UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE DIREITO
CAROLINA FIORINI RAMOS GIOVANINI

AVALIAÇÃO DE IMPACTO ALGORÍTMICO: uma abordagem contextual a partir do tripé de precaução, *accountability* e transparência

Juiz de Fora

2024

CAROLINA FIORINI RAMOS GIOVANINI

AVALIAÇÕES DE IMPACTO ALGORÍTMICO: uma abordagem contextual a partir do tripé de precaução, *accountability* e transparência

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Faculdade de Direito da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Mestra no Mestrado em Direito e Inovação, sob orientação do Prof. Dr. Sergio Marcos Carvalho de Ávila Negri.

Juiz de Fora

2024

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Giovanini, Carolina Fiorini Ramos.
Avaliação de impacto algorítmico : uma abordagem contextual a partir do tripé de precaução, accountability e transparência / Carolina Fiorini Ramos Giovanini. -- 2024.
127 f.

Orientador: Sergio Marcos Carvalho de Avila Negri
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Direito. Programa de Pós-Graduação em Direito, 2024.

1. Inteligência artificial. 2. Avaliação de impacto algorítmico. 3. Accountability. I. Negri, Sergio Marcos Carvalho de Avila, orient. II. Título.

FOLHA DE APROVAÇÃO

CAROLINA FIORINI RAMOS GIOVANINI

AVALIAÇÕES DE IMPACTO ALGORÍTMICO: uma abordagem contextual a partir do tripé de precaução, *accountability* e transparência

Dissertação apresentada ao Programa de Pós-graduação Stricto Sensu da Faculdade de Direito da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Mestre em Direito e Inovação. Na linha de pesquisa: "Direitos Humanos, Pessoa e Desenvolvimento: inovação e regulação jurídica no contexto do capitalismo globalizado", sob orientação do Prof. Dr. Sergio Marcos Carvalho de Ávila Negri

---

Orientador: Prof. Dr. Sergio Marcos Carvalho de Ávila Negri
Universidade Federal de Juiz de Fora

Profa. Dra. Joana de Souza Machado
Universidade Federal de Juiz de Fora

Profa. Dra. Clarissa Diniz Guedes
Universidade Federal de Juiz de Fora

Prof. Dr. Carlos Affonso Pereira de Souza
Universidade do Estado do Rio de Janeiro

PARECER DA BANCA

( ) APROVADO

( ) REPROVADO

Juiz de Fora, 04 de outubro de 2024

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE PÓS-GRADUAÇÃO

STRICTO SENSU

PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO

N° PPG: 2038

Formato da Defesa: ( ) presencial (x) virtual ( ) híbrido

Ata da sessão (x) pública ( ) privada referente à defesa da (x) dissertação ( ) tese intitulada "Avaliação de impacto algorítmico: uma abordagem contextual a partir do tripé de precaução, accountability e transparência", para fins de obtenção do titulo de (x)mestra(e) ( )doutor(a) em Direito, área de concentração Direito e Inovação, pelo(a) discente CAROLINA FIORINI RAMOS GIOVANINI (matrícula 102380158 - início do curso em 28/04/2022), sob orientação da Prof.(ª)Dr(ª) Sergio Marcos Carvalho de Ávila negri.

Aos 04 dias do mês de outubro do ano de 2024, às 15:30 horas, de forma não presencial, conforme Resolução nº 10/2022-CSPP e nº 16/2023-CSPP da Universidade Federal de Juiz de Fora (UFJF), reuniu-se a Banca examinadora da (X) dissertação ( )tese em epígrafe, aprovada pelo Colegiado do Programa de Pós- Graduação, conforme a seguinte composição:

Titulação Prof(a) Dr(a) / Dr(a)	Nome	Na qualidade de:
Prof(a) Dr(a) / Dr(a)	SÉRGIO MARCOS CARVAHO DE ÁVILA NEGRI	Orientador e presidente da banca
Prof(a) Dr(a) / Dr(a)	JOANA DE SOUZA MACHADO	Membro interno
Prof(a) Dr(a) / Dr(a)	CARLOS AFONSO PEREIRA DE SOUZA	Membro externo
Prof(a) Dr(a) / Dr(a)	CLARISSA DINIZ GUEDES	Membro interno
Prof(a) Dr(a) / Dr(a)	DANIEL BUCAR CERVAZIO	Suplente externo

*Na qualidade de (opções a serem escolhidas):

- Membro titular interno
- Membro titular externo
- Membro titular externo e Coorientador(a)
- Orientador(a) e Presidente da Banca
- Suplente interno
- Suplente externo
- Orientador(a)
- Coorientador(a)

*Obs: Conforme §2º do art. 54 do Regulamento Geral da Pós-graduação stricto sensu, aprovado pela Resolução CSPP/UFJF nº 28, de 7 de junho de 2023, “estando o(a) orientador(a) impedido(a) de compor a banca, a presidência deverá ser designada pelo Colegiado”.

