições relacionadas à posse ou uso de informações pessoais obtidas ilegalmente com o propósito de projetar, desenvolver, usar ou disponibilizar para uso um sistema de inteligência artificial e à disponibilização para uso de um sistema de inteligência artificial se seu uso causar sérios danos a indivíduos.(Morgan *et al.*, 2023, Canada, 2023b; Gallagher, 2024; WHITE & CASE, 2024).

Segundo Morgan *et al.* (2023), por enquanto o Science and Economic Development Canada (ISED) caracterizam os sistemas como de alto impacto, de forma exemplificativa, como aqueles que ofereçam alta gravidade e natureza dos danos potenciais aos indivíduos, considera a abrangência do uso do sistema de IA e os impactos advindos deste, evidências de risco à saúde e segurança das pessoas, os desequilíbrios econômicos e sociais que podem causar aos indivíduos afetados, considera a dificuldade que os usuários podem encontrar ao tentar optar por não utilizar os sistemas de IA e avalia até que ponto os riscos associados aos sistemas de IA já são regulamentados por outras legislações vigentes. Os princípios que guiarão as obrigações dos sistemas de "alto impacto",de acordo com o ISED, são a supervisão humana e monitoramento humano, transparência, justiça e equidade, segurança, responsabilidade, validade e robustez. Estes princípios visam proteger os indivíduos danos[47] como um todo e assegurar o desenvolvimento da IA de forma ética, segura e responsável.

As entidades sob a AIDA são obrigadas a realizar avaliações de risco, implementar medidas de mitigação, garantir o monitoramento contínuo e divulgar publicamente o funcionamento de seus sistemas de IA. Em 2024, o Projeto de Lei C-27, que inclui a AIDA, está sob consideração do comitê na Câmara dos Comuns. O progresso e a promulgação deste projeto de lei

[47] Segundo o projeto de Lei C-27 dano significa: danos físicos ou psicológicos a um indivíduo; danos à propriedade de um indivíduo ou perda econômica para um indivíduo (CANADÁ, 2022b).

dependem da evolução do processo legislativo. O documento complementar do projeto de lei oferece insights detalhados sobre as intenções legislativas e os aspectos operacionais esperados, sublinhando a importância de uma regulamentação ágil que promova tanto a inovação responsável quanto a proteção dos direitos dos cidadãos (Morgan *et al.*, 2023; Gallagher, 2024; WHITE & CASE, 2024).

De acordo com o Bill C-27 sistema de inteligência artificial "significa um sistema tecnológico que, de forma autônoma ou parcialmente autônoma, processa dados relacionados às atividades humanas por meio do uso de algoritmo genético, uma rede neural, aprendizado de máquina ou outra técnica para gerar conteúdo ou tomar decisões, recomendações ou previsões" (Parliament of Canada, 2022). Além disto, sistemas de alto impacto "significa um sistema de inteligência artificial que atende aos critérios para um sistema de alto impacto estabelecidos em regulamentos" (Parliament of Canada, 2022). Desta forma, o Bill C-27 é focado principalmente em regulamentações do setor privado, a exclusão das instituições governamentais do âmbito de aplicação do Bill C-27 é para garantir que essas entidades sejam reguladas por um quadro legal que é mais apropriado para o setor público, evitando redundâncias e promovendo uma governança clara e eficiente. Ao excluir as instituições governamentais do escopo do Bill C-27, o governo evita a duplicação ou conflito de regulamentações que já são aplicáveis ao setor público através da Lei de Privacidade (Privacy Act) (Parliament of Canada, 2022).

O objetivo da lei é estabelecer normas uniformes em todo o Canadá para regular o comércio internacional e interprovincial de sistemas de inteligência artificial, abrangendo o design, desenvolvimento e uso desses sistemas; além de proibir comportamentos relacionados à IA que possam causar danos graves às pessoas ou prejudicar seus interesses. Por conseguinte, o Bill C-27 estabelece que o conceito de saída tendenciosa refere-se ao conteúdo ou decisões geradas por IA que discriminam injustamente indivíduos com base em motivos proibidos pela Lei Canadense de Direitos Humanos e que o conceito de dano abrange danos físicos ou psicológicos a um indivíduo; danos a propriedade de um indivíduo ou perda econômica para um indivíduo (preconceito). A atividade regulamentada refere-se a qualquer atividade realizada no comércio internacional ou

interprovincial, incluindo processar dados humanos para o desenvolvimento de sistemas de inteligência artificial, desenvolver, disponibilizar ou gerir a operação de tais sistemas. Uma pessoa é considerada responsável por um sistema de IA, incluindo sistemas de alto impacto, se estiver envolvida no seu desenvolvimento, disponibilização ou na gestão de seu funcionamento no contexto do comércio internacional ou interprovincial (Parliament of Canada, 2022).

A avaliação de sistemas de alto impacto deve ser feita pela pessoa responsável por um sistema de inteligência artificial  e deve, conforme estipulado pelos regulamentos, avaliar se o sistema é classificado como de alto impacto, e, além disso, implementar medidas para identificar, avaliar e mitigar os riscos de danos ou de vieses que possam surgir do uso desse sistema. Ademais, essa pessoa deve estabelecer medidas para monitorar tanto a conformidade com as medidas de mitigação exigidas quanto a eficácia dessas medidas, assegurando que todos os requisitos regulamentares sejam devidamente cumpridos. No Canadá, o regulamento que avalia e estabelece diretrizes para sistemas de alto impacto é parte do Bill C-27, especificamente no contexto da Lei de Implementação da Carta de Direitos Digitais e Proteção da Privacidade de Dados, também conhecida como AIDA (Parliament of Canada, 2022; Canada, 2023b).

A abordagem baseada em risco do AIDA foi projetada para alinhar-se com normas internacionais emergentes em IA, como o EU AI Act, os Princípios de IA da OCDE e o RMF do NIST dos EUA, ao mesmo tempo em que se integra às estruturas legais canadenses. A definição de IA no AIDA reflete conceitos da OCDE e do EU AI Act, garantindo interoperabilidade e facilitando o acesso das empresas canadenses aos mercados globais. No contexto de IA no Canadá, a gestão de risco, avaliação de impacto e supervisão humana são orientadas por documentos cruciais, como o Bill C-27, que estabelece diretrizes regulatórias específicas para IA, incluindo sistemas de alto impacto. Enquanto regulamentos específicos do Bill C-27 ainda não foram finalizados e implementados, existem documentos que completam essa estrutura, a *Privacy Impact Assessment* (PIA) do Canadá é adaptada para avaliar riscos de privacidade relacionados à IA, enquanto o *Cyber Security Risk Management Framework* do Canadá oferece orientações para proteger infraestruturas

digitais contra ameaças cibernéticas. Além disso, princípios internacionais, como os da OCDE e o *Risk Management Framework* (RMF) do NIST, influenciam a conformidade e a interoperabilidade global, garantindo que o Canadá alinhe suas práticas com normas internacionais para uma IA segura e responsável (Parliament of Canada, 2022; Canada, 2023b).

Neste tocante, a *Directive on Automated Decision-Making* e a *Algorithmic Impact Assessment (AIA) Tool* desempenham papéis cruciais no setor público, garantindo que os sistemas de IA sejam avaliados e monitorados quanto a impactos éticos e sociais antes de sua implementação. No setor privado, o *Privacy Impact Assessment* (PIA) é fundamental para proteger a privacidade dos indivíduos, enquanto o *Cyber Security Risk Management Framework* assegura que os sistemas de IA sejam seguros contra ameaças cibernéticas. Complementando essas ferramentas, o Bill C-27 propõe uma estrutura regulatória abrangente que aborda a governança de IA, incluindo a supervisão humana e a mitigação de riscos, reforçando o compromisso do Canadá com o uso responsável e seguro dessas tecnologias (Parliament of Canada, 2022; Stahl *et al*, 2023; Canada, 2023b; Attard-Frost, Brandusescu e Lyons, 2024).

A AIDA, parte do Bill C-27 no Canadá, estabelece um framework regulatório robusto que promove boas práticas e governança no uso da IA. A AIDA exige que desenvolvedores e operadores realizem avaliações de impacto para mitigar riscos, garantindo transparência e responsabilidade no funcionamento dos sistemas de IA. Além disso, a lei enfatiza a importância da supervisão humana para prevenir danos, incentivando a adoção de códigos de conduta e melhores práticas que assegurem a conformidade com princípios éticos e a proteção dos direitos dos indivíduos. Essa abordagem visa assegurar o uso seguro e ético da IA, alinhando o Canadá com padrões internacionais de governança tecnológica (Parliament of Canada, 2022)

A AIDA, parte do Bill C-27 no Canadá, estabelece a responsabilidade civil dos desenvolvedores e operadores de sistemas de IA, especialmente os de "alto impacto", exigindo que adotem medidas rigorosas para identificar, mitigar e monitorar riscos associados ao uso dessas tecnologias. A AIDA impõe que, em caso de falhas que causem danos e que poderiam ter sido previstas ou mitigadas, os responsáveis podem ser civilmente responsabilizados,

independentemente de negligência. Além disso, o não cumprimento das obrigações estabelecidas pela lei, como a implementação de medidas de mitigação e a comunicação de incidentes, pode resultar em penalidades severas, reforçando a necessidade de uma governança cuidadosa e a proteção dos direitos dos indivíduos no uso de IA (Parliament of Canada, 2022). Neste tocante, quanto a responsabilidade dos desenvolvedores e operadores no que tange a danos causados a consumidores "com base na legislação na lei canadense de proteção ao consumidor e direitos humanos existente, a AIDA garantiria que os sistemas de IA de alto impacto atendessem às mesmas expectativas com relação à segurança e aos direitos humanos inerentes aos consumidores" (Canada, 2023b).

A Artificial Intelligence and Data Act (AIDA), parte do Bill C-27 no Canadá, estabelece diretrizes rigorosas para a comunicação de incidentes, supervisão e penalidades relacionadas ao uso de IA. A AIDA exige que operadores de sistemas de alto impacto comuniquem rapidamente qualquer incidente que possa comprometer a segurança ou os direitos dos indivíduos, assegurando uma resposta eficaz e a mitigação de danos. Além disso, a lei impõe a necessidade de supervisão contínua para garantir que os sistemas operem de forma segura e em conformidade com as regulamentações, e prevê penalidades severas para o não cumprimento dessas obrigações, reforçando a importância da governança e da responsabilidade no uso de IA (Parliament of Canada, 2022).

Assim, a AIDA prevê penalidades que incluem multas financeiras significativas para empresas que não cumprirem os requisitos, como a realização de avaliações de impacto e a comunicação de incidentes. Além disso, a AIDA prevê sanções administrativas, como ordens de cessar e desistir ou a imposição de medidas corretivas, e, em casos mais graves, pode resultar em responsabilidade criminal para os responsáveis, incluindo processos judiciais e potencial prisão. Essas medidas visam reforçar a importância da governança ética e da segurança no uso de IA (Parliament of Canada, 2022).

E por fim, a figura do Comissário para IA e dados na AIDA terá como função educação e assistência para cumprir com a AIDA e será uma figura independente para assegurar a aplicação justa das normas, o Comissário tem a função de conformidade, fiscalização e execução da lei.  A figura do Ministro

atuará como responsável na definição das políticas e direções gerais relacionadas a IA, orientando a criação e regulamentos de lei, seu papel primordial será coordenar a política e estratégia, coordenação interministerial e a supervisão legislativa. Essas diferenças estruturais e funcionais garantem que a AIDA seja implementada de maneira abrangente, com supervisão rigorosa e políticas bem informadas para promover o uso seguro e ético da IA no Canadá. As infrações regulatórias que estão na AIDA incluem multas administrativas, sanções criminais que podem incluir a prisão e medidas corretivas que podem ser em casos extremos na proibição do uso do sistema de IA (Choudhry, Wall e Reynolds, 2023; Morgan *et al.,* 2023; Canada, 2023b)

### 4.3.5 Personal Information Protection and Electronic Documents ACT (PIPEDA):

O projeto de Lei C-27 (Bill C-27), *Digital Charter Implementation Act*[48], e que tem por objetivo atualizar as leis de privacidade do Canadá, o que inclui a Personal Information Protection and Electronic Documents Act (PIPEDA). Esta atualização é importante para a regulamentação da IA porque envolve a proteção de dados pessoais usados em sistemas de IA. No Canadá o PIPEDA regula em nível federal a coleta, uso e divulgação de informações pessoais por empresas privadas, existem legislações provinciais substancialmente semelhantes em Quebec, Alberta e British Columbia que complementam essas proteções, garantindo uma abordagem coerente e abrangente à proteção de dados pessoais no Canadá (Choudhry, Wall e Reynolds, 2023; WHITE & CASE

[48] Digital Charter Implementation Act, especificamente o Projeto de Lei C-27, abrange e atualiza a Lei de Proteção de Informações Pessoais e Documentos Eletrônicos (PIPEDA). Este projeto de lei visa modernizar a estrutura de privacidade do setor privado no Canadá e está dividido em três novas leis: 1) **Consumer Privacy Protection Act (CPPA)**: Esta lei substitui a PIPEDA, atualizando a proteção de privacidade para refletir as realidades do setor privado atual, garantindo maior controle e transparência no manejo de informações pessoais por parte das organizações. 2) **Personal Information and Data Protection Tribunal Act**: Estabelece um novo tribunal para revisar as recomendações do Comissário de Privacidade do Canadá e impor penalidades administrativas. 3) **Artificial Intelligence and Data Act (AIDA)**: Cria regras específicas para o desenvolvimento e uso responsável de sistemas de inteligência artificial, incluindo medidas para mitigar riscos de danos e vieses, e estabelece um Comissário de IA e Dados para supervisionar a conformidade e monitoramento contínuo.A Digital Charter Implementation Act, 2022, introduz essas três novas leis para reforçar a proteção da privacidade e criar um marco regulatório robusto que apoie o desenvolvimento ético e seguro de tecnologias digitais e de IA no Canada (CANADA, 2022b). Disponível em:https://ised-isde.canada.ca/site/innovation-better-canada/en/canadas-digital-charter/bill-summary-digital-charter-implementation-act-2020 .

2024).

Portanto, empresas privadas que desenvolvem sistemas de IA no Canadá devem estar cientes das obrigações impostas pelo PIPEDA, especificamente no que diz respeito ao uso de informações pessoais para treinar e desenvolver sistemas de IA, bem como à coleta e uso de dados pessoais durante as interações dos consumidores com sistemas de IA. Assim, o PIPEDA tem como foco incluir a obrigação de que empresas privadas sejam obrigadas a incluir a obtenção de consentimento dos indivíduos para o uso dos dados pessoais para assegurar que estas práticas estejam de acordo com os princípios éticos de privacidade dos indivíduos (Choudhry, Wall e Reynolds, 2023).

Portanto, a Consumer Privacy Protection Act (CPPA) substituirá a atual Lei de Proteção de Informações Pessoais e Documentos Eletrônicos (PIPEDA) o projeto de Lei C-27, também conhecido como Digital Charter Implementation Act, 2022, introduz a CPPA como parte de um esforço para atualizar e modernizar as leis de privacidade no setor privado, oferecendo uma estrutura mais robusta e atualizada em comparação com o PIPEDA (OPENPARLIAMENT.CA, 2023)

### 4.3.6 Consumer Privacy Protection Act (CPPA)

A *Consumer Privacy Protection Act* (CPPA)(Lei de Proteção à Privacidade do Consumidor) tem por objetivo, na era digital, em que os sistemas de inteligência artificial vem ganhando mercado proteger os consumidores e suas informações pessoais de possíveis abusos que esta tecnologia emergente é capaz de potencializar. Esta lei foi produzida como forma de substituir a Lei de Proteção de Informações Pessoais e Documentos Eletrônicos (PIPEDA). Essa lei tem perspectiva de ser aprovada em 2024 ou no começo de 2025 (Canada, 2022b).

Além disso a lei também exige que as organizações implementem um robusto programa de gerenciamento de privacidade, abrangendo políticas, práticas e procedimentos destinados a assegurar a conformidade com a legislação. Esta exigência visa estabelecer uma base sólida para a proteção de dados pessoais dentro das organizações. Quanto ao consentimento, o CPPA reforça o consentimento, especialmente o expresso, como a principal base

legal para o processamento de informações pessoais. Contudo, a lei também estabelece exceções que permitem a coleta, uso ou divulgação de dados sem consentimento para certas atividades comerciais padrão ou quando a organização possui um interesse legítimo, desde que cumpram condições específicas. Essas disposições visam equilibrar a necessidade de proteger os dados pessoais com a flexibilidade operacional necessária para atividades legítimas (Canada, 2022b; Kardash, Polataiko e Devir, 2023).

O CPPA contém disposições relacionadas ao processamento de dados "desidentificados" e "anonimizados", esclarecendo que informações anonimizadas estão fora do escopo da lei. O projeto de lei também confere um status especial às informações pessoais de menores. Em casos de "impacto significativo" sobre indivíduos, o CPPA exigirá que empresas expliquem como previsões, recomendações ou decisões são feitas por sistemas automatizados de tomada de decisão. Além disso, as empresas devem fornecer informações sobre o tipo e a fonte dos dados pessoais utilizados (Canada, 2022b; Kardash, Polataiko e Devir, 2023).

Os indivíduos poderão solicitar que as organizações deletem suas informações pessoais, e em certas situações, essas organizações serão obrigadas a cumprir tais solicitações. Além disso, o CPPA inclui disposições que garantem aos indivíduos direitos de portabilidade de dados, permitindo-lhes ordenar a transferência de suas informações pessoais de uma organização para outra. O não cumprimento do CPPA pode expor as organizações a multas de até C$ 25 milhões e o valor correspondente a 5% da receita bruta global do ano fiscal anterior. As organizações também podem ser expostas a penalidades monetárias administrativas de até C$ 10 milhões e o valor correspondente a 3% da receita bruta global do ano fiscal anterior (Canada, 2022b; Kardash, Polataiko e Devir, 2023).

A Artificial Intelligence and Data Act (AIDA) e o Consumer Privacy Protection Act (CPPA), ambos integrantes do Bill C-27, operam de forma complementar para assegurar a proteção dos direitos dos indivíduos no Canadá no contexto digital. Enquanto o CPPA foca na proteção de dados pessoais, estabelecendo diretrizes rigorosas sobre como as organizações devem coletar, usar e compartilhar essas informações, a AIDA assegura que os sistemas de IA que utilizam esses dados sejam desenvolvidos e operados de

maneira ética e responsável. Juntos, esses atos criam um quadro regulatório abrangente que não apenas protege a privacidade dos consumidores, mas também garante que as tecnologias de IA operem de forma a mitigar riscos e respeitar os direitos humanos, formando uma base sólida para a governança digital no país (Parliament of Canada, 2022).

As empresas no Canadá já estão sendo encorajadas a incorporar as exigências do Bill C-27, uma vez que estas entrarão em vigor nos próximos anos. Essa legislação, que visa atualizar e reforçar as leis de privacidade e proteção de dados no país, terá um impacto significativo em como as organizações utilizam a IA. As novas regras exigirão que as empresas que implementam IA sejam transparentes sobre como seus algoritmos processam e analisam os dados pessoais, garantindo que essas tecnologias respeitem os direitos dos indivíduos e operem de forma ética. Além disso, o Bill C-27 introduzirá a necessidade de explicabilidade nos sistemas de IA, o que significa que as empresas terão que ser capazes de justificar as decisões tomadas por suas ferramentas automatizadas. A antecipação dessas exigências permitirá que as empresas se adaptem mais rapidamente às novas regulamentações, evitando penalidades e fortalecendo a confiança dos consumidores na adoção de tecnologias de IA (Canada, 2022b; Kardash, Polataiko e Devir, 2023).

4.3.7 Iniciativas Provinciais:

Em nível provincial, o Canadá implementou várias legislações específicas para a proteção de informações pessoais, embora estas legislações não buscam regular diretamente a IA, no entanto podem afetar o desenvolvimento ou uso da IA no Canadá, seja no âmbito federal, provincial ou territorial. Assim, de forma exemplificativa segue a lista de leis que afetam a IA: 1) **Personal Information Protection Act, SA 2003 (Alberta)**: Esta lei regula a coleta, uso e divulgação de informações pessoais por organizações privadas em Alberta, assegurando a proteção da privacidade dos indivíduos; 2) **Personal Information Protection Act (Colúmbia Britânica)**: Similar à legislação de Alberta, esta lei aplica-se à Colúmbia Britânica, fornecendo um quadro para a proteção das informações pessoais no setor privado (WHITE & CASE, 2024; grifo nosso).

Neste mesmo sentido, temos a: 3) **Act Respecting the Protection of Personal Information in the Private Sector (Quebec)**[49] esta legislação regula a proteção das informações pessoais no setor privado em Quebec. As emendas a esta lei, que entraram em vigor em setembro de 2023, agora regulam a tomada de decisão automatizada baseada no processamento de informações pessoais. Elas exigem a divulgação da ocorrência desse processamento e o fornecimento de informações sobre os motivos e fatores que levaram a uma decisão. Esta nova obrigação é inspirada no Artigo 22 do Regulamento Geral sobre a Proteção de Dados (GDPR) da União Europeia, que aborda a tomada de decisão automatizada e a definição de perfis. Estas leis provinciais complementam a legislação federal e asseguram uma abordagem abrangente à proteção de dados pessoais e à regulamentação da IA em diferentes regiões do Canadá (Choudhry, Wall e Reynolds, 2023; WHITE & CASE, 2024; grifo nosso).

Embora a regulamentação específica de IA ainda não tenha sido amplamente adotada em nível provincial, algumas províncias como Quebec estão recomendando a adoção de leis específicas para IA e já introduziram regulamentos que afetam a tomada de decisões automatizada baseada em dados pessoais. Essas iniciativas mostram o compromisso do Canadá em alinhar suas regulações de IA com as melhores práticas internacionais, garantindo que a IA seja desenvolvida e usada de forma ética e segura,

---

[49] As disposições relacionadas à tomada de decisão automatizada (ADM) na seção 12.1 do estatuto de privacidade do setor privado recentemente reformulado de Québec, o *Ato respeitando a proteção de informações pessoais no setor privado* (o *Ato do Setor Privado* ), entrarão em vigor em setembro de 2023. Essas disposições, originalmente parte do Projeto de Lei 64 , darão aos consumidores o direito à informação e o direito à objeção quando suas informações pessoais forem usadas para tomar decisões sobre eles sem julgamento humano independente. Esta disposição se aplica a decisões de ADM que são exclusivamente automatizadas. As organizações serão obrigadas a fornecer aviso do processo de ADM no momento em que a decisão for tomada; fornecer um canal para que indivíduos enviem perguntas, comentários ou reclamações a um representante que possa revisar a decisão;permitir que as pessoas solicitem a correção das informações pessoais utilizadas na decisão; e informar o indivíduo, mediante solicitação, i) as informações pessoais utilizadas na decisão; ii) os motivos, principais fatores e parâmetros que levaram à decisão; e iii) o direito do indivíduo de corrigir as informações pessoais utilizadas na decisão.Essa mudança iminente afetará organizações que usam IA em qualquer processo de tomada de decisão totalmente automatizado que envolva informações pessoais de quaisquer indivíduos em Québec, sejam eles clientes ou funcionários. Para maior clareza, os indivíduos devem ser notificados de cada processo ADM separado para o qual uma organização usa suas informações pessoais (Choudhry, Wall e Reynolds, 2023).

protegendo os direitos dos cidadãos e promovendo a inovação tecnológica (WHITE & CASE, 2024).

# 5 Arcabouço de indicações normativas do Brasil

Neste capítulo, é presenciado uma descrição e análise dos principais meios para o desenvolvimento normativo brasileiro no que tange à regulamentação da IA.

Os temas abordados, vão desde a Lei Geral de Proteção de Dados (LGPD), de 2018, perpassa o Plano Nacional de Internet das Coisas (IoT), a Estratégia Brasileira de Inteligência Artificial (EBIA), o Plano Nacional Brasileiro de Inteligência Artificial, alguns projetos de lei e algumas outras fontes normativas ou outras legislações específicas que visam regular a aplicação da IA em diferentes contextos.

O desafio é trazer um panorama consistente o suficiente para que fique clara as semelhanças e diferenças da forma com que o Brasil está lidando com essas pautas em relação aos países já descritos.

Torna-se um tópico relevante para reconhecer o impacto da IA aos olhos do governo, dos acadêmicos e do mercado que se utiliza dessas ferramentas.

## 5.1 Marco legal da IA no Brasil

De acordo com Drummond e Carneiro (2022, p.12), "o Marco Civil da internet (Lei 12.965/2014) desempenha um papel de legislação relevante e atemporal, pois os princípios da lei conferem guias interpretativos à utilização de IA no âmbito da internet". Assim, essa legislação promove a liberdade de modelos de negócios na rede, refletindo a livre iniciativa garantida pela Constituição, isso permite que empresas desenvolvam e implementem soluções de IA, incentivando a inovação e competitividade no mercado digital, obedecendo aos limites legais. Além disso, incentivos para aprimorar a infraestrutura do ecossistema digital — como a promoção de *softwares* de código aberto, interfaces de programação de aplicações (APIs), dados estruturados de maneira aberta, interoperabilidade entre sistemas e cooperação entre agentes — são essenciais para o avanço da IA e têm sido foco de várias políticas públicas, beneficiando tanto o setor público quanto o privado (Drummond e Carneiro, 2022).

O Marco Civil também fortalece uma abordagem centrada no ser humano em relação à IA, ao priorizar o respeito aos direitos humanos como princípio central, colocando a dignidade e os valores humanos no centro das discussões sobre a inovação tecnológica; prevê também “princípios, garantias, direitos e deveres, e não esgota o tratamento do assunto” (Drummond e Carneiro, 2022) Embora a proteção de dados não seja mais o foco principal da regulação, o Marco Civil estabeleceu as bases normativas e jurisprudenciais para a criação da Lei Geral de Proteção de Dados (LGPD), que possui uma ligação direta com a IA, assegurando que o desenvolvimento tecnológico ocorra de forma ética e responsável (Brasil, 2017; Drummond e Carneiro, 2022).

## 5.2 Plano Nacional Brasileiro de Inteligência Artificial

O Presidente da República Luiz Inácio Lula da Silva e a ministra da gestão em exercício, Cristina Mori, participaram da abertura da 5a Conferência Nacional de Ciência, Tecnologia e Inovação (5CNCTI), em 30/07/2024, e apresentaram o Novo Plano Brasileiro de Inteligência Artificial. O Plano consiste em tornar o Brasil um modelo global de eficiência e inovação no uso de IA no setor público. A previsão é que sejam investidos R$ 1,76 bilhão no terceiro eixo e para que isto aconteça "serão desenvolvidas soluções que melhorem significativamente a oferta e satisfação das pessoas com os serviços, com impacto no desenvolvimento e inclusão social" (Brasil, 2024a).

Os cinco eixos estruturantes do Plano Nacional de IA no Brasil, desenvolvido em 2022, são: 1) Infraestrutura e desenvolvimento de IA; 2) Difusão, formação e capacitação em IA; 3) IA para Melhoria dos serviços públicos; 4) IA para Inovação Empresarial; 5) Apoio ao Processo Regulatório e de Governança da IA. A IA para o bem de todos na visão do Brasil tem como foco a IA centrada no ser humano e acessível a todos; orientada à superação de desafios sociais, ambientais e econômicos, fundamentada no direito e na soberania nacional; transparente, rastreável e responsável e cooperativa globalmente em bases justas e mutuamente benéficas. O objetivo do plano é centrado no desenvolvimento da IA orientada à solução dos grandes desafios nacionais, sociais, econômicos, ambientais e culturais, de forma a garantir a segurança e os direitos individuais e coletivos, a inclusão social, a defesa da

democracia, a integridade da informação, a proteção do trabalho e dos trabalhadores, soberania nacional e o desenvolvimento econômico sustentável da nação (Brasil, 2024b).

Assim, o novo plano do governo estabelece a criação de um Núcleo de IA, coordenado pelo Ministério da Gestão e da Inovação (MGI), reunindo órgãos como o MCTI, Enap, Universidade de Brasília, Serpro, Dataprev e Finep, com o objetivo de fomentar o uso de inteligência artificial no setor público. Entre as principais iniciativas está a criação de uma Plataforma de IA, voltada para o desenvolvimento, treinamento e execução de modelos de IA em larga escala até 2026, com um orçamento previsto de R$ 25 milhões. O núcleo também se dedicará a identificar e estruturar projetos estratégicos de IA em 10 áreas prioritárias do governo, com a meta de desenvolver 25 projetos de alto impacto até 2026. Além disso, o plano inclui a capacitação de cerca de 115 mil servidores públicos, representando 20% do total de servidores ativos, com um investimento de R$ 7,5 milhões (Brasil, 2024a).

Além disto, o Plano Brasileiro de Inteligência Artificial (PBIA) inclui iniciativas significativas no âmbito da Infraestrutura Nacional de Dados (IND), como a implementação de uma política de governança de dados nos órgãos federais e a catalogação de 2.000 conjuntos de dados até 2027. Também prevê o fortalecimento do programa Conecta GOV.BR, visando economizar R$ 6 bilhões até 2026 ao simplificar a troca de informações entre sistemas governamentais, eliminando a necessidade de reapresentação de documentos pelos cidadãos. Desta maneira, o plano propõe a criação de um ecossistema robusto de dados públicos em uma nuvem soberana, com um investimento potencial de R$ 1 bilhão, para garantir a autonomia tecnológica, a segurança e a privacidade das informações. A Portaria nº 5.950, publicada em 2023, define diretrizes para contratações de softwares e serviços de computação em nuvem, destacando a importância da soberania dos dados e a proteção das informações sigilosas no âmbito do governo (Brasil, 2024a).

Diversos projetos estão sendo financiados por este plano para fomentar o desenvolvimento da IA no Brasil. Cita-se como exemplo sistemas de IA para a previsão de AVC e cardiopatias; projeto do cálculo bovino por câmera 3D; mapeamento para quantificação do estoque florestal do bioma Amazônico; otimização do sistema financeiro de habitação com IA; sistema de gestão

presente para evitar abandono de alunos e evasão escolar; sistemas de IA para o desenvolvimento de habilidades matemáticas para alunos do primeiro ao quinto ano e o fiscaliza da Receita Federal para otimizar julgamentos dos processos administrativos (Brasil, 2024b).

### 5.3 Estratégia Brasileira para a Transformação Digital (Portaria MCTI n. 842/2017)

A Estratégia Brasileira para a Transformação Digital (e-Digital) foi elaborada por um grupo de trabalho envolvendo diversos órgãos do governo, em colaboração com representações setoriais e com a sociedade civil, compondo visão de futuro e propondo iniciativas estratégicas**.** O E-digital procurou estar em visão congruente com os Objetivos de Desenvolvimento Sustentável da Agenda 2030 das Nações Unidas. O Brasil ocupa a posição 80 do CGI (Global Competitiveness Index), dentre 137 países que este índice compara. Depreende-se que esta posição está muito aquém da magnitude de potencial que o Brasil possui. Por isto, a agenda do e-Digital é diminuir esta posição no ranking mundial do Brasil ao longo dos próximos 5 (cinco) anos (Brasil, 2017).

O E-Digital foi conceituado com base em dois grupos temáticos: eixo de transformação digital e eixos habilitadores. Os eixos habilitadores visam criar um ambiente propício para o desenvolvimento da transformação digital sendo assim são eles: infraestrutura e acesso às TICs; pesquisa, desenvolvimento e inovação; confiança no ambiente digital; educação e capacitação profissional; dimensão internacional. Os eixos de transformação digital são economia baseada em dados; um mundo de dispositivos conectados, novos modelos de negócio e cidadania e governo (Brasil, 2017).

Um enfoque dado por esta portaria é a preocupação com uma realidade no Brasil de falta de conexão em muitos municípios brasileiros e apresenta um plano de investimento para que todos os municípios sejam atendidos com redes de transporte de alta capacidade; todos os municípios tenham atendimento banda larga móvel; grande parte da população brasileira coberta com redes de acesso de banda larga fixa com ampliação da oferta de redes de acesso de fibra ótica; áreas remotas e de difícil acesso sejam atendidas por infraestrutura de banda larga, ampla disseminação de redes de Wi-Fi em locais

públicos e instituições de pesquisa, educação, saúde e segurança sejam integradas por redes de alta velocidade. Deste modo, se o Brasil quiser ter um protagonismo internacional em relação ao desenvolvimento de novas tecnologias emergentes deverá começar por uma conexão de internet inclusiva no seu território (Brasil, 2017).

Além da internet móvel e de banda larga fixa ser escassa fora dos grandes centros, observou-se que a formação técnica e capacitação profissional são desafios de competitividade internacional, além disso 60% das infraestruturas de pesquisa brasileira declaram que o valor total de seus equipamentos e instalações não supera a faixa de R$ 500 mil, limitando o protagonismo e competitividade do brasil na produção de riqueza no setor da tecnologia. O número de artigos de pesquisadores brasileiros publicados em periódicos internacionais teve aumento de 88%, deste modo, o país manteve a 14 posição do ranking de produção científica mundial (Brasil, 2017).

De acordo com De Oliveira (2022) a implementação da tecnologia 5G começou em 2018 nos Estados Unidos, China e Coreia do Sul, sendo que no Brasil o sinal 5G começou em 5 de julho de 2022, em Brasília, ocorre que com o uso de satélites mais potentes e com o aumento da cobertura da internet, esta tecnologia garantirá um cenário de maior aumento da conectividade e como consequência impactará no desenvolvimento de novos projetos de inovação, aumentando assim os ganhos do PIB do Brasil em R$ 104 bilhões. Assim, para aumentar os benefícios da utilização da tecnologia 5G é necessário o investimento em infraestrutura para aumentar a cobertura e conectividade nos estados e municípios de maneira igualitária.

Tendo em vista os diagnósticos apresentados, a Estratégia Brasileira para a Transformação Digital, com foco no estímulo à PD&I (Pesquisa, Desenvolvimento e Inovação) no setor de TICs, deve buscar a otimização de políticas para expandir significativamente o investimento privado em PD&I, melhorar a competitividade da economia, gerar empregos de alto valor agregado e promover o desenvolvimento social. A estratégia também deve ampliar o papel das políticas públicas pelo lado da demanda, utilizando encomendas governamentais para incentivar a inovação em TICs e tecnologias correlatas, além de integrar instituições de pesquisa em todas as regiões do país a redes de alta velocidade para fomentar o intercâmbio científico e

tecnológico (Brasil, 2017).

Ademais, é necessário estabelecer um roteiro tecnológico com metas de investimento de longo prazo para atender às demandas de ciber infraestrutura de empresas intensivas em TICs e apoiar projetos de ciência, tecnologia e inovação no setor. Por fim, deve-se estimular a formação de profissionais capacitados para enfrentar os desafios das tecnologias de fronteira, como Big Data, manufatura 4.0, inteligência artificial e robótica, promovendo um ambiente de negócios juridicamente seguro e favorável à interação entre universidades, centros de pesquisa e empresas, garantindo assim novos investimentos em PD&I no setor de TICs (Brasil, 2017).

Por fim, além dos indicadores tradicionais de PD&I utilizados para avaliar o desempenho global da economia nacional—como o nível de investimento em relação ao PIB, o percentual de investimento setorial em relação ao total empresarial e o grau de formação técnica—é fundamental que o monitoramento e a avaliação das ações em PD&I, no âmbito da Estratégia Brasileira para a Transformação Digital, sejam orientados por estudos de caso específicos, abrangendo tanto empresas individuais quanto setores específicos. Adicionalmente, é imperativo que os indicadores adotados estejam fundamentados em *benchmarks* internacionais consolidados, especialmente aqueles de países que lograram estimular a consolidação dos principais atores nacionais no setor de TICs, como é o caso das nações asiáticas (Coreia do Sul, China e Japão), dos países nórdicos (Finlândia e Suécia), e de outras nações com indicadores de destaque, como Israel. Até então, no e-digital não havia  a vigência da Lei de Proteção de Dados no Brasil, e as estratégias previam a aprovação de uma lei específica que foi feita e aprovada em 2018. Outra questão bastante debatida no e-digital foi a questão de das "implicações jurídicas e éticas de aplicação de IA, Internet das Coisas e outras áreas da fronteira tecnológica" (Brasil, 2017).

Outro marco importante regulatório na transformação Digital do Brasil é o decreto 10.332/2020, que trata sobre a Estratégia de Governo Digital de 2020-2022 nos setores dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. O decreto fixou objetivos concretos para a concretização da interoperabilidade dos sistemas federais, desenvolvimento de pesquisas, parcerias com instituições de nível superior,

setor privado e terceiro setor, iniciativas com *blockchain* e políticas públicas relacionadas ao tema de segurança cibernética. Um dos objetivos deste decreto é a "implementação de recursos de IA em no mínimo 12 (doze) serviços públicos federais até 2022". (Brasil, 2017; Drummond e Carneiro, 2022).

Assim, apesar dos avanços normativos do e-digital e Decreto 10.332/2020, a OCDE preparou relatório sobre o Brasil Digital e suas implicações no setor público e fez as seguintes considerações:

> Em 2018, a OCDE preparou relatório intitulado “Digital Government Review of Brazil: Towards the Digital Transformation of the Public Sector”, no qual recomendou ao governo brasileiro a adoção de políticas públicas sobre governança digital. Destaca-se, nesse particular, a recomendação de desenvolvimento a médio prazo de um plano de ação sobre a utilização de tecnologias emergentes para melhorar a prestação dos serviços públicos – especialmente AI. A OCDE consignou que o plano de ação de AI poderia ser um documento programático complementar ao e-Digital, e poderia versar, por exemplo, sobre (i) o desenvolvimento de mecanismos de transparência e de standards éticos para permitir uma adoção responsável de tecnologias emergentes pelo setor público e (ii) o estabelecimento de uma força-tarefa para orientar a decisão do governo sobre a aplicação de AI em áreas e serviços específicos, incluindo representantes de todos os níveis de governo e da academia e do setor privado (Drummond e Carneiro, 2022; Brasil, 2017)

## 5.4 Plano de Dados Abertos do Poder Executivo Federal (Decreto n. 8.777/2016, Decreto n. 9.903/2019 e Resolução CGINDA n. 3/2017)

O Decreto definiu como "dados abertos", no artigo 2, inciso III, aqueles dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte. (Brasil, 2016; Drummond e Carneiro, 2022).

O Plano de Dados Abertos do Poder Executivo federal tem por objetivo promover a transparência pública, fomentar o controle social e incentivar a inovação por meio da disponibilização e uso dos dados públicos. Essa política visa garantir que os dados produzidos ou sob a guarda do governo federal

sejam acessíveis ao público de maneira aberta, padronizada e estruturada, permitindo que cidadãos, empresas, pesquisadores e outras entidades possam utilizar essas informações para diferentes fins, como o desenvolvimento de novos serviços, a realização de análises e a promoção da participação cidadã na gestão pública. Além disso, a política busca melhorar a eficiência e a eficácia das ações governamentais, assegurando que a informação seja tratada como um ativo estratégico na formulação de políticas públicas e na prestação de serviços à sociedade. (Brasil, 2016; Drummond e Carneiro, 2022).

Segundo o Tribunal de Contas da União existem 5 (cinco) fortes motivos para a abertura de dados na administração pública que são: porque a sociedade exige transparência na gestão pública; porque a própria sociedade pode contribuir com serviços inovadores ao cidadão; porque ajuda a aprimorar a qualidade dos dados governamentais; para viabilizar novos negócios[50]; porque é obrigatório por lei[51]:

> No Portal Brasileiro de Dados Abertos estão disponíveis para download em formato aberto diversos conjuntos de dados de relevante interesse público, tais como dados do orçamento federal; de convênios e contratos de repasse celebrados com a União; de compras públicas do governo federal; do Produto Interno Bruto (PIB); de prestação de contas das campanhas eleitorais, entre outros (TCU, 2015).

A relação entre a IA e o decreto de políticas de dados abertos do Poder Executivo Federal é bastante significativa, pois a IA depende fortemente de grandes volumes de dados para funcionar de maneira eficaz. O decreto que

---

[50] Dados de previsão do tempo providos por serviços meteorológicos públicos também possibilitam o desenvolvimento de novos negócios. A empresa americana Climate Corporation combinou mais de 30 anos de dados climáticos, 60 anos de dados sobre a produção das safras e múltiplos terabytes de informação em tipos de solo obtidos de fontes de dados públicos para oferecer serviços de consultoria a agricultores. Serviços de informações sobre o clima, a exemplo do Weather underground (www. wunderground.com) e do próprio Weather Channel (www.weather.com) nasceram a partir de dados climáticos abertos. Percebe-se que o setor privado pode fazer uso de dados abertos governamentais para gerar produtos e serviços que são comercializados à população e que até então eram inexistentes. Dessa forma, pode-se dizer que a disponibilização de dados públicos em formato aberto pelo governo potencializa um retorno positivo pois, ao serem criados novos negócios, tem-se a geração de novos empregos e, por consequência, o aumento de receita pública mediante o recolhimento de tributos (TCU, 2015).

[51] Há anos o Brasil vem gerando um arcabouço normativo direcionado à promoção da transparência e da participação social na gestão pública, abrangendo tanto diplomas legais quanto infralegais, tais como: lei complementar 101/2000 (Lei de Responsabilidade Fiscal-LRF); Lei Complementar 131/2009 (Lei da Transparência); Lei 12.527/2011 (Lei de Acesso à informação); Instrução Normativa SLTI/MP- 4/2012 que institui a Infraestrutura Nacional de Dados Abertos (Inda); Decreto 8.243/2014 que institui a Política nacional de Participação Social (TCU, 2015).

institui a política de dados abertos do Poder Executivo estabelece diretrizes para a disponibilização de dados públicos de maneira acessível, estruturada e padronizada, o que cria um ambiente propício para o desenvolvimento e aplicação de tecnologias de IA (Drummond e Carneiro, 2022).

Com o acesso a dados abertos, modelos de IA podem ser treinados para analisar informações, identificar padrões, prever tendências e oferecer soluções inovadoras em diversos setores, como saúde, educação, segurança pública, e administração pública. Além disso, a transparência e a disponibilidade dos dados promovidos pelo decreto permitem que as aplicações de IA sejam mais precisas e confiáveis, pois se baseiam em informações verificáveis e acessíveis a todos. Resta clara, a possibilidade da utilização dos dados abertos para o desenvolvimento de aplicações de IA. De acordo com o artigo 5 do Decreto a gestão da política de Dados Abertos do Poder Executivo federal será coordenada pela Controladoria-Geral da União, por meio da Infraestrutura nacional de Dados Abertos-INDA (Brasil, 2016; Drummond e Carneiro, 2022)

### 5.5 Lei Geral de Proteção de Dados (LGPD) - lei 13.709/2018

De acordo com Drummond e Carneiro (2022, p.13) um sistema de IA depende de três componentes para seu funcionamento: um *hardware*, um *software* e uma quantidade de dados, como descrito no Quadro 2.

**Quadro 2**: Três componentes básicos para o funcionamento de um sistema de IA.

| Componentes básicos de um sistema de Inteligência Artificial | Descrição |
|---|---|
| Hardware | Capacidade de processamento |
| Software | Executa as instruções (algoritmo, ML, deep learning) |
| Quantidade de dados | Informações pelas quais o software atua |

Fonte: produzido a partir de Drummond e Carneiro (2022, p.13).

A LGPD entra em cena, no cenário da proteção dos dados pessoais, evitando seu uso indevido, inclusive em sistemas de IA. A importância da Lei

está no tratamento dos dados pessoais de pessoa física, de modo a respeitar os direitos fundamentais de liberdade, privacidade e livre desenvolvimento da personalidade da pessoa natural.

Para De Oliveira (2022. p 147) a LGPD tornou-se relevante para inserir o Brasil no rol dos países que possuem uma legislação voltada à "proteção dos dados pessoais" dos indivíduos, esta lei se torna imprescindível para garantir o desenvolvimento de sistemas de IA éticos, responsáveis e confiáveis, uma vez que "estes sistemas são baseados no processamento e na transferência de dados pessoais mediante o consentimento livre e informado". Esta legislação permitiu que o Brasil recebesse maior investimento e fizesse bons negócios, pois a proteção de dados é uma tendência legislativa consolidada em âmbito internacional.

A LGPD estabelece o direito do cidadão de solicitar uma revisão e explicação de decisões automatizadas que lhe afetam, mas inicialmente a lei dispunha claramente que a revisão deveria ser feita por um humano, no entanto, após aprovação da lei, com veto presidencial, houve a remoção da palavra "humano", que fique claro: isto não obsta que seja ainda realizada por intervenção humana (Drummond e Carneiro, 2022).

A Autoridade Nacional de Proteção de Dados (ANPD), órgão regulador do tratamento dos dados pessoais, deverá indicar quando de fato é melhor a revisão por ser humano e quando poderá ser feito por meios tecnológicos. Sob um olhar comparativo da legislação de proteção de dados, o Regulamento Europeu de Proteção de Dados (GDPR) prevê o direito de requerer a intervenção humana e o direito de contestar uma decisão automatizada, a LGPD por sua vez pode não obrigar expressamente a intervenção humana, mas claramente consagra um direito de requerer revisão. Um ponto em comum na LGPD e GDPR é o princípio da transparência. Assim, no Brasil, de acordo com o artigo 20, § 1, existe a obrigação de informar os critérios e procedimentos utilizados em decisões automatizadas, por mais que não haja o termo expresso de "explicação" na legislação (Souza, Perrone e Magrani, 2021; Drummond e Carneiro, 2022).

Neste mesmo sentido, a Oliveira (2022, p.147) salienta que no artigo 20 da LGPD "garante o direito à revisão de decisões automatizadas sem

interferência humana, sendo um importante dispositivo no âmbito do debate sobre inteligência artificial e que se alinha, parcialmente, ao art 22 do GDPR, considerado um dos artigos mais citados nos tribunais europeus nos anos 2021 e 2022". Assim, a LGPD garante aos titulares dos dados a possibilidade de fiscalizar os dados pessoais e solicitar esclarecimentos sobre a utilização destes em sistemas automatizados de decisões, e se o controlador/operador não der tais informações, o titular dos dados poderá recorrer Autoridade Nacional de Proteção de Dados (ANPD) que poderá auditar a empresa e verificar o processamento de dados pessoais para a automatização de dados e a conformidade da empresa com a lei vigente.

Outra questão crucial sobre IA e LGPD é que muitas aplicações de IA dependem de dados pessoais protegidos pela lei, assim a LGPD modula as possibilidades de uso desses dados e tem por função de evitar que estes dados sejam usados indiscriminadamente na construção de perfis de comportamentos dos indivíduos e da análise de dados pessoais sensíveis (orientação sexual, religião, partido político etc) a partir da análise preditiva de dados pessoais. Assim, entende-se que a LGPD pode ser usada para prevenir danos dos sistemas de IA aos indivíduos, através do uso de ferramentas como a anonimização não reversível e pela elaboração da arquitetura de aplicações de IA norteada pela privacidade e proteção dos dados, conhecida como "*privacy by design"* (Drummond e Carneiro, 2022).

### 5.6 Plano Nacional de Internet das Coisas (IoT) (Decreto n. 9.854/2019)

A indústria dos IoT movimenta trilhões de dólares, está presente na forma de dispositivos conectados com capacidade computacional de processamento e armazenamento de dados na agricultura, pecuária, segurança, saúde e nos lares dos cidadãos. Com a proliferação de novos dispositivos conectados à Internet aptos a armazenar, coletar e tratar uma significativa quantidade de dados, tem sido recorrente a discussão sobre os usos legítimos de dados e sobre as vulnerabilidades das bases de dados geradas, por isto a necessidade do Plano Nacional de Internet das Coisas (JOTA, 2018; Decreto n. 9.854/2019; Drummond e Carneiro, 2022).

Assim, o assunto vem sendo debatido desde 2014, antes mesmo do desenvolvimento da EBIA, pois a junção de IoT e IA (chamada AIoT) possibilita operações poderosas de homem-máquina melhorando a gestão e análise de dados. Nos dispositivos de AIoT, a IA é integrada em diversos componentes da infraestrutura, como programas e sistemas, todos interconectados por meio de redes de Internet das Coisas (IoT). Para garantir que esses componentes — sejam eles de hardware, software ou plataformas — possam operar de maneira eficiente e comunicar-se de forma integrada, utilizam-se interfaces de programação de aplicativos (APIs)[52]. Dessa forma, todo o sistema funciona de maneira coesa, sem a necessidade de intervenção direta do usuário final (Drummond e Carneiro, 2022; TechTarget; 2023).

Nesse âmbito, os objetivos do Plano são: melhorar a qualidade de vida das pessoas e promover ganhos de eficiência nos serviços; promover a capacitação profissional relacionada ao desenvolvimento de aplicação de IoT e geração de emprego na economia digital; incrementar a produtividade e fomentar a competitividade das empresas brasileiras desenvolvedoras de IoT; buscar parcerias com os setores público e privado para a implementação da IoT e aumentar a integração do País no cenário internacional, por meio da participação em fóruns de padronização, da cooperação internacional em pesquisa, desenvolvimento e inovação e da internacionalização de soluções de IoT desenvolvida no País. De acordo com este decreto, os principais ambientes priorizados para aplicações de soluções de IoT são os ambientes de saúde, cidades, indústria e rural (Brasil, 2018; Brasil, 2019).

No artigo 7 do plano fica ressaltado que a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e internet das Coisas- Câmara IoT, é órgão de assessoramento destinado a acompanhar a implementação do Plano Nacional de Internet das Coisas; promover e fomentar parcerias entre entidades públicas e privadas; apoiar e propor projetos e atuar conjuntamente com órgãos e entidades

[52] APIs, ou **Interfaces de Programação de Aplicativos** (do inglês *Application Programming Interfaces*), são conjuntos de protocolos, ferramentas e definições que permitem a comunicação entre diferentes softwares ou sistemas. Elas atuam como intermediárias que facilitam a interação entre aplicações, permitindo que um software solicite e utilize funcionalidades ou dados de outro de maneira padronizada e segura (Drummond e Carneiro, 2022).

públicas para estimular o uso e o desenvolvimento de soluções de IoT. (Brasil, 2018; Brasil, 2019; ).

### 5.7 Estratégia Brasileira de Inteligência Artificial (EBIA)

A Estratégia Brasileira de Inteligência Artificial (EBIA) instituída pela portaria MCTI n. 4617, de 6 de abril de 2021 e alterada pela portaria MCTI n. 4.979, de 13 de julho de 2021, foi criada com a intenção de nortear o desenvolvimento responsável e ético da IA no Brasil. A EBIA está alinhada com as diretrizes da OCDE sendo estas "o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar; valores centrados no ser humano e na equidade; transparência e explicabilidade; robustez e proteção e; responsabilização ou prestação de contas (accountability)" (Brasil, 2021; Divino, 2022). Em De Oliveira (2022, p. 149), há uma espécie de crítica ao EBIA, pois é um documento que possui mais características de uma carta de intenções do que de uma estratégia nacional" (De Oliveira, 2022, p.149).

A "EBIA propõe o mapeamento de barreiras legais e regulatórias ao desenvolvimento de IA no Brasil, identificando aspectos da legislação brasileira que possam requerer atualizações" (Drummond e Carneiro, 2022, p. 28). Ainda que a EBIA não tenha aprofundado nos seus objetivos propostos, ela aponta caminhos e fornece caminhos para a discussão de uma IA ética nos sistemas de IA, e toca em questões de temas importantes, discutidos internacionalmente, como viés algorítmico, perda de postos de trabalho e problemas de implementação em projetos de segurança pública (Drummond e Carneiro, 2022).

Os principais objetivos, ações consideradas estratégicas para a consecução de resultados éticos e jurídicos, da EBIA são "contribuir para a elaboração de princípios éticos para o desenvolvimento e uso de IA responsáveis"; "promover investimentos sustentados em pesquisa e desenvolvimento em IA" (não se sabe de onde virão os recursos para o financiamento de projetos) ; "remover barreiras à inovação em IA"; "mapear as barreiras regulatórias para o desenvolvimento da IA"; "capacitar e formar profissionais para o ecossistema da IA"; "estimular a inovação e o desenvolvimento da IA brasileira em ambiente internacional" e "promover ambiente de cooperação entre entes públicos e privados, a indústria e os

centros de pesquisa para o desenvolvimento da Inteligência Artificial" (não se sabe como as mencionadas parcerias ocorrerão se existirão incentivos fiscais para parcerias público-privada); "elaboração de políticas de controle de qualidade de dados para treinamento dos dados" (não se sabe de quem será a competência para elaborar a política e de onde sairá o recurso); "promover abordagens inovadoras para a supervisão regulatória *sandboxes*[53] e *hubs* regulatórios" (sandbox regulatório é disciplinado na Lei Complementar de Startups 182/2021 para encorajar modelos de negócio inovadores por meio de testes em ambiente controlado) (De Oliveira, 2022; Brasil, 2021; Divino, 2022).

A estratégia possui 9 eixos temáticos[54] que caracterizam os pilares do documento no qual apresenta um diagnóstico da situação atual da IA no mundo e Brasil; destaca os desafios a serem enfrentados, oferece visão de futuro e apresenta ações estratégicas para alcançar esta visão de futuro (De Oliveira, 2022; Brasil, 2021; Divino, 2022).

De acordo com Divino (2022) os principais eixos da estratégia brasileira de IA se concentram no eixo da regulação e uso ético e governança de IA. O objetivo do eixo é o desenvolvimento de princípios éticos para a estruturação de uma IA responsável. As principais ações estratégicas deste eixo são:

> "estimular a produção de IA ética financiando projetos de pesquisa com cunho ético, principalmente nos campos de equidade/não discriminação (fairness), responsabilidade/prestação de contas (accountability) e transparência (transparência), estimular parcerias com empresas que estejam desenvolvendo e pesquisando soluções na area de IA ética; estabelecer que em licitações haja requisitos técnicos previstos para que as empresas ofereçam soluções com uma IA ética incorporada ao produto; estabelecer discussões em ambientes variados e multissetoriais, atualização legislativa de acordo com o desenvolvimento digital e mapeamento de barreiras legais e

---

[53] Já o programa da CVM de sandbox está descrito na Resolução n. 29 da CVM e na Portaria CVM/PTE n. 75 de 2020, sendo que o objetivo é incentivar o desenvolvimento de soluções inovadoras no mercado de capitais. A primeira edição (Ciclo 1) ocorreu em 2021 e contou com 33 propostas, sendo que apenas três foram selecionadas pela CVM. Não se pode olvidar que a Vórtx QR Tokenizadora foi uma das empresas escolhidas, sendo que, em 2022, realizou uma importante parceria com o Banco Itaú para “tokenizar”, pela primeira vez, debêntures negociáveis (Gusson, 2022). Dessa forma, já podemos observar as novas soluções tecnológicas apresentadas por essas empresas e a celebração de parcerias com as instituições financeira (De Oliveira, 2022)

[54] Os nove eixos temáticos são: legislação, regulação e uso ético; governança de IA, aspectos internacionais (eixos transversais); qualificações para um futuro digital; força de trabalho e capacitação; pesquisa, desenvolvimento, inovação e empreendedorismo, aplicação no poder público e segurança pública (eixos verticais)(Brasil, 2021).

> regulatórias no desenvolvimento da IA no Brasil (BRASIL, 2021)

Neste mesmo sentido do parágrafo anterior as outras ações estratégicas usadas para o desenvolvimento da IA são: estimular a transparência e divulgação responsável dos sistemas de IA de acordo com os valores democráticos e de direitos humanos; desenvolver técnicas para captar e tratar o risco de viés algorítmico; elaboração de políticas de controle de qualidade de dados para treinamento de IA; criar políticas e parâmetros para que haja a intervenção humana em contextos de IA em que os resultados dos sistemas de IA cause sérios danos ao indivíduo; incentivo de políticas de revisão apropriados em organizações públicas e privadas e por órgão públicos; criar e implementar melhores práticas ou códigos de conduta com relação a coleta, implantação e uso de dados, incentivando as organizações a melhorar sua rastreabilidade; promover supervisão regulatória como por exemplo de sandboxes e hubs regulatórios (Brasil, 2021).

Existem 3 (três) iniciativas priorizadas em relação ao desenvolvimento ético da IA no Brasil de acordo com o Ministério da Ciência, Tecnologia e Inovação que são: desenvolvimento de um *Framework*-Recomendação para uma IA ética, com mitigação de riscos e vieses; desenvolvimento de um repositório dinâmico de legislação e revisão de modelo regulatório; propor diretrizes e políticas com as melhores práticas para o desenvolvimento da IA ética (Brasil, 2021). Enquanto o governo exerce o papel da governança, as empresas privadas são responsáveis por desenvolver projetos que respeitem os direitos individuais e coletivos (Cortiz, 2020; Brasil, 2021; Divino, 2022).

De acordo com De Oliveira (2022, p. 149) a EBIA indica que o Brasil deve desenvolver um modelo de gestão de risco para a automatização de decisões:

> Nesse sentido, em setores que possuem riscos elevados, deve haver a participação humana (Ebia, 2021, p.19). Consequentemente, pode-se notar que o Brasil seguiu as diretrizes presentes na Resolução de 20 de outubro de 2020 do Parlamento Europeu (European Parliament, 2020), que estabelece recomendações para a comissão sobre o regime de responsabilidade civil para a inteligência artificial (2020/2014 INL), pois, segundo tal norma comunitária, a responsabilidade civil deverá seguir o modelo de níveis de riscos dos sistemas de inteligência artificial e as respectivas medidas a serem adotadas para minimizar o risco.

O eixo de governança de Inteligência Artificial tem por função estruturar um ecossistema de governança do uso da IA no setor público e privado de modo a observar o uso ético de IA, estruturação de modo a permitir e prevenir e eliminar vieses, algoritmos explicáveis, escolha de dados utilizados nos sistemas de IA, responsabilidade e prestação de contas. As principais ações estratégicas de governança, a título de rol exemplificativo, se apresentam de forma diversa para promover e incentivar o compartilhamento de dados, obedecendo a Lei geral de proteção de Dados (LGPD); promover o desenvolvimento de padrões voluntários e consensuais para gerenciar os riscos associados a IA; estimular que as organizações criem conselhos de revisão de dados ou comitês de ética em relação a IA; criar um observatório de IA no Brasil em contato com outros observatórios internacionais (iniciativas priorizadas); promover o estímulo de uso de conjunto de dados representativos para treinar e testar modelos; melhorar a qualidade dos dados disponíveis; estimular a divulgação do código fonte aberto capaz de verificar tendências discriminatórias no conjunto de dados e nos modelos de aprendizado de máquina; desenvolver diretrizes para elaboração de Relatório de Impacto de Proteção de Dados (RIPD); estimular diálogo social com participação multissetorial e alavancar e incentivar práticas de accountability relacionadas à IA nas organizações (Brasil, 2021).

De acordo com o EBIA (2021) a transparência e explicabilidade são princípios a serem seguidos no âmbito de todo o processo de elaboração, evidência que tem conhecimento da dificuldade de interpretação e explicação de decisões tomadas por pelos sistemas de IA, em razão do aprendizado da máquina, principalmente no que tange em redes neurais artificiais. O Ebia prevê o uso e aplicação de técnicas de acompanhamento do processo de decisão da máquina, porém não informa como e com qual recurso irá implementar estas políticas.

No eixo 5 que trata sobre força de trabalho e capacitação envolve a capacitação em nível de graduação ou pós graduação lato e stricto sensu de mão de obra direcionada a capacitação a temas de IA e Ciência de Dados. As principais ações estratégicas são estabelecer parcerias com o setor privado e com a academia para capacitação de profissionais para o mercado da IA; empresas públicas e privadas promovendo capacitação contínua de novas

forças de trabalho na IA; estimular a composição de gênero, raça, orientação sexula e outros aspectos socioculturais nas equipes de desenvolvimento da IA; reforçar políticas de interação entre empresas e academias (Brasil, 2021).

Neste sentido, com a era das tecnologias emergentes, houve uma polarização do emprego e crescimento de empregos com altos salários com maior nível de ensino em detrimento de empregos com menor nível de qualificação e baixa remuneração. Há também a complementaridade de produtividade dos trabalhadores pelas tecnologias disruptivas, em vez de deslocá-los dessas tarefas. Para este autor, o desenvolvimento da IA terá um deslocamento maior de em profissionais e técnicos mais qualificados, no entanto as medições deste fenômeno estão em andamento (Autor, Mindell e Reynolds*,* 2020). Assim, neste mesmo sentido, no Brasil onde a escolaridade formal é relativamente baixa, estes trabalhadores sofrerão ainda mais em relação a baixas remunerações a cada dia mais em detrimento de uma classe mais estudada (Arbix e Comin, 2020).

No eixo transversal sobre a Cooperação Internacional, A Parceria Global em Inteligência Artificial[55] (GPAI) é uma iniciativa do G7, sob a presidência canadense e francesa e tem como intuito reunir vários parceiros globais para preencher a lacuna entre a teoria e a prática em IA, apoiando pesquisas de IA. Os 15 membros que fazem parte do GPAI são Austrália, Canadá, França, Alemanha, Índia, Japão, México, Nova Zelândia, República da Coréia, Cingapura, Eslovênia, Reino Unido e União Europeia. Eles se juntaram a Brasil, Holanda, Polônia e Espanha em 2020 (Brasil, 2020a; GPAI, 2024).

O eixo de aplicação nos setores produtivos tem como foco aumentar a competitividade brasileira fomentando a indústria 4.0, desenvolvendo incentivos de uso da IA para pequenas e médias empresas e Startups. No eixo de aplicação no poder público a Estratégia de Governo Digital é implementar a IA em pelo menos 12 serviços públicos federais até 2022 e realizar análise de impacto nos casos de uso de IA que afetem os direitos do cidadão e do servidor público (Brasil, 2021).

[55] The Global partnership on Artificial Intelligence

## 5.8 Documentação de Projetos de Lei.

### 5.8.1 Projeto de Lei 5051/2019

O projeto de Lei 5051/2019[56] estabelece princípios para o uso da IA no Brasil e ressalta que esta tecnologia deve estar a serviço de melhorar o bem estar humano tendo como objetivo principal servir as pessoas respeitando à "dignidade humana", "à liberdade", "à democracia e à igualdade"; "respeito aos direitos humanos", "à pluralidade e à diversidade", com foco na "proteção dos dados pessoais" e a na "garantia da proteção da privacidade"; "transparência", "confiabilidade" e "possibilidade de auditorias de sistemas de IA" e a "supervisão humana" (Brasil, 2019) . Observa-se que o texto do projeto de lei foi omisso em relação a uma definição do conceito de Inteligência Artificial e aduz que uma legislação considerada robusta que quer se posicionar e se estruturar de maneira lógica frente à sociedade e a arcabouços jurídicos deve portanto conceituar o seu objeto de regulamentação (Peixoto e Coutinho, 2020).

Neste mesmo sentido Peixoto e Coutinho (2020), a coordenadora do grupo de estudos em direito e tecnologia da USP, De Oliveira (2022) ressalta que se no projeto de lei 5051/2019 não houve o debate conceitual de IA, o que dizer de temas que exigem mais consolidação de conhecimento tecnológico e jurídico do legislador como por exemplo: como garantir a explicabilidade e transparência da IA, o que dizer da estruturação de parâmetros para controle dos dados, no que tange a regulação setorial esta abordagem não seria a mais indicada no Brasil, no que diz respeito a responsabilidade civil sobre sistemas de IA ainda não ficou cristalino o debate neste projeto de Lei.

Outra crítica importante que Peixoto e Coutinho fazem em relação a PL 5051/2019 é nos seguintes termos:

> Assim, no processo analítico de aplicação do binômio adequação e necessidade, é possível construir uma crítica. Embora a regulação seja um fator relevante em atividades de inovação e fronteiriças do conhecimento humano. Já foi assim – um dia, para se regular a personalidade jurídica da empresa, a internet, as telecomunicações, as novas formas de energia,os cuidados com o meio ambiente, etc., mas seria

[56] Data da última movimentação do PL 5051/2019: 11/07/2024. Pesquisa no site do Senado Federal em 15 de agosto de 2024.

> necessário um PL para normatizar que o uso de IA deve respeitar a dignidade humana, à liberdade, à democracia e à igualdade? (Peixoto e Coutinho, 2020, p. 6)

No que tange ao mercado de trabalho, este PL trouxe o objetivo de valorização do trabalho humano em conjunto com o desenvolvimento de sistemas de IA e exige que os sistemas decisórios deverão sempre ser auxiliados à tomada de decisão auxiliados por tomada de decisão humana. Ressalta-se que quanto maior a gravidade das decisões submetidas aos sistemas de IA maior será a supervisão humana. E no que tange a responsabilidade foi simplista em resumir que a responsabilidade civil por danos decorrentes de sistemas de IA serão atribuídos a seu supervisor (Brasil, 2019). Para Fabiano e Coutinho (2020. p. 8) “limitar a ideia de responsabilizar somente o supervisor por danos produzidos por sistemas de IA seria reduzir a capacidade da IA de forma muito simplista, para os autores, seria interessante” (regular) a prática de boas práticas e princípios nos campos de responsabilidade ética e responsabilidade normativa, que devem refletir nas etapas de validação, verificação e controle e segurança no desenvolvimento e uso da IA".

As diretrizes para a atuação da União, Estados, Distrito Federal e Municípios no desenvolvimento da IA são especificados pela promoção da educação para o desenvolvimento mental, emocional e econômico; criação de políticas para a proteção e qualificação dos trabalhadores; adoção gradual da IA e uma ação na busca de regulamentar a matéria (Brasil, 2019).

Assim para De Oliveira (2022, p. 154), "ao buscar uma aprovação rápida de um projeto de lei como este, há um risco de o Poder Judiciário ficar responsável por disciplinar a matéria que possui muitas lacunas nos atuais projetos de lei, causando um ambiente de incerteza jurídica" capaz de obstaculizar o progresso do desenvolvimento tecnológico. Assim, o projeto de Lei 5051 de 2019, é muito semelhante ao Projeto de Lei n. 240 de 2020. Este PL está parado no Senado Federal e foi solicitado em 29 de junho de 2022, por meio do Requerimento n. 512 de 2022, a tramitação conjunta do PL n. 5691/2019 com o PL n. 21/2020, PL n. 5051/2019 e PL n. 872/2021 por tratarem de matéria correlata;

Assim, o Projeto de Lei 5051/2019 não inova em relação ao tema do desenvolvimento da IA no cenário brasileiro, podendo extrair estes princípios, fundamentos e objetivos do documento da OCDE que trata sobre o tema, no qual o Brasil é signatário, e considera-se que esta não atingiu o objetivo de uma legislação que é feita para uma tecnologia disruptiva como a IA, no entanto, apresenta-se como um marco regulatório, ainda em tramitação, no qual busca a proteção dos direitos dos cidadãos e o desenvolvimento econômico através do desenvolvimento de sistemas de IA. Assim, é necessário o aprofundamento dos legisladores sobre o tema de forma a poder concretizar uma legislação que possa estimular o avanço da IA e não o contrário (Peixoto e Coutinho, 2020).

### 5.8.2 Projeto de Lei 5691/2019

O Projeto de lei 5691/2019, institui a Política de Inteligência Artificial, e tem por "objetivo articular esforços e estimular a formação de um ambiente favorável à implantação de um ecossistema tecnológico que incorpore esse novo fator de crescimento" (Brasil, 2019). Os princípios da Política Nacional de IA estão centrados no "desenvolvimento inclusivo e sustentável"; "respeito à ética, aos direitos humanos e aos valores democráticos e à diversidade"; "proteção da privacidade e dos dados pessoais"; "transparência , segurança e confiabilidade"(Brasil, 2019). As diretrizes da Política Nacional de Inteligência Artificial, rol exemplificativo, centra-se em estabelecer "padrões éticos"; "promoção de crescimento inclusivo e sustentável"; "melhoria da qualidade e da eficiência dos serviços oferecidos à população"; "estímulo a investimentos públicos e privados em pesquisa de IA"; "cooperação entre entes públicos, setores públicos e privados e entre empresas"; "desenvolvimento de estratégias para incrementar o intercâmbio de informações e a colaboração entre especialistas e instituições"; "capacitação de profissionais na área da tecnologia em IA" e "valorização do trabalho humano" (Brasil, 2019).

De acordo com Argôlo dos Santos e Almeida Santos (2024), os projetos de lei no Congresso Nacional tem uma abordagem mais conceitual, sem entrar em detalhes específicos sobre a regulamentação setorial da IA, tarefa que é delegada aos ministérios especializados e agências reguladoras. No modelo regulatório do Reino Unido, entende-se que reguladores com expertise em

seus respectivos setores estão mais capacitados para identificar e gerenciar os riscos associados à IA, permitindo uma regulamentação proporcional e adequada às particularidades de cada área.

E neste mesmo sentido, ressalta que o Judiciário, Executivo e Legislativo no Brasil são grandes utilizadores de IA, em razão da grande quantidade de dados produzidos por estes. A EBIA traz diversas estratégias para a consolidação de uma regulamentação e uso ético da IA, no entanto a regulamentação federal ainda não veio, o que se observa são vários projetos de lei sobre o tema. A regulamentação federal deve ser mais do que conceitos, políticas e diretrizes deve orientar além do desenvolvimento interno a aquisição de serviços de IA para que estes três poderes prestem serviços a cada dia mais eficientes e para que empresas privadas também se beneficiem desta regulamentação. Assim, o PL 5691/2019 só traz conceitos principiológicos e de fato, não contribui para uma estrutura de governança que possa viabilizar as ações estratégicas da EBIA. (Argôlo dos Santos e Almeida Santos, 2024; De Oliveira, 2022).

Assim, o projeto de Lei 5051 de 2019, é muito semelhante ao Projeto de Lei n. 240 de 2020. Este PL está parado no Senado Federal e foi solicitado em 29 de junho de 2022, por meio do Requerimento n. 512 de 2022, a tramitação conjunta do PL n. 5691/2019 com o PL n. 21/2020, PL n. 5051/2019 e PL n. 872/2021 por tratarem de matéria correlata. Portanto, o projeto de lei 5691/2019 ainda não foi aprovado e tramita em conjunto com os anteriores (De Oliveira, 2022).

### 5.8.3 Projeto de Lei 21/2020

O Projeto de lei 21/2020[57] que estabelece princípios, direitos e deveres para o uso de IA está com a casa revisora (senado), aguardando aprovação do senado, o objetivo do projeto é **criar um marco legal ético**, transparente e responsável tanto para o setor público quanto para o setor privado. Dentre as várias iniciativas do PL 21/20 as principais são fixar princípios, diretrizes e fundamentos para o desenvolvimento da IA no Brasil (Brasil, 2020b; Colombelli,

[57] Data da última movimentação do PL 21/2020: 11/07/2024 (CTIA - Comissão Temporária Interna sobre Inteligência Artificial no Brasil). Pesquisa no site do senado em: 15 de agosto de 2024.

2024; grifo nosso)

Os principais objetivos, dispostos no artigo 3, na aplicação da IA no Brasil tem por foco o desenvolvimento econômico sustentável e inclusivo; aumento da competitividade brasileira no cenário interno e internacional; melhoria na prestação dos serviços públicos com o uso da IA; promoção da pesquisa e desenvolvimento com intuito de alcançar a inovação e a proteção do meio ambiente. Os fundamentos, dispostos no artigo 4, tem com missão alcançar o desenvolvimento científico e a inovação; livre iniciativa e concorrência; respeito a ética bem como dos direitos humanos e valores democráticos; livre manifestação do pensamento; não discriminação; estímulo à autorregulação, mediante a adoção de códigos de conduta e guia de boas práticas globais; segurança, privacidade e proteção dos dados pessoais, segurança da informação; acesso a informação; liberdade de modelos de negócios; proteção da livre concorrência e contra práticas abusivas de mercado e harmonização com as Leis 13.709/2018 (LGPD) e a lei 8.078/90 (CDC) dentre outras leis (Brasil, 2020b; Colombelli, 2024).

Os princípios norteadores para o desenvolvimento da IA no Brasil são a finalidade benéficas na consecução da implantação de sistemas de IA; centralidade do ser humano com respeito a dignidade da pessoa e aos direitos fundamentais; não discriminação mitigando o uso de sistemas de IA que possam ser usados com finalidades discriminatórias; a busca pela neutralidade dos desenvolvedores com a identificação de vieses contrários a legislação; transparência (direito das pessoas saberem que estão utilizando a IA); segurança e prevenção (adotar medidas técnicas, organizacionais e administrativas alinhadas às melhores práticas e padrões internacionais, visando gerenciar e mitigar riscos durante todo o ciclo de vida dos sistemas de inteligência artificial); inovação responsável e disponibilidade de dados (desde que não viole o direito do autor de uso de dados, bancos de dados e de textos por ele protegidos). (Brasil, 2020b; Colombelli, 2024).

Quanto ao risco atribuído a sistemas de IA, a autora De Oliveira (2022) preceitua que o PL 21/2020 não estabelece definições e classificação de riscos gerando ponto de obscuridade do projeto quanto ao tema:

> conceito de sistemas de IA apresentado no projeto de lei é contestável e exigiria mais debate e detalhamento; b) não há

> diretrizes para a elaboração de relatórios de gestão de risco por empresas que desenvolvem atividades de alto risco; c) não há classificação de risco dos sistemas de IA; d) estabelece a responsabilidade civil subjetiva como regime padrão em matéria de IA, mas não esclarece como a culpa será verificada quando houver discussões sobre a proteção do segredo industrial e comercial; e) não há explicação de como a culpa será analisada para fins de responsabilidade civil quando se está diante de danos provocados por aprendizado de redes neurais artificiais; f) não há previsão de como os dados utilizados para o treinamento de *machine learning* serão analisados pelo Poder Judiciário; g) não há clareza de como os segredos comercial e industrial serão protegidos nos casos de aquisição e uso de sistemas de IA pelo poder público, tendo em vista a efetividade do princípio da transparência na esfera pública, etc (De Oliveira, 2022, p.158 e 159).

A respeito dos riscos gerados por sistemas de IA, De Oliveira (2022, p.19) não compreende o motivo dos legisladores do Brasil não terem disciplinado os parâmetros no projeto de lei, "uma vez que a OCDE (2022) apresentou um Quadro Geral para a classificação de sistemas de IA, em 2022 com o objetivo de facilitar o trabalho de legisladores, políticos e juristas". Assim, a classificação de sistemas é parte fundamental para se estruturar uma gestão de riscos e para se chegar a estruturação factível de supervisão humana. Portanto, a PL21/2020 se perfaz de conceitos vagos e normas pouco estruturantes para alcançar o objetivo de governança proposto pelo EBIA.

Uma consideração e crítica importante a respeito de uma abordagem de regulação baseada no risco no que diz respeito a sistemas de IA feita por Hartmann Peixoto (2019):

> Uma abordagem da IA baseada no risco, argumentando que este debate tem sido preponderante sobre as discussões que buscam enfrentar os problemas apresentados pela tecnologia, negligenciando os inúmeros benefícios já experimentados pelo seu uso. Estar atento aos riscos é uma coisa, outra, muito distinta é basear a regulação pelos riscos (Hartmann Peixoto, 2019)

As principais diretrizes dispostas no PL 21/2020 ao disciplinar a IA são: intervenção subsidiária (regras específicas deverão ser criadas para o uso de sistemas de IA apenas quando absolutamente necessárias); gestão baseada em risco (Os sistemas de IA devem ser avaliados com base nos riscos específicos que apresentam, e a necessidade de regulação e o nível de intervenção devem ser proporcionais à gravidade e à probabilidade desses riscos), participação social contendo agentes de vários setores;

responsabilidade (salvo disposição legal em contrário a responsabilidade será subjetiva) e quando a utilização de sistemas de inteligência artificial envolver relações de consumo a responsabilidade será objetiva (o que evidencia o caráter protecionista e consumerista da legislação) (Brasil, 2020b).

A matéria referente a IA compete privativamente à União legislar e normatizar sobre a matéria em todo o território nacional. As principais diretrizes, segundo o projeto de lei, para atuação da União, Estados, Distrito Federal e Municípios, rol exemplificativo, são a promoção da interoperabilidade tecnológica dos sistemas de IA usados pelo poder público, estímulo e capacitação das pessoas ao mercado de trabalho frente aos novos desafios desta tecnologia; estímulo à criação de mecanismos de governança transparente e colaborativa com participação de todos os setores; promoção da cooperação internacional (Brasil, 2020a).

Uma importante ponderação feita por Kaufman (2022) sobre a dúvida de se ter e desenvolver uma regulamentação macro ou uma regulamentação setorial são os benefícios de que cada tipo de regulamentação trará para o sistema lógico jurídico levando em conta a eficácia concreta da lei para o país. Para Kaufman (2022. p.167) "se cada país tem seu Banco Central, que regula todo o funcionamento do sistema financeiro, qual o sentido de outro órgão definir e fiscalizar os procedimentos de concessão de crédito com Inteligência Artificial?". No entanto, para que uma regulamentação setorial aconteça de forma eficaz é necessário que a legislação macro esteja bem alicerçada em conceitos robustos, taxonomia e classificação de risco (De Oliveira, 2022).

No que tange a responsabilidade civil em sistemas de IA, a regra disposta no PL 21/2020 é a responsabilidade subjetiva. No entanto, nas relações de consumo há a previsão de indenização em relação a danos causados por sistemas de IA, sendo esta configurada como responsabilidade objetiva. Existe uma crítica da comunidade jurídica a respeito da responsabilidade, regra geral, ser subjetiva e temem por um difícil trajeto realizado pela vítima provar a culpa é dos agentes que trabalham no desenvolvimento de sistemas de IA (De Oliveira, 2022).

É imperioso destacar, após o debate sobre o projeto, de acordo com De Oliveira (2022) que se o projeto de lei n. 21/2020 for aprovado haverá uma

judicialização de temas de responsabilidade do Poder Executivo, e isto é fator de insegurança jurídica e de impedimento do avanço tecnológico:

> O Projeto de Lei n. 21 de 2020, o qual visa a ser o Marco Legal da Inteligência Artificial, não cumpre a sua função, sendo lacunoso e generalista em sua integralidade, ou seja, esse projeto de lei não regula o fenômeno emergente da inteligência artificial. O risco de aprovar um projeto de lei dessa maneira é delegar ao Poder Judiciário a regulamentação da matéria de IA, já que a lei é omissa em diversos momentos. Até ser consolidada a jurisprudência sobre as diversas temáticas que circundam a IA, será instaurado um ambiente de insegurança, o que poderá afastar investimentos em *startups* e empresas de IA no país (De Oliveira, 2022, p.158).

Assim, observa-se que até este marco temporal, os Projeto de Lei 5051/2019; Projeto de Lei 5691/2019; Projeto de Lei 21/2020, não satisfazem aos requisitos de uma lei que reúna requisitos para uma estratégia legislativa nacional de IA robusta, direcionada e efetiva e, que talvez a solução seja uma discussão aprofundada sobre o tema entre poder público e setores específicos e gabaritados- "centros de pesquisa na área da IA que poderão fornecer a base para construirmos um caminho em que o Brasil se torne competitivo no plano tecnológico" (Peixoto e Coutinho, 2020; De Oliveira, 2022, p.159).

### 5.8.4 Projeto de Lei 2338/2023

O projeto de lei de autoria do senador Rodrigo Pacheco, dispõe sobre o uso da Inteligência Artificial, e pode ser considerado o mais avançado em relação a todos os projetos de lei no Brasil sobre IA. Em 03 de fevereiro de 2022, os Projetos de Lei n. 5051/2019, PL 21/2020 e PL 872/21 passaram a tramitar conjuntamente no senado e em seguida foi instituída a Comissão de Juristas destinada a subsidiar a elaboração de uma minuta de substituição deles, o relatório Final da comissão de juristas informa que foi ouvido pessoas dos mais diversos setores da sociedade para a análise do fenômeno e implicações desta tecnologia emergente na sociedade. Eis que, a PL 2338 surgiu, este projeto de lei tem como objetivo estabelecer princípios, regras e diretrizes e fundamentos para o desenvolvimento da IA e também "uma abordagem baseada em riscos com uma modelagem regulatória baseada em direitos, estabelecendo instrumentos de governança e fiscalização para que sejam prestadas contas e promova o escrutínio individual e social em relação

aos sistemas de inteligência artificial" (Brasil, 2023; Oliveira Neto, 2023; Junior e Nunes, 2023 p.7775).

O projeto de lei tem como objetivo combinar o desenvolvimento e a inovação da IA de forma segura e responsável com a proteção dos direitos fundamentais em benefício da pessoa humana (art 1º). Os fundamentos basilares para a implementação de uma IA responsável estão alicerçados na "centralidade da pessoa humana", "respeito aos direitos humanos e aos valores democráticos"; o "livre desenvolvimento da personalidade"; "a proteção ao meio ambiente e o desenvolvimento sustentável"; "igualdade, a não discriminação, a pluralidade e o respeito aos direitos trabalhistas"; "desenvolvimento tecnológicos e a inovação"; "a livre iniciativa, a livre concorrência e a defesa do consumidor"; "privacidade, proteção de dados e a autodeterminação legislativa informativa"; "promoção da pesquisa e do desenvolvimento com finalidade de estimular a inovação nos setores produtivos e no poder público e acesso à informação e à educação, e a conscientização sobre os sistemas de inteligência artificial e suas aplicações" (Brasil, 2023; Oliveira Neto, 2023).

Os princípios elencados no artigo 3º do projeto de Lei 2338/2023 se amoldam aos princípios que a OCDE estabeleceu em 2019 para promover o uso responsável da IA sendo eles- "o crescimento inclusivo, desenvolvimento sustentável e bem estar"; "autodeterminação e liberdade de decisão de escolha"; "participação humana no ciclo da inteligência artificial e supervisão humana efetiva"; "não discriminação"; "justiça, equidade e inclusão"; "transparência, explicabilidade, inteligibilidade e auditabilidade"; "confiabiliade e robustez dos sistemas de inteligência artificial e segurança da informação"; "devido processo legal, contestabilidade e contraditório"; "rastreabilidade de decisões durante o ciclo de vida de sistemas de inteligência artificial como meio de prestação de contas e atribuição de resposabilidade a uma pessoa natural ou jurídica"; "prestação de contas, responsabilização e reparação integral de danos"; "prevenção, precaução e mitigação de riscos sitêmicos derivados de usos intencionais ou não intencionais e de defeitos não previstos de sistemas de IA" e por fim "não maleficência e proporcionalidade entre os métodos empregados e as finalidade determinadas e legítimas dos sistemas de IA" (OECD, 2019; Brasil, 2023; Oliveira Neto, 2023).

Este projeto de lei conceituou os sistemas de IA como – "sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real". O projeto de lei trouxe outros conceitos importantes-fornecedor de sistema de IA, operador de IA, agentes de IA, autoridade competente, discriminação, discriminação indireta e mineração de dados-serão tratados ao decorrer do trabalho em tópicos correlacionados e específicos ao tema (Brasil, 2023)

No capítulo 2, artigo 5º, dispõe que pessoas afetadas por decisões de sistemas de IA tem direito à informação prévia se forem colocados para interagir com algum sistema de IA ( inciso I); direito a explicação sobre decisão, recomendação o previsão de tomada por sistemas de IA (inciso II); direito de contestar decisões ou previsões de sistemas de IA que produzem efeitos jurídicos ou que impactam de maneira significativa os interesses dos afetados (inciso III); direito à determinação e à participação humana em decisões de sistemas de IA, levando-se em conta o contexto e o estado da arte do desenvolvimento tecnológico (inciso IV); direito à não discriminação e a correção de vieses discriminatórios diretos, indiretos[58], ilegais ou abusivos (inciso V); e direito à privacidade e à proteção de dados pessoais, nos termos da lei pertinente (VI) (Brasil, 2023).

Sobre o direito de receber informações claras sobre a contratação ou utilização do sistema de inteligência artificial, no artigo 7º, evidencia-se a importância de esclarecer ao cidadão o caráter automatizado da interação e da

[58] Artigo 4, inciso VI e VII, do projeto de lei traz o conceito de discrimanção e discriminação indireta: discriminação: qualquer distinção, exclusão, restrição ou preferência, em qualquer área da vida pública ou privada, cujo propósito ou efeito seja anular ou restringir o reconhecimento, gozo ou exercício, em condições de igualdade, de um ou mais direitos ou liberdades previstos no ordenamento jurídico, em razão de características pessoais como origem geográfica, raça, cor ou etnia, gênero, orientação sexual, classe socioeconômica, idade, deficiência, religião ou opiniões políticas. discriminação indireta: discriminação que ocorre quando normativa, prática ou critério aparentemente neutro tem a capacidade de acarretar desvantagem para pessoas pertencentes a grupo específico, ou as coloquem em desvantagem, a menos que essa normativa, prática ou critério tenha algum objetivo ou justificativa razoável e legítima à luz do direito à igualdade e dos demais direitos fundamentais (Brasil, 2023).

decisão em processos ou produtos que afetem a pessoa (inciso I); "descrição geral do sistema, tipos de decisões, recomendações ou previsões que se destina a fazer e consequências de sua utilização para a pessoa" (inciso II); "identificação dos operadores[59] do sistema de inteligência artificial e medidas de governança adotadas no desenvolvimento e emprego do sistema pela organização" (inciso III); "papel do sistema de inteligência artificial e dos humanos envolvidos no processo de tomada de decisão, previsão ou recomendação" (inciso IV); "categorias de dados pessoais utilizados no contexto do funcionamento do sistema de inteligência artificial" (inciso V); "medidas de segurança, de não-discriminação e de confiabilidade adotadas, incluindo acurácia, precisão e cobertura" (inciso VI); e "outras informações definidas em regulamento" (inciso VII). Nota-se o que é mais preocupante é disposto no § 2º no qual as pessoas que são expostas a sistemas de emoções ou a sistemas de categorização biométrica deverão ser avisadas sobre a utilização não possuindo nenhuma vedação sobre a utilização deste sistemas de forma clara. O § 3º  de forma bastante genérica aponta que os sistemas de IA desenvolvidos a grupos de vulneráveis deverão ser desenvolvidos de forma que este público saiba entender o funcionamento da IA e seus direitos (Brasil, 2023).

Assim sendo, a pessoa afetada por sistemas de IA poderão solicitar, de acordo com artigo 8º, explicação sobre decisão ou recomendação, tendo informações a respeito dos critérios e procedimentos utilizados e incluindo informações sobre a lógica do sistema; grau de contribuição de sistemas de IA para a tomada de decisão; dados processados para a decisão e quando apropriados a ponderação aplicado à pessoa afetada; mecanismos por meio dos quais a pessoa pode contestar a decisão; possibilidade de solicitar intervenção humana. O prazo para a disponibilização destas informações é de 15 dias a contar da solicitação, podendo ser prorrogado por igual período (Brasil, 2023).

---

[59] Artigo 4, inciso III, operador de sistema de inteligência artificial: pessoa natural ou jurídica, de natureza pública ou privada, que empregue ou utilize, em seu nome ou benefício, sistema de inteligência artificial, salvo se o referido sistema for utilizado no âmbito de uma atividade pessoal de caráter não profissional (Brasil, 2023).

Em sequência, sobre o direito de contestar decisões o artigo 9º do Projeto de Lei 2338/2023 garante aos indivíduos afetados por decisões de sistemas de inteligência artificial o direito de contestar e solicitar revisão de decisões, recomendações ou previsões que tenham impactos jurídicos significativos ou que afetem de maneira relevante seus interesses. Além disso, assegura o direito de corrigir dados incompletos, inexatos ou desatualizados, bem como de solicitar a anonimização, bloqueio ou eliminação de dados tratados de forma inadequada. O artigo também abrange a contestação de decisões baseadas em inferências discriminatórias, irrazoáveis ou que violam a boa-fé objetiva, incluindo aquelas fundadas em dados inadequados, métodos imprecisos ou que não consideram devidamente as características individuais dos afetados (Brasil, 2023).

Por conseguinte, o artigo 10 prevê que, quando decisões, previsões ou recomendações de sistemas de inteligência artificial resultarem em efeitos jurídicos significativos ou impactarem de forma relevante os interesses de uma pessoa, inclusive fazendo a perfilização de perfis, esta poderá solicitar a intervenção ou revisão humana. Entretanto, caso seja comprovadamente impossível implementar tal intervenção, o responsável pelo sistema deverá adotar medidas alternativas eficazes, garantindo a reanálise da decisão contestada, considerando os argumentos da parte afetada e assegurando a reparação de eventuais danos. Ademais, o artigo 11 estabelece que, em situações onde as decisões de IA tenham impactos irreversíveis ou envolvam riscos à vida ou à integridade física, é necessário que haja um envolvimento humano significativo no processo decisório, com a decisão final sendo determinada por um humano, E por fim, o capítulo 2 encerra-se com a proibição de utilizar os sistemas de IA de modo a acarretar discriminação direta, indireta ilegal ou abusiva em razão do gênero, cor ou etnia, orientação geográfica, raça, cor, classe econômica e social, religião e opinião política (dados pessoais sensíveis) (Brasil, 2023).

O capítulo 3 aborda sobre a categorização dos riscos sobre os sistemas de IA, ponto não aprofundado pelo Projeto de Lei 21/2020 consideradas normas de risco pouco estruturantes para garantir o objetivo proposto pela EBIA. Deste modo, o PL 2338/2023 no artigo 13 versa que, antes de sua comercialização ou uso em serviço, todo sistema de inteligência artificial deve

passar por uma avaliação preliminar conduzida pelo fornecedor, a fim de classificar seu grau de risco, conforme os critérios estabelecidos nesta legislação. Adicionalmente, para sistemas de inteligência artificial de propósito geral, essa avaliação incluirá as finalidades ou aplicações previstas na lei. O fornecedor deve documentar e registrar essa avaliação para garantir a responsabilidade e a prestação de contas, especialmente se o sistema não for classificado como de alto risco. A autoridade competente possui o poder de reclassificar o sistema, exigindo, se necessário, uma avaliação de impacto algorítmico. Caso a reclassificação identifique o sistema como de alto risco, será obrigatória a realização de tal avaliação e a implementação de medidas de governança, sem prejuízo das sanções aplicáveis em casos de avaliações fraudulentas, incompletas ou inverídicas (Brasil, 2023; Oliveira Neto, 2023; Junior e Nunes, 2023).

Ademais, sob o aspecto de classificação dos riscos de forma detalhada, o projeto de lei classifica em a) risco excessivo (art 14, 15 e 16) e b) alto risco (art 17 e 18). Nos sistemas de IA de risco excessivo- o artigo 14 estabelece a proibição da implementação e uso de sistemas de inteligência artificial que empreguem técnicas subliminares para induzir comportamentos prejudiciais à saúde ou segurança das pessoas, conectando essa restrição com a necessidade de proteger os fundamentos legais e éticos que regem o uso dessas tecnologias. Além disso, veda o uso de IA que explore vulnerabilidades de grupos específicos, como idosos ou pessoas com deficiência, de modo a prejudicar sua saúde ou segurança, e proíbe o poder público de utilizar IA para avaliar ou classificar indivíduos com base em comportamento social ou atributos pessoais de maneira ilegítima ou desproporcional, ressaltando a importância da proporcionalidade e da legitimidade no acesso a bens, serviços e políticas públicas (Brasil, 2023; Oliveira Neto, 2023; Junior e Nunes, 2023).

Da mesma maneira, no artigo 15, estabelece diretrizes rigorosas para o uso de sistemas de identificação biométrica à distância em espaços públicos no contexto de atividades de segurança pública. Ele determina que tal utilização só é permitida quando houver uma previsão legal específica estabelecida por uma lei federal e que essa aplicação esteja acompanhada de uma autorização judicial. Esse requisito é especialmente relevante em situações que envolvem a persecução penal, em casos de crimes cuja pena máxima de reclusão seja

superior a dois anos, na busca de vítimas de crimes ou pessoas desaparecidas, ou em situações de flagrante delito. O texto exige que a legislação preveja medidas proporcionais e necessárias para atender ao interesse público, garantindo o devido processo legal e a revisão das inferências algorítmicas por agentes públicos antes da tomada de qualquer ação. No mesmo sentido, finalmente, o artigo 16 delega à autoridade competente a tarefa de regulamentar os sistemas de inteligência artificial classificados como de risco excessivo, estabelecendo um elo entre a necessidade de supervisão rigorosa e a mitigação de potenciais danos associados ao uso dessas tecnologias (Brasil, 2023).

Visto o anterior, analisar-se-á os sistemas de alto risco (artigo 17 e 18), assim, são classificados como sistemas de inteligência artificial de alto risco aqueles que desempenham funções críticas em diversas áreas. Primeiramente, incluem-se os sistemas utilizados na gestão de infraestruturas essenciais, como controle de trânsito e redes de abastecimento de água e eletricidade, devido à sua importância para a segurança pública (inciso I). Além disso, sistemas aplicados na educação e formação profissional, incluindo aqueles que determinam o acesso a instituições ou que monitoram o desempenho dos estudantes (inciso II) , também são considerados de alto risco, pois afetam diretamente o futuro e as oportunidades de indivíduos. Em adição, sistemas de IA utilizados em processos de recrutamento, avaliação de candidatos e gestão de trabalhadores são classificados como de alto risco, devido ao seu potencial impacto significativo nas vidas e carreiras das pessoas envolvidas (inciso III). A regulamentação também inclui sistemas que avaliam a elegibilidade para serviços essenciais, como assistência pública e seguridade social, devido à sua capacidade de influenciar o acesso a direitos fundamentais (inciso IV) (Brasil, 2023; Junior e Nunes, 2023).

Outros sistemas que entram nessa classificação são aqueles usados para avaliar a capacidade de crédito (inciso V), para determinar prioridades em serviços de emergência (bombeiros e assistência médica- inciso VI), e para auxiliar na administração da justiça (inciso VII). Estes têm um potencial impacto direto na segurança financeira e física das pessoas. Também são considerados de alto risco os veículos autônomos (inciso VIII), particularmente quando seu uso pode representar riscos à integridade física de pessoas, e aplicações na

área da saúde, onde a IA pode auxiliar diagnósticos e procedimentos médicos, influenciando diretamente a vida dos pacientes (IX). Além disso, a regulamentação abrange sistemas biométricos de identificação (X), investigações criminais e segurança pública, especialmente em avaliações individuais de riscos e na previsão de infrações(inciso XI). Sistemas de IA utilizados para análises complexas de grandes conjuntos de dados, em investigações policiais ou administrativas, também são incluídos devido à sua capacidade de descobrir padrões ou prever crimes, o que pode ter implicações significativas para a privacidade e os direitos individuais e a gestão da migração e controle de fronteiras (inciso XIV)(Brasil, 2023).

Neste sentido, Junior e Nunes (2023, p. 7779) esclarece que independente da proposta elaborada pela comissão de juristas brasileiros sobre a classificação de risco se faz importante analisar os sistemas de IA e sua possível classificação de risco sob uma perspectiva internacional no qual a OCDE elaborou um documento, "Framework for the Classification of AI systems" (OECD, 2022b), para auxiliar na classificação de riscos da IA, o documento conta com 37 perguntas a serem respondidas por aqueles interessados em determinar o risco da implementação de uma IA específica, assim:

> Perguntas como “os resultados do sistema podem impactar os direitos fundamentais?”, “a interrupção da função ou atividade do sistema afetaria os serviços essenciais? ” e “quão autônomas são as ações que o sistema e que papel os humanos desempenham? ” fazem com que, respondidas de maneira concisa e direta, possibilite que se enxergue os desafios trazidos por cada software em sua individualidade (Junior e Nunes, 2023. p. 7779)

Por fim, de acordo com o artigo 18, a responsabilidade pela atualização da lista de sistemas de inteligência artificial classificados como de risco excessivo ou alto risco é atribuída à autoridade competente. Essa atualização deve ser conduzida com base em determinados critérios, como a implementação em larga escala do sistema, considerando o número de pessoas afetadas, a extensão geográfica, e a duração e frequência do uso (inciso I). Além disso, se o sistema puder impactar negativamente o exercício de direitos e liberdades ou a utilização de serviços, ou ainda, se possuir um alto potencial danoso de ordem material, moral ou discriminatório, ele deverá ser

incluído na lista (inciso II) (Brasil, 2023; Oliveira Neto, 2023; Junior e Nunes, 2023).

Outros critérios incluem a afetação de grupos vulneráveis específicos e a possibilidade de que os resultados prejudiciais do sistema sejam irreversíveis ou de difícil reversão (inciso IV). Também se considera a inclusão de sistemas que tenham causado danos materiais ou morais anteriormente (inciso V) , que apresentem baixo grau de transparência, explicabilidade ou auditabilidade, dificultando seu controle ou supervisão (inciso VI e VII). O alto nível de identificabilidade dos titulares dos dados, especialmente quando envolve o tratamento de dados genéticos e biométricos, também é um critério relevante (inciso VIII). Ademais, a expectativa de confidencialidade no uso de dados pessoais no sistema é um fator que a autoridade deve considerar (inciso IX). Finalmente, o parágrafo único determina que qualquer atualização dessa lista deve ser precedida de consulta ao órgão regulador setorial competente, se existente, além de consultas e audiências públicas, e de uma análise de impacto regulatório. Isso garante que a atualização seja realizada de forma transparente e com a devida consideração dos impactos sociais e jurídicos (Brasil, 2023).

O Capítulo 4, nos artigos 19 a 26 -que desenvolve disposições gerais, medidas de governança para sistemas de IA de alto risco e avaliação de impacto algorítmico sobre sistemas de IA-medidas altamente importantes para estabelecer sistemas de IA confiáveis e responsáveis. Conforme o artigo 19, os agentes responsáveis por sistemas de inteligência artificial devem implementar estruturas de governança e processos internos que garantam a segurança dos sistemas e a proteção dos direitos das pessoas afetadas, em conformidade com a legislação aplicável. Essas estruturas devem incluir medidas de transparência tanto na interação com os usuários, utilizando interfaces claras homem-máquina, quanto nas práticas de governança e gestão de dados, assegurando a mitigação de vieses e a conformidade com as normas de proteção de dados. O parágrafo 1º e 2º enfatiza que essas medidas de governança devem ser aplicadas ao longo de todo o ciclo de vida do sistema de IA, desde a sua concepção inicial até a sua descontinuação, garantindo a segurança e a conformidade legal durante toda a existência do sistema e que a documentação técnica dos sistemas de IA de alto risco deve ser elaborada

antes de sua comercialização ou uso, e mantida atualizada durante toda a sua utilização (Brasil, 2023; Oliveira Neto, 2023).

A supervisão humana dos sistemas de IA de alto risco deve ser orientada para prevenir ou minimizar os riscos aos direitos e liberdades das pessoas. Isso implica que os supervisores humanos devem ser capazes de compreender as capacidades e limitações do sistema, identificar e resolver rapidamente quaisquer anomalias ou disfuncionalidades, e, se necessário, decidir pela não utilização do sistema, ignorar seus resultados ou até mesmo interromper seu funcionamento (Brasil, 2023).

A avaliação de impacto algorítmico  é uma obrigação dos agentes que operam sistemas de inteligência artificial classificados como de alto risco, conforme determinado na avaliação preliminar. Essa avaliação deve ser conduzida por profissionais ou equipes com conhecimento técnico, científico e jurídico, garantindo independência funcional para a realização do relatório. A autoridade competente deve ser notificada sobre esses sistemas de alto risco, por meio do compartilhamento tanto da avaliação preliminar quanto da avaliação de impacto algorítmico (artigo 22).O artigo 23 reforça a necessidade de que essa avaliação seja realizada por profissionais qualificados e com independência, e, em casos específicos, a autoridade competente pode exigir que a auditoria seja conduzida por equipes externas ao fornecedor do sistema (Brasil, 2023; Oliveira Neto, 2023).

Já o artigo 24 detalha a metodologia da avaliação de impacto, que deve incluir etapas como preparação, cognição dos riscos, mitigação e monitoramento contínuo dos riscos identificados. Ademais, o parágrafo 1º do artigo 24 destaca que a avaliação deve registrar os riscos conhecidos, os benefícios do sistema, a probabilidade e gravidade de consequências adversas, lógica do funcionamento, processo e resultado de testes e avaliações a impactos a direitos, treinamento e ações de conscientização dos riscos, medidas de transparência ao público e as medidas de mitigação adotadas. O princípio da precaução é enfatizado, especialmente para sistemas que possam gerar impactos irreversíveis. Os agentes de IA, após a introdução no mercado de sistema de IA de alto risco que observarem risco inesperado a direitos de pessoas naturais, obrigam-se a comunicar às autoridades competentes e às pessoas afetadas pelo sistema de IA (Brasil, 2023; Oliveira

Neto, 2023).

O artigo 25 esclarece que a avaliação de impacto deve ser um processo contínuo, com atualizações periódicas ao longo de todo o ciclo de vida do sistema, e prevê a participação pública na revisão dessas avaliações, cabendo à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto. Por fim, o artigo 26 garante que as conclusões das avaliações de impacto sejam públicas, respeitando os segredos industrial e comercial, e devem incluir informações sobre a finalidade do sistema, medidas de mitigação implementadas e seu patamar residual, e a participação de diferentes segmentos sociais afetados. Isso assegura a transparência e a responsabilidade na utilização de sistemas de IA de alto risco (Brasil, 2023; Oliveira Neto, 2023).

O capítulo V, destina-se à responsabilidade civil, portanto, o Art. 27 estabelece que fornecedores ou operadores de sistemas de inteligência artificial são responsáveis pela reparação completa de danos patrimoniais, morais, individuais ou coletivos causados por esses sistemas, independentemente do grau de autonomia do sistema. Em particular, para sistemas classificados como de alto risco ou risco excessivo, a responsabilidade é objetiva, sendo imputada conforme a participação no dano, enquanto para sistemas de menor risco, a culpa é presumida, invertendo-se o ônus da prova em favor da vítima(BRASIL, 2023; Oliveira Neto, 2023).

Por outro lado, o Art. 28 isenta de responsabilidade os agentes de inteligência artificial que possam comprovar que não colocaram o sistema em circulação ou que o dano foi causado exclusivamente por ação da vítima, de terceiros, ou por um evento fortuito. O Art. 29 complementa ao afirmar que, nas relações de consumo, as regras do Código de Defesa do Consumidor continuam aplicáveis, garantindo a harmonia entre as normativas existentes e as novas disposições legais referentes à inteligência artificial(Brasil, 2023; Oliveira Neto, 2023).

No que tange ao capítulo VI, Código de Boas Práticas e Governança, artigo 30, os agentes de inteligência artificial têm a possibilidade de formular, individualmente ou por meio de associações, códigos de boas práticas e governança. Esses códigos devem estabelecer diretrizes claras sobre a organização, regime de funcionamento, procedimentos, e tratamento de

reclamações das pessoas afetadas. Além disso, devem incluir normas de segurança, padrões técnicos, e medidas específicas para cada contexto de aplicação, além de ações educativas e mecanismos de supervisão e mitigação de riscos, assegurando a gestão adequada dos riscos associados aos sistemas de IA (Brasil, 2023; Oliveira Neto, 2023).

Assim, ao desenvolver essas regras de boas práticas, é fundamental considerar a finalidade, a probabilidade e a gravidade dos riscos e benefícios decorrentes, alinhando-se à metodologia prevista no art. 24 desta Lei (avaliação de impacto). Os desenvolvedores e operadores de IA também podem implementar programas de governança que demonstrem seu compromisso com normas e boas práticas, adaptados à escala e ao potencial de suas operações,integrados à estrutura geral de governança, conte com planos de resposta para possíveis resultados prejudiciais dos sistemas de IA e seja atualizado constantemente com base em informações de monitoramento contínuo e avaliações periódicas. A adesão voluntária a esses códigos de boas práticas pode ser vista como um indicativo de boa-fé e pode ser considerada pela autoridade competente na aplicação de sanções administrativas. A autoridade competente pode, ainda, realizar análises de compatibilidade dos códigos de conduta com a legislação vigente, promovendo sua aprovação, publicização e atualização periódica (Brasil, 2023).

Outro aspecto relevante, o Capítulo VII, aborda sobre a Comunicação de Incidentes Graves, nesta diapasão Os agentes de inteligência artificial devem notificar a autoridade competente sobre a ocorrência de incidentes de segurança graves, especialmente aqueles que impliquem risco à vida ou à integridade física das pessoas, interrupção de operações críticas de infraestrutura, danos significativos à propriedade ou ao meio ambiente, bem como graves violações aos direitos fundamentais, conforme estabelecido pelo regulamento. Além disso, essa comunicação deve ser realizada dentro de um prazo razoável, conforme definido pela autoridade competente. Após a notificação, a autoridade avaliará a gravidade do incidente e, se necessário, poderá exigir que o agente adote medidas para reverter ou mitigar os efeitos do incidente (Brasil, 2023).

Sob o mesmo ponto de vista, o capítulo VII aborda sobre Supervisão e Fiscalização. Neste tocante, o Poder Executivo será responsável por designar uma autoridade competente para supervisionar e garantir a implementação e fiscalização desta Lei, com a incumbência de proteger direitos fundamentais e assegurar que os sistemas de inteligência artificial operem de maneira conforme os princípios estabelecidos. Essa autoridade terá a responsabilidade de promover a elaboração, atualização e implementação da Estratégia Brasileira de Inteligência Artificial, promover boas práticas e elaborar normas regulamentares, incluindo procedimentos para avaliação de impacto algorítmico e certificação de sistemas de alto risco (Brasil, 2023; Oliveira Neto, 2023).

Além disso, a autoridade terá o papel de fiscalizar e aplicar sanções em casos de descumprimento, podendo também solicitar informações detalhadas sobre o uso de sistemas de IA por entidades públicas. A coordenação com órgãos reguladores de setores específicos será fundamental para garantir o cumprimento das normas, e todas as regulamentações deverão ser precedidas por consultas públicas e análises de impacto regulatório. Essa interação inclui a possibilidade de estabelecer condições especiais para pequenas empresas e startups, visando adaptar a regulação às suas particularidades e promover a inovação de forma responsável (Brasil, 2023; Oliveira Neto, 2023).

Por conseguinte, o Art. 34 estabelece que a autoridade competente, juntamente com os órgãos e entidades públicas responsáveis pela regulação de setores específicos da atividade econômica e governamental, coordenarão suas ações para garantir a plena execução desta Lei. Essa coordenação é essencial para assegurar que as diretrizes legais sejam respeitadas em todas as áreas afetadas pela inteligência artificial. Além disso, o §1º determina que a autoridade competente manterá um fórum permanente de comunicação com esses órgãos e entidades, promovendo cooperação técnica para facilitar suas funções regulatórias, fiscalizatórias e sancionatórias. No §2º, é previsto que, em ambientes regulatórios experimentais (sandbox regulatório) que envolvam IA e sejam conduzidos por órgãos públicos, a autoridade competente deve ser informada e poderá se manifestar sobre o cumprimento dos objetivos e princípios desta Lei. O Art. 35, por sua vez, exige que todas as normas e regulamentos emitidos pela autoridade competente sejam precedidos por consultas e audiências públicas, bem como por análises de impacto

regulatório(Brasil, 2023; Oliveira Neto, 2023).

Sobretudo a de se destacar, a consagração de um aspecto legislativo de cunho punitivo para àqueles que não se encaixarem nas exigências legais. Deste modo, os agentes de inteligência artificial que violarem as normas estabelecidas por esta Lei estão sujeitos a diversas sanções administrativas aplicadas pela autoridade competente, que variam de advertências até multas substanciais, limitadas a R$ 50 milhões por infração ou até 2% do faturamento da empresa, no caso de pessoas jurídicas. Outras sanções incluem a publicização das infrações, proibição de participação em sandboxes regulatórios por até cinco anos, e a suspensão ou proibição de atividades relacionadas ao sistema de IA em questão. A aplicação dessas sanções será feita de forma gradativa ou cumulativa, considerando fatores como a gravidade da infração, a condição econômica do infrator, a boa-fé, e a adoção de boas práticas e de governança e medidas corretivas (BRASIL, 2023; Oliveira Neto, 2023).

Ademais, antes ou durante o processo administrativo, a autoridade competente pode adotar medidas preventivas, incluindo multas cominatórias, observado o limite de R$ 50.000.000,00 (cinquenta milhões de reais) por infração, sendo, no caso de pessoa jurídica de direito privado, de até 2% (dois por cento) de seu faturamento,para evitar danos irreparáveis ou garantir a eficácia do processo. Essas sanções administrativas são complementares às sanções civis ou penais previstas em outras legislações, como o Código de Defesa do Consumidor e a Lei Geral de Proteção de Dados. Em casos de infrações envolvendo sistemas de risco excessivo, haverá a aplicação obrigatória de multas e, para pessoas jurídicas, a suspensão parcial ou total de suas atividades. A autoridade competente deverá definir, por meio de regulamento, os procedimentos e critérios para a aplicação dessas sanções, garantindo transparência e fundamentação detalhada em todos os seus aspectos(Brasil, 2023; Oliveira Neto, 2023).

E por fim, nos artigos 38 a 45, a autoridade competente pode autorizar ambientes regulatórios experimentais, conhecidos como *sandboxes* regulatórios, para incentivar a inovação em inteligência artificial, conforme requisitos estabelecidos por lei e regulamentação. As entidades interessadas devem apresentar projetos que demonstrem inovação tecnológica, ganhos de

eficiência e benefícios sociais, incluindo um plano de descontinuidade para garantir a viabilidade pós-experimentação. A autoridade também regulamentará o processo de solicitação e pode limitar ou interromper o funcionamento dos *sandboxes*, sempre considerando a preservação de direitos fundamentais e a proteção de dados pessoais. Participantes desses ambientes continuam responsáveis por quaisquer danos causados durante a experimentação(BRASIL, 2023; Oliveira Neto, 2023).

### 5.9 Sistema Nacional de Processamento de Alto Desempenho-SINAPAD (Decreto n. 5.156/2004)

O SINAPAD, unidade de pesquisa do MCTIC, desde 2004, tem a importante função de prestar "serviços, sob demanda, à academia, governo e setor privado, fomentar e apoiar a formação de pessoal especializado, transferir conhecimento e tecnologia e difundir a cultura e aplicação de processamento de alto desempenho (art 1), incluindo aplicações de IA". Atualmente, o SINAPAD conta com 9 (nove) Centros Nacionais de Processamento de Alto Desempenho[60]. O SINAPAD foi criado para apoiar a execução de projetos de pesquisa e desenvolvimento que demandam grande capacidade computacional, permitindo a realização de simulações complexas, processamento massivo de dados e análises que seriam inviáveis em sistemas computacionais convencionais (Drummond e Carneiro, 2022. p. 28-29).

### 5.10 White paper *"Unpacking AI Procurement in a Box: insights from implementation"*

Este estudo foi elaborado pelo Fórum Econômico Mundial, em parceria com o centro para a quarta Revolução Industrial do Brasil (C4IR Brasil) e a comunidade internacional para atualizar as diretrizes que orientam a IA responsável e ética pelo setor público no cenário internacional. Este trabalho discute as ações necessárias para tornar o projeto " AI Procurement in a Box" mais acessível às nações do Sul Global (World Economic Forum, 2022).

---

[60] Recursos computacionais de alto desempenho dos CENAPADs: https://www.lncc.br/sinapad/resources.php?pg=resources

O "**AI Procurement in a Box**", é considerado um conjunto de ferramentas, para colocar em prática sistemas IA de forma segura, responsável e ética, implementadas em todo o mundo. No Brasil, existem dois projetos pilotos em andamento, o Hospital das Clínicas (HC) da Universidade de São Paulo e o metrô de São Paulo. No metrô de São Paulo o objetivo era adquirir um sistema de manutenção preditiva alimentado por IA com o foco no monitoramento dos trilhos. Os principais desafios e conquistas deste projeto piloto foram: primeira avaliação de impacto algorítmico no Brasil, conforme o EBIA, e a criação de um conselho independente de especialistas para avaliar riscos do projeto (World Economic Forum, 2022).

No Hospital das Clínicas o desafio do projeto piloto foi a integração de mais de 60 sistemas de tecnologia da informação (TIC), tendo como foco a aquisição de uma IA com função de avaliar o viés e a qualidade dos dados em um nível mais alto de responsabilidade algorítmica e transparência. Os principais desafios e conquistas do projeto foram a mudança de visão da instituição no que tange a reestruturação de departamentos e criação de novos, desenvolvimento de uma nova plataforma para integrar outros sistemas e melhor padrões de privacidade e anonimato e principalmente a promoção de uma cultura de envolvimento das pessoas ao longo de todo o ciclo de vida dos dados (World Economic Forum, 2022).

De acordo com este relatório, a fase zero que exige requisitos para a adoção generalizada de IA/ML no setor público depende de uma maturidade técnica dos envolvidos para se obter resultados benéficos quanto ao desenvolvimento da IA. E ressalta que no Brasil "as entidades governamentais municipais são muito menos propensas a adotar tecnologias de IA/ML do que aquelas em nível federal ou estadual, principalmente devido a barreiras de maturidade técnica e prontidão", esta realidade tem haver com o acesso desigual dos dados e à infraestrutura de TI diferente em relação aos diferentes níveis de governo. O HC de São Paulo demonstra bem o caminho desta maturidade digital, no qual integrou o seu sistema para aumentar a disponibilidade de dados para os modelos de dados, criou um Hub de inovação para conectar pesquisadores e empresários e criou o In. Lab laboratório de IA para pesquisar e desenvolver modelos de IA para entrar em produção (World Economic Forum, 2022).

Outra questão que se faz necessário para as Nações do Sul Global, inclusive o Brasil, é estimular a confiança pública da IA na utilização do setor público implementando ferramentas de auto avaliação de impacto algorítmica, certificações e subvenções para uma IA ética. Uma das recomendações mais significativas é a utilização de uma avaliação de impacto algorítmica (AIA), como utilizada pela Diretiva Canadense sobre Tomada de Decisão Automatizada, que tem por objetivo analisar os potenciais riscos e impactos negativos de cada projeto relacionado a IA, criando estratégias para mitigar esses riscos. Em 2022, o Instituto de Direito Europeu publicou as suas Regras Modelo sobre Avaliação de Impacto de Sistemas Algorítmicos de Tomada de Decisão utilizados pela Administração Pública. Assim como na AIA canadense, as regras sugeridas categorizam os sistemas automatizados de tomada de decisão (ADMS) em categorias escalonadas, sendo que aqueles que correm maior risco necessitam de mais supervisão. O metrô de São Paulo fez a primeira avaliação de Impacto Algorítmico do Brasil adaptado às leis e regulamentações brasileiras se baseando no AIA canadense (World Economic Forum, 2022).

A supervisão humana é necessária para reduzir preconceitos e perigos nas ferramentas de IA. A estratégia do "human-in-the-loop", significa que os humanos devem avaliar, revisar e permanecer responsáveis pelos algoritmos. A Comissão Europeia adota a supervisão nas suas propostas de regulamento sobre IA fazendo com que os sistemas de IA de alto risco continuem responsáveis ao longo do tempo. Por mais que a estratégia "human-in-the-loop" não seja totalmente segura quanto a possíveis falhas, no entanto é uma estratégia que deve ser validada para que os sistemas sejam construídos o mais confiável possível Assim, de acordo com World Economic Forum (Fórum Econômico Mundial) os dois projetos pilotos no Brasil foram importantes para apoiar as nações do Sul Global a adotarem medidas eficazes e regulamentações na implementação da IA, ética, segura e responsável, tendo como observatório o que está sendo feito de melhor em países desenvolvidos. (World Economic Forum, 2022).

## 5.11 Resolução 332/2020 CNJ (promoção do uso ético da IA no judiciário)

A resolução 332/2020 do CNJ é considerada um marco regulatório no que tange a aplicação de sistemas IA no Judiciário. Tendo em vista que, no ordenamento jurídico não possui normas específicas de governança e parâmetros éticos para o desenvolvimento de IA, a resolução dispões sobre critérios "éticos, transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e objetiva contribuir com agilidade, coerência e isonomia na prestação jurisdicional". Outra questão importante, é que no Considerando da Resolução expressa que será considerado o contido na Carta Europeia de Ética sobre Uso da IA em sistemas Judiciais e seus ambientes (Brasil, 2020c; Drummond e Carneiro, 2022).

No artigo 5 e 6 da Resolução expressa-se que os modelos de IA devem oportunizar segurança jurídica assegurando o tratamento dos casos jurídicos iguais de maneira "absolutamente igual" e ressalta a importância de que no desenvolvimento e treinamento de modelos de inteligência artificial a exigência de dados que possuem o maior grau de representatividade, considerando a preocupação de resguardar os dados pessoais sensíveis, disposto na Lei 13.709/2028, e o segredo de justiça. O objetivo da implantação de sistemas de IA no Judiciário deve ser a promoção de igualdade, a não discriminação, a pluralidade e ao atendimento a sua finalidade de consecução de um julgamento justo com "condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos" (Brasil, 2020c).

No que diz respeito ao aspecto anti discriminação a artigo 7, ressalta a importância dos sistemas de IA antes de serem colocados em produção passarem por um homologação para avaliar possíveis indícios com aspectos de preconceitos ou vieses capazes de gerar tendências discriminatórias. Se identificado estes vieses discriminatórios no sistema de IA, deverá passar por uma correção e se não for possível a correção para a eliminação do aspecto discriminatório o sistema de IA deverá ser descartado e inutilizado, devendo constar a motivação de sua descontinuação. Uma importante crítica feita por Drummond e Carneiro (2022, p. 31) a respeito do artigo 7 é a de que a "resolução não possui ainda critérios técnicos para testar e acompanhar a

conformidade da performance dos algoritmos, conforme já relatado em " O Futuro da IA no Judiciário Brasileiro". Assim, Drummond e Carneiro (2022) finaliza com a consideração e seu texto de que, assim, como feito internacionalmente, o CNJ deverá criar outras diretrizes para que os sistemas de IA sejam avaliados periodicamente com níveis de pontuação para que seja avaliado os riscos dos sistemas de IA para ser implementado no Judiciário brasileiro. (Brasil, 2020c; Brehm *et al.*, 2020; Drummond e Carneiro, 2022)

A explicabilidade dos sistemas de IA mostra-se presente no artigo 8, VI, "fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de IA, especialmente quando essa for de natureza judicial". Apesar do título do Capítulo da Resolução ser "publicidade e transparência" nota-se o conceito de explicabilidade em um dos incisos (VI) para garantir a transparência no uso de IA pelo Poder Judiciário. A discussão sobre os critérios a serem utilizados pelo CNJ na explicabilidade não estão claros e precisarão de novos debates e critérios formais frente aos desafios que estes sistemas de IA promovem (Brasil, 2020c; Drummond e Carneiro, 2022).

No que tange ao tema Governança dos sistemas de IA o artigo 10 dispõe que os órgão envolvidos em projetos de IA deverão comunicar ao CNJ sobre seus projetos e todo a cadeia de desenvolvimento até a finalização e a consequente colocação em produção dos sistemas de IA e os resultados pretendidos com o projeto; depósito do modelo de IA na plataforma Sinapses[61], possuir a interface de programação de aplicativo (API) que permite a interoperabilidade com outros sistemas e a utilização de código aberto (*open source*) que traz inúmeros benefícios como a possibilidade de compartilhamento de conhecimento, melhoramento de tecnologias bem como a

[61] De acordo com a Resolução do CNJ, artigo 3, III, Sinapses: solução computacional, mantida pelo Conselho Nacional de Justiça, com o objetivo de armazenar, testar, treinar, distribuir e auditar modelos de Inteligência Artificial (BRASIL, 2020).SINAPSES, a “fábrica para modelos de IA” tem sido identificada pelo CNJ como um possível componente de uma estratégia de governança de IA. SINAPSES permitirá que os tribunais que utilizam o PJe e aqueles que não tenham times in-house de tecnologia possam escalar o uso de algoritmos em suas operações. A ferramenta está sendo elaborada apenas no Processo Judicial Eletrônico (PJe) e disponibilizada a outros tribunais para que a reutilizem, adaptem e até adicionem seus próprios algoritmos ao sistema. Os tribunais que construírem suas ferramentas de IA através do PJe também poderão incorporar os algoritmos por si desenvolvidos de volta ao sistema. Dessa forma, SINAPSES torna-se uma plataforma aberta de desenvolvimento de IA, na qual os tribunais usufruirão de diferentes sistemas, enquanto expandem outros. (Brehm et al., 2020)

interoperabilidade entre Tribunais e entre Judiciário e sociedade civil. Em relação ao projeto Sinapse, embora muitos dos algoritmos vêm sendo criados pelo time dos próprios Tribunais, existe a possibilidade de algoritmos serem criados por empresas ou startups, o Jusbrasil é um exemplo de empresa privada que tem um banco de dados com informações jurídicas. Assim, O CNJ deve esclarecer os métodos para que a comunidade jurídica se beneficie dos avanços feitos na iniciativa privada (Brasil, 2020c; Brehm *et al.*, 2020; Drummond e Carneiro, 2022).

Nos termos do artigo 13 da Resolução, os bancos de dados para treinamento da IA deverão ser preferencialmente governamentais, por serem consideradas mais seguras e aptas ao treinamento dos modelos de IA. O sistema de IA deverá impedir que os dados recebidos sejam alterados antes de colocados em modelos de treinamento, e observa que é imprescindível que seja mantido cópia (dataset) de cada versão de modelo desenvolvida. E ressalta a importância de uma segurança no que diz respeito ao armazenamento e a execução dos modelos de IA, conforme padrões de segurança da informação (artigo 16). (Brasil, 2020c; Brehm *et al.*, 2020; Drummond e Carneiro, 2022)

As decisões nos modelos automatizados de IA devem ser supervisionados por seres humanos (servidores públicos) e esta possibilidade de revisão garante que não haja qualquer vinculação à solução apresentada pela Inteligência Artificial. Assim, a resolução garante que esteja claro e que seja informada ao cidadão a interação dele com uma ferramenta de IA e de seu caráter não vinculativo (artigo 17 e 18). O artigo 19, vai um pouco mais além e abre uma prerrogativa de uso de sistemas de IA auxiliando na produção de elaboração de decisão judicial, desde que sejam observados os critérios e as técnicas utilizadas e que seja capaz de explicar (explicabilidade) os passos que conduziram ao resultado. Neste caso, o sistema de IA deve permitir a supervisão do magistrado competente (Brasil, 2020c; Brehm *et al.*, 2020; Drummond e Carneiro, 2022).

Em matérias penais, evita-se a propagação da utilização de modelos preditivos para perpetuar decisões com viés racial e social, assim como já retratado em sistemas de IA utilizados no Poder Judiciário internacionalmente. No artigo 23, os sistemas de IA não poderão ser usados em relação a modelos

de decisões preditivas, no entanto, o seu uso em aplicações de soluções destinados a "cálculos de pena, prescrição, verificação de reincidência, mapeamentos, classificação e triagem dos autos para fins de gerenciamento do acervo são permitidos". Ressalta-se que "os modelos de IA destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização". Os órgãos do Poder Judiciário poderão realizar cooperação técnica com outras instituições, públicas ou privadas ou sociedade civil, para o desenvolvimento colaborativo de modelos de IA" (Brasil, 2020c; Drummond e Carneiro; Brehm *et al.*, 2020).

Para finalizar, segundo Brehm *et al.*(2020, p. 13-14), às ferramentas de IA no Poder Judiciário tem a função de "desde a classificação de processos até evitar que os servidores públicos realizem tarefas repetitivas ou mesmo para fornecer recomendações para o julgamento de uma demanda". Eis, assim, uma lista de ferramentas em produção no Judiciário brasileiro - o que chamou mais atenção dos autores foram o projeto Victor (STF) e Hércules (TJAL) que foram criados em parceria com a Universidade de Brasília e Universidade Federal de Alagoas, este tipo de parceria com centros de referência em educação vem sendo realizado internacionalmente como no Reino Unido- projeto Turing Center)- A ferramenta Victor objetiva "simplificar o reconhecimento de padrões em textos jurídicos (normalmente em um documento em PDF) apresentados perante o STF.

A ferramenta Sócrates "produz um exame automatizado de cada recurso encaminhado ao STJ e decisões prévias do processo, recomenda fontes normativas e precedentes jurídicos, e fornece uma recomendação de ação"; O tribunal de Justiça do Acre tem a ferramenta que se chama LEIA, "ferramenta vinculada ao e-SAJ, e não ao PJe, que lê PDFs e visa conectar cada processo a precedentes dos tribunais superiores. Outros tribunais que utilizam o e-SAJ como TJSC e o TJSP estão criando modelos semelhantes". O tribunal de Justiça de Alagoas com o sistema Hércules "ferramenta utilizada para evitar que o servidor público realize tarefas repetitivas, como classificar se um documento é um pedido de bloqueio de bens ou suspensão do processo"; Tribunal de Justiça de Minas Gerais (TJMG) ferramenta radar "Identifica e separa recursos que lidam com matérias jurídicas semelhantes ou possuem

precedentes nos Tribunais Superiores ou em Incidentes de Resolução de Demandas Repetitivas (IRDR)" (Brehm *et al.*, 2020, p. 13-14).

## 5.12 Política de Dados no Poder Judiciário (Resolução CNJ n. 331/2020 e Recomendação CNJ n. 74/2020)

A Resolução 331/2020 tem por objetivo instituir a Base Nacional de Dados Processuais do Poder Judiciário- DataJud. Assim para fins da Resolução 331/2020 "O DataJud será alimentado com dados e metadados[62] processuais relativos a todos os processos físicos ou eletrônicos, públicos ou sigilosos (...)". Outra importante informação a respeito do Datajud é que "a carga inicial do DataJud conterá, no mínimo, os processos que estejam em tramitação no Poder Judiciário e os que tenham sido baixados a partir de 1º de janeiro de 2020". Cabe ao CNJ a responsabilidade de cuidar dos dados enviados ao DataJud e de sua confidencialidade, quando for o caso (Brasil, 2020d; Drummond e Carneiro; Brehm *et al.*, 2020).

A Resolução 331/2020 do CNJ considera API como "um conjunto de instruções e padrões de sistemas que possibilitem integração e intercâmbio de dados" (Brasil, 2020d) . Neste sentido, Brehm *et al.* (2020, p. 21) preceitua que:

> Um dos modos mais comuns para que os dados possam ser arquivados de maneira apropriada é ter uma Interface de Programação de Aplicativos (API) bem definida. As APIs definem o tipo de formatação e a metodologia de coleta de dados que deverá ser utilizada. De fato, parece que o SINAPSES já está utilizando APIs para facilitar o uso dos algoritmos elaborados dentro de sua estrutura. Não obstante, também deveria haver APIs para facilitar a integração de algoritmos “caseiros” dentro do sistema SINAPSES.

Na Recomendação do CNJ 74/2020 a respeito da Resolução CNJ 331/2020 resolve que a Recomendação "terá como função estabelecer diretrizes para avaliação e implementação de medidas à governança do acesso e uso massificado de dados no Poder Judiciário, exceto o Supremo Tribunal Federal". E por fim, a Recomendação explícita "a importância do desenvolvimento da tecnologia, em particular de técnicas de inteligência artificial, para a sistematização e processamento de informações sobre a produção jurídica dos tribunais". Assim, fica claro o compromisso do CNJ com

[62] De acordo com artigo 2, I, da resolução 331/2020, metadados processuais são informações estruturadas dos processos judiciais.

o DataJud de forma a poder obter uma política de dados abertos, assegurando que o desenvolvimento da tecnologia no Poder Judiciário deve acontecer assegurando os direitos já previstos em legislações vigentes e o respeito aos direitos fundamentais previstos na Constituição. A intenção da Resolução é preparar o Poder Judiciário para o desenvolvimento de ferramentas de IA através da integração de dados e de dados disponibilizados de forma aberta (Brasil, 2020d; Drummond e Carneiro; Brehm *et al.*, 2020).

# 6. Quais as contribuições legislativas internacionais de aplicação da Inteligência Artificial - União Europeia, Estados Unidos e Canadá - se ajustam à realidade legislativa brasileira para uma IA ética e responsável.

## 6.1. Quadro de semelhanças e contribuições legislativas internacionais entre Brasil e União Europeia.

| ASSUNTOS | UNIÃO EUROPEIA | BRASIL |
|---|---|---|
| **Qual a estratégia nacional para desenvolvimento de IA ética e responsável?** | Plano Coordenado sobre IA. Sendo que o "*Ethics Guidelines for Trustworthy AI*" é um documento que aponta as diretrizes éticas para o desenvolvimento da IA. | Plano Nacional de Inteligência Artificial de 2022 e Estratégia Brasileira de Inteligência Artificial de 2021. |
| **Legislação centralizada ou descentralizada e setorial em IA** | A regulação é centralizada e harmonizada através do *AI Act.* | Regulação da IA setorial ou descentralizada, o PL 2338/23 busca centralizar esta regulação após a sua vigência. |
| **Gestão de risco em IA** | No *AI Act*, há uma categorização específica de sistemas de IA que são considerados de "alto risco" e que, portanto, estão sujeitos a requisitos mais rigorosos de conformidade e gestão de risco. | De forma similar, o Projeto de Lei 2338/2023 propõe uma classificação dos sistemas de IA conforme o nível de risco que apresentam, com foco em sistemas que possam causar danos significativos, seja em termos de segurança, privacidade, ou discriminação. Esses sistemas de maior risco estão sujeitos a controles e obrigações mais rigorosas, incluindo a necessidade de realizar avaliações detalhadas de impacto e implementar medidas específicas de mitigação de riscos. |
| **Avaliação de** | A Lei *AI Act* fala em | A PL2338/23 fala de |

| | | |
|---|---|---|
| **impacto e monitoramento contínuo** | avaliação de impacto. Assim, o termo "avaliação de impacto" é mais abrangente e foca nos aspectos gerais do uso da IA do que o utilizado pelo Brasil. O AI Act exige a avaliação de impacto antes de colocados no mercado, além de monitoramento contínuo após a implementação de sistemas de IA, conforme artigo 71. | avaliação de impacto de "algoritmo". Esse tipo de avaliação é voltado para entender como um algoritmo pode influenciar decisões automatizadas e seu impacto sobre os direitos dos cidadãos, como privacidade, igualdade e não discriminação. |
| **Supervisão humana em sistemas de IA de alto risco** | O *AI Act* da União Europeia oferece um detalhamento significativo sobre a supervisão humana, especialmente para sistemas de IA considerados de "alto risco". | De forma semelhante, o PL 2338/23 sublinha a importância da supervisão humana em sistemas de inteligência artificial, especialmente aqueles considerados de "alto risco".<br>-Em ambas as legislações, a supervisão humana é vista como um mecanismo essencial para garantir que as decisões automatizadas possam ser monitoradas, corrigidas ou anuladas por humanos, quando necessário, para evitar ou mitigar danos.<br>-Em ambas as legislações preveem que as medidas de supervisão humana devem ser proporcionais ao nível de risco e autonomia do sistema de IA.<br>-Em ambas as legislações, há uma clara há uma clara atribuição de responsabilidade aos desenvolvedores e operadores de sistemas de IA para garantir que mecanismos de supervisão humana estejam adequadamente |

| | | |
|---|---|---|
| | | implementados antes que os sistemas sejam colocados em operação.<br>- Os dois marcos regulatórios sublinham a importância da supervisão humana como uma medida ética que protege os direitos fundamentais dos indivíduos. |
| **Boas práticas e governança** | **-**O *AI Act* incentiva o desenvolvimento de códigos de conduta baseados nas melhores práticas da indústria para a aplicação voluntária dos requisitos regulamentares, especialmente para sistemas de IA de que não são classificados como de alto risco, visando orientar as organizações na implementação ética e responsável de IA.<br>-O AI Act coloca uma forte ênfase na transparência e na responsabilidade (accountability), exigindo que as empresas mantenham registros detalhados e adotem práticas que garantam a rastreabilidade e a supervisão dos sistemas de IA.<br><br>-A governança no AI Act está fortemente ligada à proteção dos direitos fundamentais, garantindo que os sistemas de IA sejam desenvolvidos e implementados de maneira a respeitar a dignidade humana, a privacidade, e outros direitos essenciais. O AI Act estabelece mecanismos de | -O Projeto de Lei 2338/2023 no Brasil também incentiva a adoção de códigos de conduta como parte de uma autorregulação mais ampla, promovendo o uso ético e responsável de IA. As diretrizes devem estar alinhadas às melhores práticas, princípios de transparência e proteção de direitos.<br>-De forma semelhante, o projeto de lei brasileiro exige que as empresas sejam transparentes sobre o uso de IA, especialmente em aplicações que possam afetar os direitos fundamentais dos cidadãos. A responsabilização também é um componente chave, assegurando que as empresas sejam responsáveis pelos impactos de seus sistemas de IA.<br>-O projeto de lei 2338/2023 também enfatiza a importância da governança ética, exigindo que os sistemas de IA sejam projetados e usados de forma a proteger os direitos dos indivíduos, incluindo a privacidade, a não discriminação, e a |

| | | |
|---|---|---|
| | governança, como a criação de um conselho de IA para supervisionar a aplicação dessas normas.<br><br>-O AI Act permite que as práticas regulatórias evoluam conforme as tecnologias de IA se desenvolvem, incentivando a adaptação das normas de acordo com as melhores práticas e os avanços técnicos. | transparência. A criação de um órgão regulador ou a atribuição de responsabilidades a órgãos existentes para supervisionar a conformidade também faz parte dessa abordagem.<br>-O projeto de lei brasileiro também adota uma postura flexível, permitindo que as regulamentações sejam atualizadas para refletir as melhores práticas e novas descobertas no campo da IA, garantindo que a governança acompanhe a evolução tecnológica. |
| **Responsabilidade Civil por danos causados por sistemas de IA** | **-** O *AI Act* aplica o princípio da responsabilidade objetiva, onde a responsabilidade pelos danos causados por sistemas de IA não depende da comprovação de culpa ou negligência. No AI Act, essa abordagem é claramente estabelecida para garantir que os consumidores possam ser compensados por danos causados por defeitos ou falhas em sistemas de IA, sem a necessidade de provar que o provedor agiu de forma inadequada<br>-o AI Act exige que os provedores de IA adotem medidas imediatas de mitigação de riscos e suspendam testes em caso de incidentes graves | - De forma semelhante, o Projeto de Lei 2338/2023 também adota essa abordagem, buscando proteger os consumidores e outras partes afetadas pelos riscos associados ao uso de IA.<br><br>-O Projeto de Lei 2338/2023 também impõe obrigações aos operadores para que monitorem e mitiguem riscos associados ao uso de IA. Ambos os textos legislativos enfatizam que os operadores e provedores devem ser proativos na gestão de riscos, e qualquer falha em |

| | | |
|---|---|---|
| | O AI Act, no Artigo 31, exige que os organismos notificados tenham seguro de responsabilidade adequado, ou que o Estado-Membro assuma a responsabilidade. Essa medida visa garantir que qualquer dano causado por falhas na avaliação de conformidade seja compensado. | fazê-lo pode resultar em responsabilização objetiva.<br>-Embora o Projeto de Lei 2338/2023 possa não detalhar explicitamente a necessidade de seguro de responsabilidade para todos os operadores, ele também busca assegurar que os danos sejam reparados, mantendo a coerência com o princípio da responsabilidade objetiva. |
| **Legislação de IA que protege consumidores por danos causados por sistemas de IA** | -Ambos os marcos regulatórios enfatizam a proteção dos direitos dos consumidores e dos direitos fundamentais. No AI Act, isso é reforçado pelo Artigo 69, que aplica a responsabilidade objetiva conforme a Diretiva 85/374/EEC sobre produtos defeituosos.<br>-O AI Act, embora proteja os consumidores, faz isso dentro de um contexto regulatório mais amplo, focado na segurança e conformidade dos sistemas de IA como um todo. O foco particular é em sistemas de alto risco. Embora a proteção ao consumidor seja uma parte importante da legislação. | -Já no Projeto de Lei 2338/2023, a responsabilidade objetiva é utilizada para garantir que os operadores de IA sejam responsáveis pelos danos causados por suas tecnologias, assegurando que os direitos dos consumidores sejam preservados, mesmo sem a necessidade de comprovação de culpa.<br>-o Projeto de Lei 2338/2023 pode ser considerado mais diretamente orientado para a proteção dos consumidores. Ele parece integrar de forma mais explícita e abrangente as questões de proteção ao consumidor em seu texto. |
| **Supervisão e Comunicação de Incidentes** | -O AI Act atribui às autoridades de vigilância do mercado dos Estados-Membros a responsabilidade de supervisionar os sistemas de IA, investigando incidentes e assegurando a conformidade com as normas estabelecidas. | -Da mesma forma, o Projeto de Lei 2338/2023 propõe que órgãos reguladores específicos sejam responsáveis pela fiscalização e supervisão dos sistemas de IA no Brasil. Esses órgãos seriam encarregados de monitorar a conformidade com as leis, investigar |

| | | |
|---|---|---|
| | -A legislação da União Europeia exige que os provedores de sistemas de IA de alto risco comuniquem imediatamente às autoridades de vigilância do mercado qualquer incidente que possa comprometer a saúde, segurança ou os direitos fundamentais, permitindo uma resposta rápida e eficaz.<br><br>-O AI Act concede às autoridades de vigilância do mercado o poder de impor sanções, como multas e outras medidas administrativas, aos provedores de sistemas de IA que não cumprirem os requisitos legais ou que causarem danos devido a falhas em seus sistemas.<br><br>-A legislação europeia prevê a cooperação entre as autoridades de vigilância do mercado e outras entidades, como o "European Artificial Intelligence Board", para assegurar uma aplicação harmonizada e eficaz do regulamento em toda a União Europeia. | incidentes e aplicar sanções em caso de não conformidade.<br>-No Brasil, o projeto de lei também prevê a obrigação de comunicação de incidentes aos órgãos reguladores competentes. Esta comunicação é fundamental para garantir que as autoridades possam tomar medidas rápidas para mitigar danos e proteger os consumidores e a sociedade em geral.<br>-Similarmente, o Projeto de Lei 2338/2023 estabelece que os órgãos reguladores brasileiros podem aplicar sanções em casos de não conformidade ou de danos causados por sistemas de IA, incluindo a aplicação de multas e outras penalidades.<br>-No Brasil, o projeto de lei também enfatiza a importância da coordenação entre diferentes órgãos reguladores e entidades, garantindo uma abordagem integrada e coerente para a supervisão e fiscalização dos sistemas de IA. |
| **Penalidades** | -O AI Act, estabelece multas expressivas para infrações graves às regulamentações. As multas podem chegar a até 35 milhões de euros ou 7% do faturamento anual global, dependendo da gravidade da infração. | **-**Embora o valor específico das multas no Brasil não seja exatamente igual ao do *AI Act* europeu, o projeto de lei brasileiro prevê penalidades proporcionais à gravidade da infração, que podem incluir multas consideráveis, bem como |

| | | |
|---|---|---|
| | -Ambos os marcos legais utilizam um sistema de escalonamento das penalidades, onde as infrações mais graves, como o uso de IA em práticas proibidas ou que resultem em danos significativos aos direitos fundamentais, são penalizadas de forma mais severa.<br>-Tanto o AI Act quanto o Projeto de Lei 2338/2023 têm um enfoque significativo na proteção dos direitos fundamentais. As penalidades são estruturadas de forma a garantir que o uso da IA não infrinja direitos como a privacidade, a não discriminação e a segurança dos indivíduos | outras sanções administrativas.<br>-O Projeto de Lei 2338/2023 também adota uma abordagem semelhante, com multas e sanções que variam conforme a gravidade da infração e o potencial dano causado.<br><br>-Tanto o AI Act quanto o Projeto de Lei 2338/2023 têm um enfoque significativo na proteção dos direitos fundamentais. As penalidades são estruturadas de forma a garantir que o uso da IA não infrinja direitos como a privacidade, a não discriminação e a segurança dos indivíduos |

**Quadro 3:** Semelhanças e contribuições legislativas internacionais entre Brasil e União Europeia. Fonte: Próprio autor.

### 6.2. Quadro de semelhanças e contribuições legislativas internacionais entre Brasil e Estados Unidos.

| ASSUNTOS | Estados Unidos | BRASIL |
|---|---|---|
| **Qual a estratégia nacional para desenvolvimento de IA ética e responsável?** | Sim. A National AI Initiative Act de 2020 estabelece o plano estratégico e a estrutura legal para o avanço da IA nos Estados Unidos, enquanto a Executive Order on Safe, Secure, and Trustworthy AI fornece diretrizes específicas para garantir que esse avanço seja seguro, confiável e ético | Plano Nacional de Inteligência Artificial de 2022 e Estratégia Brasileira de Inteligência Artificial de 2021. |
| **Legislação centralizada ou** | A regulamentação da inteligência artificial nos | -Regulação da IA setorial ou |

| | | |
|---|---|---|
| **descentralizada e setorial em IA** | Estados Unidos é caracterizada por uma abordagem setorial e fragmentada, sem uma lei federal única que regule todos os aspectos da IA. A regulação ocorre principalmente através de agências federais específicas e legislações estaduais, com várias iniciativas legislativas em andamento para abordar as complexidades e desafios emergentes da IA<br>-Brasil adotando uma postura mais proativa e centralizada.<br>- Em relação ao desenvolvimento e à regulação de sistemas de IA os Estados Unidos vem adotando uma tendência na autorregulação e na regulação setorial.<br><br>-Não há até o momento nenhum projeto de lei centralizado e abrangente que regule a IA. | descentralizada, o PL 2338/23 busca centralizar esta regulação após a sua vigência.<br><br>- Em relação ao desenvolvimento e à regulação de sistemas de IA o Brasil vem adotando uma tendência mais proativa e centralizada.<br><br>-No Brasil existe a PL 2338/23. |
| **Gestão de risco em IA** | -O PL 2338/23 tem algumas semelhanças com os documentos da NIST, especialmente no que diz respeito à regulamentação e à governança de sistemas de IA. Assim como o framework da NIST, o PL 2338/23 busca estabelecer diretrizes para o desenvolvimento e uso responsável de IA, enfatizando princípios como transparência, responsabilidade e segurança. Ambos os documentos têm o objetivo de mitigar riscos associados à IA, como | O PL 2338/23, tem algumas semelhanças com os documentos da NIST, especialmente no que diz respeito à regulamentação e à governança de sistemas de IA. Assim como o framework da NIST, o PL 2338/23 busca estabelecer diretrizes para o desenvolvimento e uso responsável de IA, enfatizando princípios como transparência, responsabilidade e segurança. Ambos os documentos têm o |

| | | |
|---|---|---|
| | discriminação e vieses, além de garantir que os sistemas sejam seguros e confiáveis.<br><br>-Por outro lado, enquanto a NIST foca em um framework de gerenciamento de riscos mais abrangente e adaptável para diferentes indústrias e aplicações, a Lei 2338/23 é mais prescritiva, com ênfase específica na regulação de sistemas de IA de alto risco e a proteção de grupos vulneráveis, inspirando-se também na legislação europeia, como o AI Act da União Europeia. | objetivo de mitigar riscos associados à IA, como discriminação e vieses, além de garantir que os sistemas sejam seguros e confiáveis.<br><br>-Por outro lado, enquanto a NIST foca em um framework de gerenciamento de riscos mais abrangente e adaptável para diferentes indústrias e aplicações, a Lei 2338/23 é mais prescritiva, com ênfase específica na regulação de sistemas de IA de alto risco e a proteção de grupos vulneráveis, inspirando-se também na legislação europeia, como o AI Act da União Europeia. |
| **Avaliação de impacto e monitoramento contínuo** | -Atualmente, a avaliação de impacto nos Estados Unidos é geralmente orientada por diretrizes setoriais ou frameworks voluntários, como o *AI Risk Management Framework* desenvolvido pelo National Institute of Standards and Technology (NIST). Este framework sugere que as empresas realizem avaliações de impacto focadas em identificar riscos associados à segurança, privacidade, discriminação e direitos civis. O AI RMF não categoriza diretamente os sistemas em "alto risco". O AI RMF fala explicitamente sobre monitoramento contínuo de sistemas de IA. | O Brasil, através do Projeto de Lei 2338/23, está buscando uma abordagem centralizada e abrangente para a regulação da IA, impondo requisitos uniformes de avaliação de impacto e monitoramento contínuo. Nos Estados Unidos, a regulação é mais fragmentada, com padrões variando entre setores e a dependência de frameworks voluntários. |

| | | |
|---|---|---|
| | -Nos Estados Unidos, muitas práticas de avaliação e monitoramento são guiadas por diretrizes voluntárias, exceto em setores altamente regulamentados que possuem documentos setoriais de gestão de risco complementares ao *AI Risk Management Framework* (ex: saúde, finanças, energia, educação e aeroespacial).<br>-O NIST foca em um framework mais abrangente e adaptável para diferentes indústrias e aplicações.<br><br>-Quando o projeto de lei Algorithmic Accountability Act entrar em vigor o AI Risk Framework continuará sendo um documento complementar importante na avaliação de impacto.<br>-O Algorithmic Accountability Act, projeto de lei, que exige que empresas realizem avaliações de impacto e auditorias de algoritmos de sistemas de IA sob pena de serem responsabilizadas pela falta de transparência. Apesar do texto não falar sobre monitoramento contínuo, o conceito está implícito nas exigências de auditorias regulares. | -No Brasil, a avaliação de impacto e o monitoramento contínuo, conforme proposto pelo Projeto de Lei 2338/23, seriam obrigatórios para sistemas de alto risco.<br><br>-a Lei 2338/23 é mais prescritiva, com ênfase específica na regulação de sistemas de IA de alto risco e a proteção de grupos vulneráveis<br><br>-No Brasil, o PL 2338/23 aborda sobre as regras de avaliação de impacto e monitoramento contínuo.<br><br>-O projeto de lei 2338/23 prevê regras de avaliação de impacto de algoritmo e monitoramento contínuo em seu texto. |
| **Supervisão humana em sistemas de IA** | -Ambos os países reconhecem a importância da intervenção humana em sistemas críticos de IA e buscam regulamentar o uso dessas tecnologias de forma que proteja a | -Ambos os países reconhecem a importância da intervenção humana em sistemas críticos de IA e buscam regulamentar o uso dessas tecnologias |

| | | |
|---|---|---|
| | sociedade contra riscos potenciais, ao mesmo tempo em que aproveita os benefícios que a IA pode oferecer.<br><br>-Nos EUA, a supervisão humana é fortemente incentivada ou exigida em setores regulamentados, como saúde, finanças, defesa e transportes, onde a IA pode assistir nas decisões, mas os humanos mantêm a autoridade final, especialmente em casos de alto risco.<br>-O monitoramento contínuo também é uma prática padrão nos EUA, onde os sistemas de IA são supervisionados e auditados regularmente para garantir que continuem operando dentro dos limites de segurança e conformidade regulatória.<br><br>-O AI RMF aborda a gestão de riscos em sistemas de IA e inclui a orientação sobre supervisão humana<br>-O projeto de lei- *Algorithmic Accountability Act*- não menciona explicitamente a supervisão humana contínua, mas as exigências de auditorias e avaliações de impacto podem envolver supervisão humana para garantir que os algoritmos estejam operando de forma justa e em conformidade com as normas. | de forma que proteja a sociedade contra riscos potenciais, ao mesmo tempo em que aproveita os benefícios que a IA pode oferecer.<br>-O projeto de lei prevê a obrigatoriedade de supervisão humana em sistemas de IA, especialmente em sistemas classificados como de "alto risco".<br><br>-O projeto de lei propõe que a supervisão humana não seja estática, mas que haja um monitoramento contínuo dos sistemas de IA, permitindo que os operadores humanos intervenham quando necessário e que ajustes sejam feitos para manter a conformidade e a segurança ao longo do tempo. |

| | | |
|---|---|---|
| **Boas práticas e governança** | **-** O AI RMF, o framework enfatiza a importância de boas práticas como a transparência, justiça e segurança nos sistemas de IA, com orientações para governança ao longo de seu ciclo de vida.<br><br>- No aspecto da governança: Incentiva processos robustos de governança para garantir supervisão contínua, transparência e responsabilidade, mitigação de riscos e conformidade regulatória e resposta a novos riscos. Existem outros documentos que abordam boas práticas e governança, mas este trabalho focou mais no AI RMF.<br><br>-O Algorithmic Accountability Act exige que as empresas adotem práticas robustas para a gestão ética e responsável dos sistemas automatizados. A lei promove transparência, responsabilidade e mitigação de riscos, assegurando o uso justo e legal da IA. | -De forma O Projeto de Lei 2338/2023 no Brasil também incentiva a adoção de boa práticas incluindo a transparência, justiça e segurança no desenvolvimento e implantação de sistemas de IA.<br><br>-No PL 2338/23 de forma semelhante em relação a governança, incentiva a supervisão contínua, transparência e responsabilidade, mitigação de riscos e conformidade regulamentar<br><br>-De forma semelhante, o PL 2338/23 exige práticas de avaliação de impacto, transparência, responsabilidade e mitigação de riscos. |
| **Responsabilidade Civil por danos causados por sistemas de IA** | **-**A responsabilidade civil relacionada a sistemas de IA nos Estados Unidos é abordada dentro das estruturas legais tradicionais (direito civil), como negligência e responsabilidade por produtos defeituosos, em vez de uma legislação federal específica. | - No **PL 2338/23**, a responsabilidade civil dos desenvolvedores de IA é predominantemente objetiva para sistemas de alto risco, refletindo a seriedade dos possíveis impactos desses sistemas na sociedade. Esse tipo de |

| | | |
|---|---|---|
| | Desenvolvedores podem ser responsabilizados se não tomarem precauções adequadas ou se a IA for considerada defeituosa. Há também debates sobre a "lacuna de responsabilidade", que aborda dificuldades em atribuir responsabilidade por danos causados por sistemas de IA autônomos, destacando a necessidade de novas abordagens regulatórias à medida que a IA se torna mais complexa. | responsabilidade visa proteger os usuários e terceiros, assegurando que os desenvolvedores e operadores sejam diligentes na criação e implementação de sistemas de IA, especialmente aqueles com potencial de causar danos significativos. |
| **Legislação de IA que protege consumidores por danos causados por sistemas de IA** | -As leis dos EUA tendem a proteger menos diretamente os consumidores, com base em responsabilidade subjetiva e sem uma estrutura federal específica para IA, o que pode dificultar o acesso dos consumidores à compensação por danos, isto quando comparado ao PL 2338/23. | -O Projeto de Lei 2338/23 do Brasil oferece uma proteção mais abrangente e direta aos consumidores em relação aos danos causados por sistemas de IA, principalmente devido à responsabilidade objetiva para sistemas de alto risco e às exigências de avaliações de impacto e supervisão contínua. |
| **Supervisão e Comunicação de Incidentes** | -Nos Estados Unidos, a supervisão e a comunicação de incidentes relacionados a sistemas de IA são tratadas de forma setorial e fragmentada, com regulamentações específicas aplicáveis a diferentes indústrias. A supervisão contínua é geralmente incentivada por meio de frameworks como o **AI RMF** do NIST, e a comunicação de incidentes é exigida por agências | - O **Projeto de Lei 2338/23** no Brasil oferece uma abordagem mais detalhada e estruturada para a supervisão e a comunicação de incidentes em sistemas de IA, especialmente quando comparado à abordagem fragmentada dos Estados Unidos. Ele exige supervisão contínua, intervenção |

| | | |
|---|---|---|
| | reguladoras setoriais, como FTC dentre outras. Embora não haja uma legislação unificada que aborde esses aspectos para todos os sistemas de IA, os regulamentos setoriais e as diretrizes de melhores práticas desempenham um papel crucial na governança de IA nos EUA. | humana quando necessário, e a comunicação obrigatória de incidentes tanto às autoridades quanto aos usuários afetados. Essas medidas são projetadas para garantir que os sistemas de IA operem de maneira segura, transparente e responsável, proporcionando uma proteção robusta para os cidadãos contra os potenciais riscos associados ao uso de IA. |
| **Penalidades** | -Nos Estados Unidos, as penalidades impostas a empresas ou indivíduos que causarem danos com sistemas de IA são geralmente determinadas por meio de leis setoriais e regulamentos específicos, em vez de uma legislação unificada sobre IA. A aplicação dessas penalidades depende do setor em que o sistema de IA está sendo utilizado, e várias agências governamentais têm a autoridade para fazer cumprir essas normas. O FTC pode impor multas e exigir que as empresas corrijam práticas prejudiciais ou enganosas que envolvam IA, especialmente em relação à privacidade do consumidor e proteção contra discriminação.<br>As penalidades podem incluir multas substanciais, sanções administrativas, e | **-**O Projeto de Lei 2338/23 no Brasil impõe penalidades que incluem responsabilidade objetiva, multas, sanções, e obrigações de reparação para desenvolvedores e operadores de sistemas de IA de alto risco. As penalidades são aplicadas de forma centralizada por autoridades competentes, focando na proteção robusta dos direitos dos cidadãos. |

| | | |
|---|---|---|
| | obrigações de reparação, dependendo da gravidade do dano e das circunstâncias envolvidas. | |

**Quadro 4:** Semelhanças e contribuições legislativas internacionais entre Brasil e União Europeia. Fonte: Próprio autor.

### 6.3 Quadro de semelhanças e contribuições legislativas internacionais entre Brasil e Canadá.

| ASSUNTOS | CANADÁ | BRASIL |
|---|---|---|
| **Qual a estratégia nacional para desenvolvimento de IA ética e responsável?** | -"Pan-Canadian Artificial Intelligence Strategy" (Estratégia Pan-Canadense de Inteligência Artificial), lançada em 2017. | -Plano Nacional de Inteligência Artificial de 2022 e Estratégia Brasileira de Inteligência Artificial de 2021. |
| **Legislação centralizada ou descentralizada e setorial em IA** | -A regulação é centralizada, especialmente em nível federal, através da criação de leis como o Bill C-27 e a Artificial Intelligence and Data Act (AIDA). | Regulação da IA setorial ou descentralizada, o PL 2338/23 busca centralizar esta regulação após a sua vigência. |
| **Gestão de risco em IA** | -No AIDA, parte do Bill C-27, faz referência a "sistemas de alto impacto," mas a lei em si não define de forma específica e detalhada quais sistemas de IA que se enquadram nessa categoria. Em vez disso, a AIDA estabelece que regulamentos futuros, a serem desenvolvidos pelo governo, definirão os critérios específicos para classificar um sistema de IA como de "alto impacto."<br>- O Canadá está desenvolvendo uma estrutura regulatória para a IA, baseada no Bill C-27, AIDA, que inclui diretrizes | -De forma similar, o Projeto de Lei 2338/2023 propõe uma classificação dos sistemas de IA conforme o nível de risco que apresentam, com foco em sistemas que possam causar danos significativos, seja em termos de segurança, privacidade, ou discriminação. Esses sistemas de maior risco, "alto risco", estão sujeitos a controles e obrigações mais rigorosas, incluindo a necessidade de realizar avaliações detalhadas |

| | | |
|---|---|---|
| | de avaliação de riscos e proteção de privacidade. Ferramentas como o Privacy Impact Assessment (PIA) e o Cyber Security Risk Management Framework são usadas para mitigar riscos, enquanto normas internacionais, como as da OCDE e do NIST, influenciam as práticas para garantir uma IA segura e responsável. | de impacto e implementar medidas específicas de mitigação de riscos.<br>-O projeto de lei 2338/23 define o que são sistemas de IA de alto risco: são aqueles que afetam significativamente os direitos fundamentais, são utilizados em contextos críticos e que tenham impacto na vida das pessoas. |
| **Avaliação de impacto e monitoramento contínuo** | -O Bill C-27 e a AIDA estabelecem a base legal para a avaliação de impacto e monitoramento contínuo de IA no Canadá. Diretrizes detalhadas geralmente vêm de documentos complementares, como orientações do governo, o Privacy Impact Assessment (PIA), o Cyber Security Risk Management Framework, e padrões internacionais como o RMF do NIST. Esses frameworks auxiliam na implementação prática das exigências legais. | -A PL2338/23 fala de avaliação de impacto de "algoritmo". Esse tipo de avaliação é voltado para entender como um algoritmo pode influenciar decisões automatizadas e seu impacto sobre os direitos dos cidadãos, como privacidade, igualdade e não discriminação.<br>Sobre a avaliação de impacto e monitoramento contínuo, o Brasil segue uma linha semelhante ao *AI Act* exigindo avaliação de impacto sobre privacidade e direitos fundamentais e monitoramento contínuo. |
| **Supervisão humana em sistemas de IA** | -A AIDA prevê que a supervisão humana em sistemas de IA, especialmente aqueles classificados como de alto impacto, envolve a implementação de medidas para identificar, avaliar e mitigar riscos de | - Tanto a AIDA do Canadá quanto o PL 2338/2023 do Brasil enfatizam a necessidade de supervisão humana em sistemas de IA de alto risco, garantindo que decisões automatizadas |

| | | |
|---|---|---|
| | danos ou resultados enviesados que possam surgir do uso desses sistemas. Além disso, os responsáveis devem monitorar continuamente a conformidade dessas medidas e avaliar sua eficácia. A lei estabelece a necessidade de intervenção humana em sistemas automatizados para assegurar que decisões críticas possam ser supervisionadas e, se necessário, corrigidas ou revertidas. | possam ser monitoradas, corrigidas ou revertidas por humanos. Ambos os textos legislativos exigem medidas para identificar e mitigar riscos, com monitoramento contínuo para assegurar conformidade e segurança. |
| **Boas práticas e governança** | **-**O AIDA, prevê diretrizes para boas práticas e governança de sistemas de IA. Estas incluem a obrigatoriedade de que os responsáveis por sistemas de IA implementem medidas robustas para identificar, avaliar e mitigar riscos, com uma ênfase significativa em garantir a transparência, a segurança e a responsabilidade. Além disso, a AIDA permite que o Ministro emita diretrizes adicionais e regulamentos para apoiar a adoção dessas boas práticas e assegurar a conformidade com os requisitos da lei. | -Tanto a AIDA do Canadá quanto o PL 2338/2023 do Brasil destacam a importância de transparência, segurança e supervisão na governança de IA. Ambos exigem que os sistemas de IA sejam operados de forma responsável, com foco na proteção dos direitos fundamentais e na promoção de uma inovação ética, garantindo a responsabilidade e mitigando riscos associados ao uso da IA. |
| **Responsabilidade Civil por danos causados por sistemas de I A** | **-** A responsabilidade civil dos desenvolvedores e operadores de sistemas de IA de alto impacto, conforme a AIDA, é objetiva. Isso significa que podem ser responsabilizados por danos causados, independentemente de culpa ou negligência, | - De forma semelhante, o Projeto de Lei 2338/2023 também adota essa abordagem, buscando proteger os consumidores e outras partes afetadas pelos riscos associados ao uso de IA. |

| | | |
|---|---|---|
| | apenas pelo fato de o dano ter ocorrido. Isso destaca a importância de uma governança rigorosa e medidas preventivas para mitigar riscos e proteger direitos. | |
| **Legislação de IA que protege consumidores por danos causados por sistemas de IA** | -Tanto a **AIDA** no Canadá quanto o **Projeto de Lei 2338/23** no Brasil têm como objetivo proteger os consumidores contra danos causados por sistemas de IA. Ambos os marcos legais estabelecem responsabilidades para desenvolvedores e operadores de IA, assegurando que, em caso de danos, os consumidores tenham mecanismos de proteção e recurso. Embora os detalhes específicos possam variar entre as legislações, ambos os textos visam garantir que os sistemas de IA sejam seguros e operem em conformidade com os direitos dos consumidores, oferecendo proteção robusta contra possíveis danos. | -Tanto a **AIDA** no Canadá quanto o **Projeto de Lei 2338/23** no Brasil têm como objetivo proteger os consumidores contra danos causados por sistemas de IA. Ambos os marcos legais estabelecem responsabilidades para desenvolvedores e operadores de IA, assegurando que, em caso de danos, os consumidores tenham mecanismos de proteção e recurso. Embora os detalhes específicos possam variar entre as legislações, ambos os textos visam garantir que os sistemas de IA sejam seguros e operem em conformidade com os direitos dos consumidores, oferecendo proteção robusta contra possíveis danos. |
| **Supervisão e Comunicação de Incidentes** | -A **AIDA** exige que responsáveis por sistemas de IA de alto impacto adotem supervisão contínua, monitorem a conformidade com medidas de mitigação de riscos e comuniquem rapidamente ao Ministro | -O Projeto de Lei 2338/2023 do Brasil estabelece que os desenvolvedores e operadores de sistemas de Inteligência Artificial (IA), especialmente aqueles classificados como de alto risco, têm |

| | | |
|---|---|---|
| | qualquer incidente que possa causar danos. Essas medidas garantem a operação segura e supervisionada dos sistemas de IA. | a obrigação de garantir a supervisão contínua dos sistemas para identificar e mitigar riscos potenciais. Além disso, o projeto de lei prevê a comunicação obrigatória de incidentes que possam causar danos significativos aos indivíduos ou à sociedade. Essa comunicação deve ser feita de forma rápida e detalhada às autoridades competentes, permitindo uma resposta eficaz para mitigar os efeitos do incidente e proteger os direitos das pessoas afetadas. |
| **Penalidades** | -As penalidades incluem: **Multas Administrativas**: Organizações podem ser multadas em valores significativos, que podem chegar ao maior valor entre $10 milhões ou 3% da receita global anual da organização, para infrações menos graves. Para infrações mais graves, as multas podem alcançar até $25 milhões ou 5% da receita global anual, o que for maior. **Multas Criminais**: Em casos de violações mais sérias, as penalidades criminais podem ser aplicadas, incluindo multas e penas de prisão. As multas podem chegar aos mesmos valores mencionados para as infrações administrativas. | **-**O Projeto de Lei 2338/2023 do Brasil prevê uma série de sanções administrativas para infrações relacionadas ao uso de sistemas de IA, incluindo advertência, multa de até R$ 50 milhões por infração (ou até 2% do faturamento da empresa), publicização da infração, proibição de participação em sandbox regulatório, suspensão ou proibição do uso do sistema de IA, e proibição de tratamento de determinadas bases de dados. As penalidades são aplicadas de forma gradativa, considerando fatores como a |

| | | |
|---|---|---|
| | Além disso, os indivíduos responsáveis podem enfrentar até cinco anos de prisão. | gravidade da infração, reincidência, cooperação do infrator, e a adoção de boas práticas. Além disso, essas sanções não excluem a obrigação de reparação integral dos danos causados. |

**Quadro 5:** Semelhanças e contribuições legislativas internacionais entre Brasil e Canadá. Fonte: Próprio autor.

# CONCLUSÃO

A aplicação da Inteligência Artificial traz inúmeras vantagens, como a automação de tarefas repetitivas, aumento da produtividade e aprimoramento da tomada de decisões por meio da análise de grandes volumes de dados. Além disso, a IA pode melhorar a qualidade de serviços em áreas como saúde, educação, segurança, transporte e legislativa promovendo inovação, personalização de experiências e eficiência operacional. Essas tecnologias também contribuem para o avanço científico e econômico, criando novas oportunidades de emprego em setores emergentes e impulsionando a competitividade global.

No entanto, o uso da IA também apresenta desvantagens significativas, como a possível perda de empregos devido à automação, riscos à privacidade e segurança dos dados, e o viés algorítmico que pode levar a decisões injustas ou discriminatórias.

A falta de transparência em muitos sistemas de IA, a dependência excessiva da tecnologia, e a ausência de regulamentação adequada levantam questões éticas e morais sobre seu uso responsável. Além disso, desafios ambientais e o impacto da IA em países em desenvolvimento podem ampliar desigualdades e suscitar preocupações sobre soberania e segurança nacional.

Apesar dessas desvantagens, os benefícios da IA tendem a superar os riscos. A tecnologia também permite que empresas e governos otimizem recursos, reduzam custos e melhorem serviços. Ao melhorar a eficiência, personalizar experiências e abrir novas oportunidades de negócios, a IA tem o potencial de transformar positivamente a sociedade, tornando-a mais conectada, produtiva e adaptável às mudanças rápidas do mundo moderno.

No entanto, para que esses benefícios sejam plenamente realizados e os impactos negativos mitigados, é fundamental que, antes da criação de regulamentações tradicionais, haja uma tecnorregulação robusta que garanta o desenvolvimento ético e responsável da inteligência artificial.

Nos Estados Unidos, a abordagem para o desenvolvimento da IA tem sido caracterizada por uma tecnorregulação robusta, onde o foco é promover diretrizes flexíveis e princípios éticos sem a necessidade de uma lei formal

específica para a IA. Essa estratégia permite que a inovação tecnológica obtenha orientações claras para o uso responsável, sem limitar a competitividade das empresas em um mercado global em constante evolução. O governo dos EUA tem enfatizado a autorregulação pelas próprias indústrias de tecnologia, apoiando a criação de padrões voluntários e promovendo a colaboração entre setor privado, academia e agências governamentais para enfrentar os desafios éticos e de segurança da IA. Dessa forma, os Estados Unidos conseguem equilibrar o incentivo à inovação com uma governança adaptativa, permitindo o desenvolvimento dinâmico da IA.

Em contraste com a abordagem dos Estados Unidos, a União Europeia, o Canadá e o Brasil estão adotando estratégias mais regulamentadoras e normativas para o desenvolvimento e uso da IA. A União Europeia, por exemplo, com a *AI Act*, um marco regulatório abrangente que visa garantir o uso ético e seguro da tecnologia, classificando os sistemas de IA por níveis de risco e impondo obrigações específicas conforme sua aplicação. O Canadá segue uma linha semelhante, com o desenvolvimento de políticas públicas e estruturas regulatórias voltadas para a IA, focando na proteção de dados e nos direitos humanos, enquanto incentiva a inovação responsável. No Brasil, iniciativas legislativas, como o Projeto de Lei da Regulamentação da Inteligência Artificial (PL 2338/23), procuram estabelecer diretrizes éticas e práticas seguras, embora ainda estejam em fase de debate e elaboração. Comparativamente, esses países apostam em marcos legais mais definidos, buscando equilibrar o avanço tecnológico com a proteção dos direitos fundamentais e o bem-estar social.

É incerto se o direito, em sua forma tradicional, será capaz de responder a todos os problemas complexos e multifacetados advindos dos sistemas de IA. Embora a criação de uma governança regulatória robusta possa proporcionar um quadro normativo claro e seguro, protegendo direitos fundamentais e promovendo o uso responsável da tecnologia, ela também enfrenta desafios significativos. A velocidade de inovação na IA frequentemente supera o ritmo da legislação, o que pode tornar as regulamentações rapidamente obsoletas ou excessivamente restritivas, prejudicando a competitividade e a inovação. Além disso, uma abordagem regulatória rígida pode sufocar o desenvolvimento tecnológico e a

experimentação, limitando o potencial da IA de gerar benefícios econômicos e sociais. Por outro lado, a ausência de uma regulação adequada pode resultar em riscos elevados, como violações de privacidade, discriminação algorítmica e impactos negativos na segurança e na soberania nacional. Assim, o direito deve evoluir de maneira dinâmica e adaptativa, balanceando o incentivo à inovação com a proteção de direitos e a mitigação de riscos, e reconhecendo que nem todos os desafios da IA podem ser solucionados apenas com normas jurídicas tradicionais.

Nesse contexto, aduz que é fundamental adotar um tripé de ética, tecnorregulação e governança regulatória para orientar o desenvolvimento e a aplicação dos sistemas de IA de maneira equilibrada e eficaz. A ética fornece os princípios norteadores que asseguram o respeito aos direitos humanos, promovendo a justiça, a transparência e a responsabilidade no uso da tecnologia. A tecnorregulação permite uma abordagem flexível e adaptativa, capaz de responder rapidamente às inovações e aos desafios emergentes sem sufocar o potencial criativo e disruptivo da IA. Já a governança regulatória deve ser implementada em um momento oportuno, evitando tanto a rigidez prematura que possa engessar o desenvolvimento da tecnologia quanto a falta de controle que possa permitir abusos e riscos indesejados. Esse tripé garante que os países possam maximizar os benefícios da inteligência artificial, protegendo os interesses e direitos dos cidadãos, enquanto estimulam o avanço tecnológico e a inovação de forma sustentável e responsável.

# REFERÊNCIAS


ALBEN; Alexander. **When Artificial Intelligence and Big Data Collide-How Data Aggregation and Predictive Machines Threaten our Privacy and Autonomy.** School of Law, University of California, Los Angeles. V.1, Issue 1, Fall 2020. Página 2. Disponível em: https://www.researchgate.net/publication/346745060_When_Artificial_Intelligence_and_Big_Data_Collide-How_Data_Aggregation_and_Predictive_Machines_Threaten_our_Privacy_and_Autonomy. Acesso em: 29 ago. 2024.

ALENCAR, Ana Catarina. **Inteligência Artificial, Ética e Direito.** Saraiva Educação SA, 2022. Disponível em: https://play.google.com/books/reader?id=Yl9oEAAAQBAJ&pg=GBS.PT5&hl=pt. Acesso em: 30 ago. de 2024.

AMARO JR, Edson *et al.* Utilização de Inteligência Artificial em Saúde: lições aprendidas durante o enfrentamento ao surto de COVID-19. **Panorama setorial da Internet**, v. 2, n. 12, p. 1-11, 2020. Disponível em: https://cetic.br/media/docs/publicacoes/6/20200908170853/panorama_setorial_ano-xii_n_2_Ano%20XII%20-%20N.%202%20-%20inteligencia_artificial_e_saude.pdf. Acesso em: 29 ago. 2024.

ANGWIN, J. *et al.* Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. **ProPublica** [Online], May 23, 2016. Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 21 ago. 2024.

ANNONI Alessandro; BENCZUR Peter; BERTOLDI Paolo *et al.*, **Artificial Intelligence**: A European Perspective, Craglia, M. editor(s), EUR 29425 EN, Publications Office of the European Union, Luxembourg, 2018.

APPIO; Francesco; LA TORRE; Davide; LAZZERI; Francesca; MASRI; Hatem; SCHIAVONE; Francesco. **Impact of Artificial Intelligences in business and Society Opportunities and Challenges**. Routledge, 2024. Disponível em: https://www.routledge.com/Impact-of-Artificial-Intelligence-in-Business-and-Society-Opportunitie/Appio-La-Torre-Lazzeri-Masri-Schiavone/p/book/9781032303413?_gl=1*sj781y*_ga*NjY0NzgzOTYzLjE2OTg5MzExMjU.*_ga_0HYE8YG0M6*MTY5ODkzMTEyNi4xLjEuMTY5ODkzMTM1NC4wLjAuMA..#. Acesso em: 29 ago. 2024.

APPIO, Francesco Paolo et al. Artificial Intelligence: Technological Advancements and Methodologies. In: **Impact of Artificial Intelligence in Business and Society.** Routledge. 2023, p. 13-81.

ARBIX. Glauco; COMIN. Alvaro A. A pandemia, a tecnologia e o trabalho no meio da encruzilhada. Panorama Setorial da Internet, numero 4, dezembro de 2020, ano 12. Disponível em:

https://cetic.br/media/docs/publicacoes/6/20201223152932/panorama_setorial_ano-xii_n_4_inteligencia_artificial_trabalho_O%20trabalho%20do%20futuro_moldando%20a%20tecnologia%20e%20as%20instituicoes.pdf.pdf. Acesso em: 29 ago. 2024.

ARGÔLO DOS SANTOS, Paulo; ALMEIDA SANTOS, João. Os Desafios da Regulamentação da Inteligência Artificial (IA) no Brasil em Relação a Alguns Países Desenvolvidos. **Revista FSA**, v. 21, n. 6, 2024.

ATTARD-FROST, Blair; BRANDUSESCU, Ana; LYONS, Kelly. The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. **Government Information Quarterly**, v. 41, n. 2, p. 101929, 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S0740624X24000212. Acesso em: 29 ago. 2024.

AUTOR. David; MINDELL. David; REYNOLDS. Elisabeth. O Trabalho do futuro: moldando a tecnologia e as instituições. **Panorama Setorial da Internet**. numero 4. dezembro. 2020, ano 12. Disponível em: https://cetic.br/media/docs/publicacoes/6/20201223152932/panorama_setorial_ano-xii_n_4_inteligencia_artificial_trabalho_O%20trabalho%20do%20futuro_moldando%20a%20tecnologia%20e%20as%20instituicoes.pdf.pdf. Acesso em: 29 ago. 2024.

AYINLA, Benjamin Samson *et al.* Ethical AI in practice: Balancing technological advancements with human values. **International Journal of Science and Research Archive**, v. 11, n. 1, p. 1311-1326, 2024. Disponível em: https://ijsra.net/content/ethical-ai-practice-balancing-technological-advancements-human-values. Acesso em: 29 ago. 2024.

BAYAMLIOĞLU, Emre; LEENES, Ronald. The 'rule of law' implications of data-driven decision-making: a techno-regulatory perspective. **Law, Innovation and Technology**, v. 10, n. 2, p. 295-313, 2018. Disponível em: https://www.tandfonline.com/doi/full/10.1080/17579961.2018.1527475. Acesso em: 29 ago. 2024.

BENDER, Emily M.; FRIEDMAN, Batya. Data statements for natural language processing: Toward mitigating system bias and enabling better science. **Transactions of the Association for Computational Linguistics**, v. 6, p. 587-604, 2018. Pagina 594. Disponível em: https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00041/43452/Data-Statements-for-Natural-Language-Processing. Acesso em: 29 ago. 2024.

BOHR, Adam; MEMARZADEH, Kaveh. The rise of artificial intelligence in healthcare applications. In: **Artificial Intelligence in healthcare**. Academic Press, 2020. p. 25-60. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780128184387000022. Acesso em: 29 ago. 2024.

BOLUKBASI, Tolga et al. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. **Advances in neural information processing systems**, v. 29, 2016. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html. Acesso em: 29 ago. 2024.

BOYD, Karen L. Datasheets for datasets help ML engineers notice and understand ethical issues in training data. **Proceedings of the ACM on Human-Computer Interaction**, v. 5, n. CSCW2, p. 1-27, 2021.

BRASIL. Conselho Nacional de Justiça. Resolução Nº 332/2020. Dispõe sobre ética, a transparência e a Governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. DJe/CNJ, nº 274, de 25/08/2020, p. 4-8.. Brasília. 2020c. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 29 ago. 2024.

BRASIL. Conselho Nacional de Justiça. Resolução Nº 331/2020. Institui a Base Nacional de Dados do Poder Judiciário-Datajus como fonte primária de dados do Sistema de Estatística do Poder Judiciário-SIESPJ para tribunais nos incisos II a VII do art. 92 da Constituição Federal. DJe/CNJ, nº 274, de 25/08/2020, p. 2-4. Brasília. 2020d. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3428. Acesso em: 29 ago. 2024.

BRASIL. Conselho Nacional de Justiça. Recomendação n. 74, de 21 de setembro de 2020. Brasília. 2020. Disponível: https://atos.cnj.jus.br/files/original172205202009225f6a32bd3f21d.pdf. Acesso em: 29 ago. 2024.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. Institui a política de Dados Abertos do Poder Executivo Federal. Diário Oficial da União: Brasília, DF, 12 maio de 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 29 ago. 2024.

BRASIL. Decreto nº 9.854, de 25 de junho de 2019. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas. Diário Oficial da União: seção 1, Brasília, DF, 26 jun. 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d9854.htm. Acesso em: 29 ago. 2024.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. Cooperação Internacional em Inteligência Artificial. 2020a. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/cooperacao_internacional_inteligencia_artificial. Acesso em: 29 ago. 2024.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Estratégia Brasileira de Inteligência Artificial. 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosin

teligenciaartificial/ebia-diagramacao_4-979_2021.pdf. Acesso em: 29 ago. 2024.

BRASIL. Ministério da Ciência, Tecnologia e Inovação - Conselho Nacional de Ciência e Tecnologia. IA para o bem de todos. Proposta de Plano Brasileiro de Inteligência Artificial 2024-2028. Publicado em 29 jul 2024. 2024b. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasileiro-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos/ia_para_o_bem_de_todos.pdf/view. Acesso em: 21 ago. 2024.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. Estratégia Brasileira para Transformação Digital. E-Digital. f.106. Brasília, 2018. Disponível em: https://www.gov.br/governodigital/pt-br/estrategia-de-governanca-digital/eDigital.pdf. Acesso em: 29 ago. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Novo Plano Brasileiro de Inteligência Artificial prevê o investimento de R$ 1,76 bi para melhoria de serviços públicos. Publicado em 30 jul. 2024. 2024a. Disponível em: https://www.gov.br/gestao/pt-br/assuntos/noticias/2024/julho/novo-plano-brasileiro-de-inteligencia-artificial-preve-o-investimento-de-r-1-76-bi-para-melhoria-de-servicos-publicos. Acesso em: 21 ago. 2024.

BRASIL. Senado Federal. Projeto de Lei da Câmara dos Deputados nº 21, de 2020. 2020b. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1723640822719&disposition=inline. Acesso em: 29 ago. 2024.

BRASIL. Senado Federal. Projeto de Lei do Senado Federal nº 5691, de 2019. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=8009064&ts=1723640980812&disposition=inline. Acesso em: 29 ago. 2024.

BRASIL. Senado Federal. Projeto de Lei n. 21/2020. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil. 2020. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1720798342741&disposition=inline. Acesso em: 29 ago. 2024.

BRASIL. Senado Federal. Projeto de Lei n. 2338/23. Brasília. 2023. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1723640838762&disposition=inline. Acesso em: 29 ago. 2024.

BRASIL. Tribunal de Contas da União. 5 motivos para a abertura de dados na Administração Pública. Brasília. 2015. Disponível em: https://portal.tcu.gov.br/data/files/81/55/71/DB/A592C710D79E7EB7F18818A8/5_motivos_abertura_dados_administracao_publica.PDF. Acesso em: 29 ago. 2024.

BREHM, Katie et al. **O futuro da IA no sistema judiciário brasileiro.** Mapeamento, Integração e Governança da IA. Orientação: André Corrêa D'Almeida. Tradução de Matheus Drummond e Matheus de Souza Depieri. p, v. 8, 2020.

CANADA. Algorithmic Impact Assessment tool. 30 de maio de 2024. Disponível em: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html. Acesso em: 29 ago. 2024.

CANADA. BILL C-27 summary: Digital Charter Implementation Act, 2022b. Disponível em: https://ised-isde.canada.ca/site/innovation-better-canada/en/canadas-digital-charter/bill-summary-digital-charter-implementation-act-2020. Acesso em: 29 ago. 2024.

CANADA. Directive on Automated Decision-Making. 25 de abril de 2023. 2023a. Disponível em: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592. Acesso em: 29 ago. 2024.

CANADA. Pan-Canadian Artificial Intelligence Strategy. Government of Canada. 2022a. Disponível em: https://ised-isde.canada.ca/site/ai-strategy/en#pillar1 Acesso em: 20 ago 2024.

CANADA. The Artificial Intelligence and Data Act (AIDA)-Companion document. 2023b. Disponível em: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document. Acesso em: 29 ago. 2024.

CAPEL, Tara; BRERETON, Margot. What is human-centered about human-centered AI? A map of the research landscape. In: **Proceedings of the 2023 CHI conference on human factors in computing systems**. 2023. p. 1-23.Disonível em: https://dl.acm.org/doi/abs/10.1145/3544548.3580959. Acesso em: 29 ago. 2024.

CARRILLO, Margarita Robles. Artificial intelligence: From ethics to law. **Telecommunications policy**, v. 44, n. 6, p. 101937, 2020.

CHENG, Lu; VARSHNEY, Kush R.; LIU, Huan. Socially responsible ai algorithms: Issues, purposes, and challenges. **Journal of Artificial Intelligence Research**, v. 71, p. 1137-1181, 2021. Disponível em: https://www.jair.org/index.php/jair/article/view/12814. Acesso em: 29 ago. 2024.

CHOLLET, François. A definition of intelligence for the real world. **Journal of Artificial General Intelligence**, v. 11, n. 2, p. 27-30, 2020. Disponível em: https://intapi.sciendo.com/pdf/10.2478/jagi-2020-0003#page=28. Pag. 8. Acesso em: 29 ago. 2024.

CHOUDHRY, Mavra; WALL, Nic; REYNOLDS, Molly. **Guide to artificial intelligence regulation in Canada**. 27 de abril de 2023. Disponível em: https://www.torys.com/our-latest-thinking/publications/2023/04/guide-to-artificial-intelligence-regulation-in-canada. Acesso em: 29 ago. 2024.

CHOWDHARY, KR1442; CHOWDHARY, K. R. Natural language processing. **Fundamentals of artificial intelligence**, p. 603-649, 2020. Disponível em: https://link.springer.com/chapter/10.1007/978-81-322-3972-7_19. Acesso em: 29 ago. 2024.

COLOMBELLI, Wagner Godinho. **Regulamentação da IA (Inteligência Artificial) na administração pública brasileira: análise do Projeto de Lei n° 21 de 2020 e Projeto de Lei n° 2338 de 2023**. 2024. Trabalho de Conclusão de Curso.

CORRÊA, Pedro Barros Nunes Studart. **De ontologia e deontologia**: a regra principiológica do direito contemporâneo. 2013. Disponível em: https://bdm.unb.br/bitstream/10483/6810/1/2013_PedroBarrosNunesStudartCorrea.pdf. Acesso em: 29 ago. 2024.

CORTIZ. Diogo. Inteligência Artificial: equidade, justiça e consequências. **Panorama setorial da Internet**. maio. 2020, ano 12. Disponível em: https://cetic.br/media/docs/publicacoes/6/20200626161010/panorama_setorial_ano-xii_n_1_inteligencia_artificial_equidade_justi%C3%A7a.pdf. Acesso em: 29 ago. 2024.

COWLS, Josh *et al.* Designing AI for social good: Seven essential factors. **Available at SSRN 3388669**, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388669. Acesso em: 29 ago. 2024.

DA SILVEIRA, Paulo Antônio Caliendo Velloso. **Ética e Inteligência Artificial**: da possibilidade filosófica de Agentes Morais Artificiais. 2021. Disponível em: https://tede2.pucrs.br/tede2/handle/tede/9534. Acesso em: 29 ago. 2024.

DE OLIVEIRA, Cristina Godoy Bernardo. Desafios da regulação do digital e da inteligência artificial no Brasil. **Revista USP**, n. 135, p. 137-162, 2022. Disponível em: https://www.revistas.usp.br/revusp/article/view/206257/189893. Acesso em: 29 ago. 2024.

DIVINO, Stephano Bruno Santos. **Estratégia Brasileira de Inteligência Artificial (EBIA) e políticas públicas propostas para efetivação dos eixos legislação e uso ético de IA**. v.15, 2022. Disponível em: https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php/acer

vo/buscador.html?task=detalhes&source=&id=W4315559662. Acesso em: 29 ago. 2024.

DOS SANTOS, Kalani Sobrinho; TORRES, Leonardo Guimarães. O Uso Da Tecnologia Na Atividade Administrativa Do Estado. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 10, n. 4, p. 1413-1426, 2024.

DRUMMOND, Matheus; CARNEIRO, João Víctor. **Panorama regulatório de Inteligência Artificial no Brasil**. Rio de Janeiro. 2022. Instituto de Tecnologia & Sociedade do Rio. Disponível em: https://itsrio.org/wp-content/uploads/2022/04/Relatorio-Panorama-IA.pdf. Acesso em: 29 ago. 2024.

EKBIA, Hamid *et al.* Big data, bigger dilemmas: A critical review. **Journal of the Association for Information Science and Technology**, v. 66, n. 8, p. 1523-1545, 2015. Página 1536. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23294. Acesso em: 15 set. 2024.

ELIAS, Paulo Sá. **Algoritmos, Inteligência Artificial e o Direito**. Conjur, novembro, 2017. Disponível em: https://www.conjur.com.br/dl/al/algoritmos-inteligencia-artificial.pdf. Acesso em: 29 ago. 2024.

ENARSSON, Therese; ENQVIST, Lena; NAARTTIJÄRVI, Markus. Approaching the human in the loop–legal perspectives on hybrid human/algorithmic decision-making in three contexts. **Information & Communications Technology Law**, v. 31, n. 1, p. 123-153, 2022. Disponível em: https://www.tandfonline.com/doi/full/10.1080/13600834.2021.1958860. Acesso em: 29 ago. 2024.

EUROPEAN COMMISSION. Coordinated Plan on Artificial Intelligence 2021 Review. COM(2021) 205 final. Brussels, 21.4.2021. Disponível em: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review. Acesso em: 29 ago. 2024.

EUROPEAN COMMISSION. Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (Text with EEA relevance). Official Journal of the European Union. 29.6.2023 Disponível em: https://eur-lex.europa.eu/eli/reg/2023/1230/oj. Acesso em: 29 ago. 2024.

EUROPEAN COMMISSION. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. COM/2020/64 final. Brussels, 19.2.2020. 2020b . Disponível em: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52020DC0064. Acesso em: 29 ago. 2024.

EUROPEAN COMMISSION. White paper on Artificial Intelligence: a European approach to excellence and trust. COM (2020). Publication date 19 February 2020. 2020a. Disponível em: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Acesso em: 29 ago. 2024.

EUROPEAN PARLIAMENT. European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf. Acesso em: 29 ago. 2024.

FELZMANN, Heike *et al.* Towards transparency by design for artificial intelligence. **Science and engineering ethics**, v. 26, n. 6, p. 3333-3361, 2020.

FERRAZ JUNIOR, Tercio Sampaio. **Introdução ao estudo do direito**: técnica, decisão, dominação / Tercio Sampaio Ferraz Junior. 4. ed. São Paulo: Atlas, 2003.

FERRER, Xavier *et al.* Bias and discrimination in AI: a cross-disciplinary perspective. **IEEE Technology and Society Magazine**, v. 40, n. 2, p. 72-80, 2021.

FINNIS, John. **Fundamentals of ethics**. Georgetown University Press, 1983.

FIRLEJ, Mikolaj; TAEIHAGH, Araz. Regulating human control over autonomous systems. **Regulation & governance**, v. 15, n. 4, p. 1071-1091, 2021. Disponivel em: https://onlinelibrary.wiley.com/doi/full/10.1111/rego.12344. Acesso em: 29 ago. 2024.

FLORIDI, Luciano. On the Brussels-Washington consensus about the legal definition of Artificial Intelligence. **Philosophy & Technology**, v. 36, n. 4, p. 87, 2023. Disponível em: https://link.springer.com/article/10.1007/s13347-023-00690-z. Acesso em: 29 ago. 2024.

FRANÇA, Gustavo. Finnis entre Aristóteles e Kant. **Synesis**, v. 14, n. 1, p. 37-61, jan/jul 2022, ISSN 1984-6754.

GALLAGHER, Michael M. Canada's Artificial Intelligence and data Act (AIDA) 2024: A Comprehensive Guide. 11 de abril de 2024. Disponível em: https://www.mondaq.com/canada/intellectual-property/1452828/canadas-artificial-intelligence-and-data-act-aida-2024-a-comprehensive-guide

GATT, Lucilla *et al.* The possible relationships between law and ethics applied to AI| Le possibili relazioni tra legge ed etica applicate all'IA. **European Journal of Privacy Law & Technologies**, n. 2, 2023. Disponível em:

https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1889. Acesso em: 29 ago. 2024.

GAWIEJNOWICZ, Stanisław. **Models and algorithms of time-dependent scheduling.** Berlin: Springer, 2020.

GEBRU, Timnit *et al.* Datasheets for datasets. **Communications of the ACM**, v. 64, n. 12, p. 86-92, 2021. Disponível: https://dl.acm.org/doi/fullHtml/10.1145/3458723. Acesso em: 29 ago. 2024.

GPAI. Global Partnership on Artificial Intelligence. The Global partnership on Artificial Intelligence. Disponível em: https://gpai.ai/. Acesso em: 29 ago. 2024.

GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. **Computer Law & Security Review,** v. 45, p. 105681, 2022.

GURSOY, Furkan; KENNEDY, Ryan; KAKADIARIS, Ioannis. A critical assessment of the algorithmic accountability act of 2022. **Available at SSRN 4193199**, 2022. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4193199. Acesso em: 29 ago. 2024.

GUTIERREZ; Anabel *et al.* Using privacy calculus theory to explore entrepreneurial directions in mobile location-based advertising: Identifying intrusiveness as the critical risk factor. **Computers in Human Behavior**, v. 95, p. 295-306, 2019. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0747563218304552?via%3Dihub. Pagina 11. Acesso em: 29 ago. 2024.

HAMON, Ronan; JUNKLEWITZ, Henri; SANCHEZ, Ignacio. Robustness and explainability of artificial intelligence. **Publications Office of the European Union**, v. 207, p. 2020, 2020. Disponível em: https://ai-watch.ec.europa.eu/system/files/2022-01/dpad_report.pdf. Acesso em 01 de julho de 2024.

HARTMANN PEIXOTO. Fabiano. **Direito e Inteligência Artificial**: Referenciais básicos com comentários à resolução CNJ 332/2020. Alteridade. v.2, 2020a.

HARTMANN PEIXOTO, Fabiano. **Inteligência Artificial e Direito**: Convergência Ética e estratégica, 1. Ed. Curitiba: Alteridade Editora, 2020. 2020b Disponivel em: https://play.google.com/books/reader?id=AMHoDwAAQBAJ&pg=GBS.PT2&hl=pt. Acesso em: 29 ago. 2024.

HARTMANN PEIXOTO, Fabiano; SILVA, Roberta Zumblick Martins da. Inteligência Artificial e Direito. **Coleção Direito, Racionalidade e Inteligência Artificial**. Curitiba: Alteridade, 2019.

HASSIJA, Vikas *et al.* Interpreting black-box models: a review on explainable artificial intelligence. **Cognitive Computation**, v. 16, n. 1, p. 45-74, 2024.

HERZOG, Christian. On the risk of confusing interpretability with explicability. **AI and Ethics**, v. 2, n. 1, p. 219-225, 2022. Disponível em: https://link.springer.com/article/10.1007/s43681-021-00121-9. Acesso em: 29 ago. 2024.

HICKOK, Merve. Lessons learned from AI ethics principles for future actions. **AI and Ethics**, v. 1, n. 1, p. 41-47, 2021.

INPLP. International Network of Privacy Law Professionals. Canada's proposed Artificial Intelligence and Data Act (AIDA). 24 de janeiro de 2023. Disponível em: https://inplp.com/latest-news/article/canadas-proposed-artificial-intelligence-and-data-act-aida/. Acesso em: 29 ago. 2024.

JUNIOR, Claudio do Nascimento Mendonça; NUNES, Dierle José Coelho. Desafios e oportunidades para a regulação da inteligência artificial: a necessidade de compreensão e mitigação dos riscos da IA. **Revista Contemporânea**, v. 3, n. 07, p. 7753-7785, 2023. Disponível em: https://ojs.revistacontemporanea.com/ojs/index.php/home/article/view/1146/726 . Acesso em: 29 ago. 2024.

KANT, Immanuel. **Fundamentação da metafísica dos costumes**. Leya, 2023.

KANT, Immanuel. **Sobre a pedagogia**. EDITORA VOZES, 2021.

KARDASH, Adam; POLATAIKO, Maryna; DEVIR, Gemma. The wave of privacy and data legislation reform will continue in 2024. Dezembro de 2023. Disponível em: https://legaloutlook.ca/the-wave-of-privacy-and-data-legislation-reform-will-continue-in-2024/. Acesso em: 29 ago. 2024.

KAUFMAN. **Desmistificando a inteligência artificial**. 1. ed. Belo Horizonte – MG. Autêntica. 2022.

KHARITONOVA, Yu S.; SAVINA, V. S.; PAGNINI, Fabrizio. Civil liability in the development and application of artificial intelligence and robotic systems: basic approaches. **Perm U. Herald Jurid. Sci**., v. 58, p. 683, 2022.

KIRAT, Th *et al.* Fairness and Explainability in Automatic Decision-Making Systems. A challenge for computer science and law. **EURO journal on decision processes**, v. 11, p. 100036, 2023. Disponível em: https://www.sciencedirect.com/science/article/pii/S2193943823000092. Acesso em: 29 ago. 2024.

KOULU, Riikka. Proceduralizing control and discretion: Human oversight in artificial intelligence policy. **Maastricht Journal of European and Comparative Law**, v. 27, n. 6, p. 720-735, 2020. Disponível em: https://journals.sagepub.com/doi/full/10.1177/1023263X20978649. Acesso em: 29 ago. 2024.

LA TORRE, Davide *et al.* **Impact of artificial intelligence in business and society**: Opportunities and Challenges. Routledge, 2023.

MADHAVAN, Raj *et al.* Toward trustworthy and responsible artificial intelligence policy development. **IEEE Intelligent Systems**, v. 35, n. 5, p. 103-108, 2020. DOI: 10.1109/MIS.2020.3019679. Disponível em: https://ieeexplore.ieee.org/abstract/document/9237282. Acesso em: 29 ago. 2024.

MAGRANI, Eduardo. **Entre Dados e Robôs**: Ética e privacidade na era da hiperconectividade. Arquipélago editorial LTDA. Porto Alegre-RS, 2019, p. 220.

MANDIEGA, Tambiama André. EU guidelines on ethics in artificial intelligence: Context and implementation. 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf. Acesso em: 29 ago. 2024.

MARTIN-BARITEAU, Florian; SCASSA, Teresa. Artificial Intelligence and the law in Canada. Artificial Intelligence and the Law in Canada (Toronto; LexisNexis Canada, 2021), 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3734675. Acesso em: 29 ago. 2024.

MCMILLAN-MAJOR, Angelina; BENDER, Emily M.; FRIEDMAN, Batya. Data statements: From technical concept to community practice. **ACM Journal on Responsible Computing**, v. 1, n. 1, p. 1-17, 2024.

MELO, Ana Karolina Acris. **Regulação da Inteligência Artificial.** Benchmarking de países selecionados. ENAP, Evidência Express (EvEx), dezembro de 2022. Disponível em: https://repositorio.enap.gov.br/bitstream/1/7419/1/2022.12.08%20-%20Regula%C3%A7%C3%A3o%20da%20Intelig%C3%AAncia%20Artificial.pdf. Acesso em: 29 ago. 2024.

MITTELSTADT, Brent Daniel et al. The ethics of algorithms: Mapping the debate. **Big Data & Society**, v. 3, n. 2, p. 2053951716679679, 2016. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951716679679. Acesso em: 29 ago. 2024.

MOHAMMAD, Saif. Ethics Sheets for AI Tasks. In: **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. 2022. p. 8368-8379. Disponível em: https://aclanthology.org/2022.acl-long.573/. Acesso em: 29 ago. 2024.

MÖKANDER, Jakob *et al.* The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other?. **Minds and Machines**, v. 32, n. 4, p. 751-758, 2022. Disponível em: https://link.springer.com/article/10.1007/s11023-022-09612-y. Acesso em: 29 ago. 2024.

MORGAN, Charles S. *et al.* One Step Closer to AI Regulations in Canada: The AIDA Companion Document. 28 de março de 2023. Disponível em: https://www.mccarthy.ca/en/insights/blogs/techlex/one-step-closer-ai-regulations-canada-aida-companion-document. Acesso em: 29 ago. 2024.

NEMITZ, Paul. Constitutional democracy and technology in the age of artificial intelligence. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2133, p. 20180089, 2018. Disponível em: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0089. Acesso em: 29 ago. 2024.

NIKOLINAKOS, Nikos Th. Ethical principles for trustworthy AI. In: **EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies-The AI Act**. Cham: Springer International Publishing, 2023. p. 101-166.

NILSSON, Nils J. **Artificial intelligence**: a new synthesis. Morgan Kaufmann, 1998.

NIST. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023. Disponível em: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. Acesso em: 29 ago. 2024.

NIST. National Institute of Standards and Technology. A Proposal for Identifying and Managing Bias within Artificial Intelligence. 2021. Disponível em: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf. Acesso em: 29 ago. 2024.

OECD. Organization for Economic Cooperation and Development. Evolving with innovation: The 2024 OECD AI Principles update. 2024c. Disponível em: https://oecd.ai/en/wonk/evolving-with-innovation-the-2024-oecd-ai-principles-update. Acesso em: 29 ago. 2024.

OECD. Organization for Economic Cooperation and Development. Generative AI. 2024a. disponível em: https://oecd.ai/en/genai/issues/benefits. Acesso em: 29 ago. 2024.

OECD. Organization for Economic Cooperation and Development. OECD AI Principles overview. 2024b. Disponível em: https://oecd.ai/en/ai-principles. Acesso em: 29 ago. 2024.

OECD. Organization for Economic Cooperation and Development. OECD Framework for the Classification of AI systems. **OECD Digital Economy Papers**, No. 323, 2022b. Disponível em: https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html. Acesso em: 29 ago. 2024

OECD. Organization for Economic Cooperation and Development. The OECD Framework for the Classification of AI systems. OCDE.AI Policy Observatory. 2022a. Disponível em: https://wp.oecd.ai/app/uploads/2022/02/Classification-2-pager-1.pdf. Acesso em: 21 ago. 2024.

OLIVEIRA NETO, Esclepíades de. Paradigma jurídico-político-econômico-administrativo do estado moderno e o exame dos indicadores de responsividade da regulação do uso da inteligência artificial no Brasil. 2023. Disponível em: http://www.realp.unb.br/jspui/bitstream/10482/47288/1/EsclepiadesDeOliveiraNeto_DISSERT.pdf. Acesso em: 29 ago. 2024.

OLORUNFEMI, Oluwabukunmi Latifat et al. Towards a conceptual framework for ethical AI development in IT systems. **Computer Science & IT Research Journal**, v. 5, n. 3, p. 616-627, 2024. Disponível em: https://www.fepbl.com/index.php/csitrj/article/view/910. Acesso em: 29 ago. 2024.

OLSEN, Henrik Palmer *et al.* **What's in the box?** The legal requirement of explainability in computationally aided decision-making in public administration. 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402974. Acesso em: 29 ago. 2024.

OPENPARLIAMENT.CA. BILL C-27 Digital Charter Implementation Act, 2022. Em comissão (Câmera), em 24 de abril de 2023. Disponível em: https://openparliament.ca/bills/44-1/C-27/?page=1. Acesso em: 29 ago. 2024.

PAGALLO, Ugo *et al.* New Technologies and Law: Global Insights on the Legal Impacts of Technology, Law as Meta-Technology and Techno Regulation. **New Technologies and Law**. Draft Version, 2015. Disponível em: https://lawschoolsgloballeague.com/wp-content/uploads/2021/11/New-Technologies-and-Law-Research-Group-Paper-20151.pdf. Acesso em: 29 ago. 2024.

PARDAU, Stuart L. The California Consumer Privacy Act: Towards a European-Style Privacy Regime in the United States. **J. Tech. L. & Pol'y 68 (2018-2019)**. Disponível em: https://scholarship.law.ufl.edu/jtlp/vol23/iss1/2/. Acesso em: 20 ago 2024.

PARINANDI, Srinivas *et al.* Investigating the politics and content of US State artificial intelligence legislation. **Business and Politics**, v. 26, n. 2, p. 240-262, 2024. Disponível em: https://www.cambridge.org/core/journals/business-and-politics/article/investigating-the-politics-and-content-of-us-state-artificial-intelligence-legislation/B603D28F79C554463680B22F3CA8F805. Acesso em: 29 ago. 2024.

PARLIAMENT OF CANADA. **BILL C-27**. 16 de junho de 2022. Disponível em: https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading. Acesso em: 29 ago. 2024.

PEIXOTO, Fabiano Hartmann; COUTINHO, Marina de Alencar Araripe. Inteligência Artificial e regulação. **Revista Em Tempo**, v. 19, n. 1, 2020.

PETERS, Dorian *et al.* Responsible AI — two frameworks for ethical design practice. **IEEE Transactions on Technology and Society**, v. 1, n. 1, p. 34-47, 2020. Disponível em: https://ieeexplore.ieee.org/abstract/document/9001063. Pagina 36. Acesso em: 29 ago. 2024.

PRAMOD, Akshara; NAICKER, Harsh Sankar; TYAGI, Amit Kumar. Machine learning and deep learning: Open issues and future research directions for the next 10 years. **Computational analysis and deep learning for medical care: Principles, methods, and applications**, p. 463-490, 2021.

PROTÁSIO; Aline Vieira Tomás; FARIA; Carolina Lemos de; HARTMANN PEIXOTO; Fabiano. **Projeto Simplificar 5.0**: Legal Design e Inteligência Artificial Ampliando o Acesso à justiça. RDP, Brasília, vol. 19, n. 102, 263-287.2022.

RAMOS, Gabriela; SQUICCIARINI, Mariagrazia; LAMM, Eleonora. Making AI ethical by design: The UNESCO perspective. **Computer**, v. 57, n. 2, p. 33-43, 2024. Disponível em: https://ieeexplore.ieee.org/abstract/document/10417786. Acesso em: 29 ago. 2024.

RENDA, Andrea *et al.* **Study to support an impact assessment of regulatory requirements for artificial intelligence in Europe**. European Commission: Brussels, Belgium, 2021. Disponível em: https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-COM-Impact-Assessment-3-21-April.pdf. Acesso em: 29 ago. 2024.

ROSTAMZADEH, Negar et al. Healthsheet: development of a transparency artifact for health datasets. In: **Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency**. 2022. p. 1943-1961.

SALATINO, Angelo A. *et al.* The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. **Data Intelligence**, v. 2, n. 3, p. 379-416, 2020.

SAMOILI, S. *et al.* **AI Watch.** Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020

SESTINO; Andrea; PRETE; Maria Irene; PIPER; Luigi; GUIDO; Gianluigi. Internet of Things and Big Data as enablers for business digitalization strategies. **Technovation**, v. 98, p. 102173, 2020. Disponível em: https://www.sciencedirect.com/science/article/pii/S0166497220300456. Acesso em: 29 ago. 2024.

SINGH, Richa; VATSA, Mayank; RATHA, Nalini. Trustworthy AI. In: **Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data**. 2021. p. 449-453.

SCHRECK, Martha. SCHRECK, Martin; CHARKOUDIAN, Stephen G. US **Artificial Intelligence Regulations**: Wach List for 2023. Goodwin, april 12, 2023. Disponível em: https://www.goodwinlaw.com/en/insights/publications/2023/04/04_12-us-artificial-intelligence-regulations. Acesso em: 29 ago. 2024.

SOUZA, Carlos Affonso; PERRONE, Christian; MAGRANI, Eduardo. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, D. et al (Coords.). **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

STAHL, Bernd Carsten *et al.* A systematic review of artificial intelligence impact assessments. **Artificial Intelligence Review**, v. 56, n. 11, p. 12799-12831, 2023. Disponível em: https://link.springer.com/article/10.1007/s10462-023-10420-8. Acesso em: 29 ago. 2024.

STERZ, Sarah *et al.* On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. In: **The 2024 ACM Conference on Fairness, Accountability, and Transparency.** 2024. p. 2495-2507. Disponível em: https://dl.acm.org/doi/abs/10.1145/3630106.3659051. Acesso em: 29 ago. 2024.

SWAMINATHAN, Akshay et al. Natural language processing system for rapid detection and intervention of mental health crisis chat messages. **NPJ Digital Medicine**, v. 6, n. 1, p. 213, 2023.

TAYE, Mohammad Mustafa. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers, v. 12, n. 5, p. 91, 2023.

TECHTARGET. Artificial Intelligence of Things (AIoT). July. 2023. Disponível: https://www.techtarget.com/iotagenda/definition/Artificial-Intelligence-of-Things-AIoT. Acesso em: 29 ago. 2024.

THE COUNCIL OF STATE GOVERNMENTS. Artificial Intelligence in the States: Emerging Legislation. 2023. Disponível em: https://www.csg.org/2023/12/06/artificial-intelligence-in-the-states-emerging-legislation/. Acesso em: 29 ago. 2024.

THE WHITE HOUSE. Accelerating America's Leadership in Artificial Intelligence. 2019b. Disponível em: https://trumpwhitehouse.archives.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/. Acesso em: 29 ago. 2024.

THE WHITE HOUSE. AI That Reflects American Values. 2020a. Disponível em: https://trumpwhitehouse.archives.gov/articles/ai-that-reflects-american-values/. Acesso em: 29 ago. 2024.

THE WHITE HOUSE. American Artificial Intelligence Iniciative: Year one annual report. 2020b. Disponível em: https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf. Acesso em: 29 ago. 2024.

THE WHITE HOUSE. Executive Order on AI. 2019a. Disponível em: https://trumpwhitehouse.archives.gov/ai/executive-order-ai/. Acesso em: 29 ago. 2024.

THE WHITE HOUSE. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 30 de outubro de 2023. Disponível em: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Acesso em: 29 ago. 2024.

THIEBES, Scott; LINS, Sebastian; SUNYAEV, Ali. Trustworthy artificial intelligence. **Electronic Markets**, v. 31, p. 447-464, 2021.

TIMMERS, Paul. AI challenging sovereignty and democracy. **Turkish Policy Quarterly**, v. 20, n. 4, p. 45-55, 2021. Disponivel em: file:///Users/cleideane/Downloads/ai-challenging-sovereignty-and-democracy_en_6556%20(2).pdf. Acesso em: 29 ago. 2024.

TZIMAS, Themistoklis. The Ontology of AI. In: Legal and ethical challenges of artificial intelligence from an international law perspective. **Springer Nature**, 2021. https://doi.org/10.1007/978-3-030-78585-7_3. Acesso em: 29 ago. 2024.

UNESCO. United Nations Educational, Scientific and Cultural Organization. Recommendation on the Ethics of Artificial Intelligence. 43 pages Adopted on 23 November 2021. UNESCO. 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em: 29 ago. 2024.

VAN NOORDEN, Richard; PERKEL, Jeffrey M. AI and science: what 1,600 researchers think. **Nature**, v. 621, n. 7980, p. 672-675, 2023.

WACHTER, Sandra. The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law. **Tul. L. Rev**., v. 97, p. 149, 2022.

WANG, Hong *et al.* Crash mitigation in motion planning for autonomous vehicles. **IEEE transactions on intelligent transportation systems**, v. 20, n. 9, p. 3313-3323, 2019b. Disponível em: https://ieeexplore.ieee.org/abstract/document/8617711. Acesso em: 29 ago. 2024.

WANG, Jun *et al.* Safety of autonomous vehicles. **Journal of advanced transportation**, v. 2020, p. 1-13, 2020. Disponivel em: https://www.hindawi.com/journals/jat/2020/8867757/. Acesso em: 29 ago. 2024.

WANG; Pei. On Defining Artificial Intelligence. **Journal of Artificial General Intelligence,** 10(2) 1-37, 2019a. Disponível em: https://sciendo.com/downloadpdf/journals/jagi/10/2/article-p1.pdf. Acesso em: 29 ago. 2024.

WHITE & CASE. **AI Watch**: Global regulatory tracker-Canada. 13 de maio de 2024. Disponível em: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-canada. Acesso em: 29 ago. 2024.

WHITE, Jeffrey. Autonomous Reboot: Kant, the categorical imperative, and contemporary challenges for machine ethicists. **AI & SOCIETY**, v. 37, n. 2, p. 661-673, 2022.

WORLD ECONOMIC FORUM. Unpacking AI Procurement in a Box: Insights from Implementation. White paper. may. 2022. Disponível em: https://www3.weforum.org/docs/WEF_Unpacking_AI_Procurement_in_a_Box_2022.pdf. Acesso em: 29 ago. 2024.

WRIGHT, Jasmine; VERITY, Andrej. **Artificial Intelligence Principles for vulnerable populations in Humanitarian Contexts**. DH Network, Jan 2020, p. 6.

ZUIDERVEEN BORGESIUS, Frederik. Discrimination, artificial intelligence, and algorithmic decision-making. **Council of Europe, Directorate General of Democracy**, p. 42, 2018. Disponível em: https://pure.uva.nl/ws/files/42473478/32226549.pdf. Acesso em: 29 ago. 2024.

ZHU, Xiaojin; GOLDBERG, Andrew B. **Introduction to semi-supervised learning.** Springer Nature, 2022.