AVALIAÇÃO DA BANCA EXAMINADORA

Tendo o(a) senhor(a) Presidente declarado aberta a sessão, mediante o prévio exame do referido trabalho por parte de cada membro da Banca, o(a) discente procedeu à apresentação de seu Trabalho de Conclusão de Curso de Pós-graduação Stricto sensu e foi submetido(a) à arguição pela Banca Examinadora que, em seguida, deliberou sobre o seguinte resultado:

(x) APROVADO

( ) REPROVADO, conforme parecer circunstanciado, registrado no campo Observações desta Ata e/ou em documento anexo, elaborado pela Banca Examinadora

Novo título da Dissertação/Tese (só preencher no caso de mudança de título):

Observações da Banca Examinadora caso haja necessidade de anotações gerais sobre a dissertação/tese e sobre a defesa, as quais a banca julgue pertinentes

Nada mais havendo a tratar, o(a) senhor(a) Presidente declarou encerrada a sessão de Defesa, sendo a presente Ata lavrada e assinada pelos(as) senhores(as) membros da Banca Examinadora e pelo(a) discente, atestando ciência do que nela consta.

INFORMAÇÕES

Para fazer jus ao título de mestre(a)/doutor(a), a versão final da dissertação/tese, considerada Aprovada, devidamente conferida pela Secretaria do Programa de Pós-graduação, deverá ser tramitada para a PROPP, em Processo de Homologação de Dissertação/Tese, dentro do prazo de 60 dias a partir da data da defesa. Após o envio dos exemplares definitivos, o processo deverá receber homologação e, então, ser encaminhado à CDARA.

Esta Ata de Defesa é um documento padronizado pela Pró-Reitoria de Pós-Graduação e Pesquisa. Observações excepcionais feitas pela Branca Examinadora poderão ser registradas no campo disponível acima ou em documento anexo, desde que assinadas pelo(a) Presidente(a).

Esta Ata de Defesa somente poderá ser utilizada como comprovante de titulação se apresentada junto á Certidão da Coordenadoria de Assuntos e Registros Acadêmicos da UFJF (CDARA) atestando que o processo de confecção e registro do diploma está em andamento.

Documento assinado eletronicamente por **Sergio Marcos Carvalho de Avila Negri**, **Professor(a)**, em 13/12/2024, às 18:26, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Carolina Fiorini Ramos Giovanini**, **Usuário Externo**, em 15/01/2025, às 23:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Joana de Souza Machado**, **Professor(a)**, em 15/01/2025, às 23:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Carlos Affonso Pereira de Souza**, **Usuário Externo**, em 17/01/2025, às 11:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Clarissa Diniz Guedes**, **Professor(a)**, em 17/01/2025, às 12:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2018625** e o código CRC **8009DE1E**.

AGRADECIMENTOS

Em primeiro lugar, agradeço àquelas que cuidaram de mim e, por vezes, colocaram minha vida à frente de suas próprias. À minha mãe, Cibelle, por não ter medido esforços para que eu concretizasse meus sonhos e por ter incentivado meu interesse nos estudos e na pesquisa. À minha avó, Nilma (*in memoriam*), por ter me feito sentir a criança mais amada do mundo. À madrinha Dorica (*in memoriam*) e à Maria, por terem sido um suporte em momentos difíceis.

Agradeço ao meu pai, Guilherme, por sempre apoiar e comemorar minhas escolhas e vitórias. Agradeço, também, ao meu irmão, Ian, que ocupa um lugar de muito carinho em meu coração.

Agradeço à Laura, por todo o companheirismo e amor que esteve presente em minha vida nos últimos anos e, acima de tudo, por entender meus medos e me ajudar a enfrentá-los. E, também, à Helena por ter sido um ponto de apoio, carinho e amor.

Agradeço ao Professor Sergio Negri, por ter me orientado desde a graduação e, principalmente, por ter nutrido em mim a paixão pela pesquisa e por ter me mostrado a importância da educação. Agradeço, também, aos professores com quem tive contato ao longo de toda a minha formação, especialmente no Colégio Técnico da Universidade Federal Rural do Rio de Janeiro (CTUR), por terem me apoiado na jornada de estudos.

Agradeço a sorte de ter amigos presentes em minha vida, em especial à Clarinha e ao Dodô, que me acompanham com carinho desde o ensino médio; aos meus amigos de faculdade, Larissa, Duda, Amanda, Letícia, Fernanda, Mariah, Marco Túlio, Daniel e Nathan — que, além de amigo, foi meu colega de pesquisa desde o início e pudemos produzir e crescer juntos.

Agradeço à Universidade Federal de Juiz de Fora, pelo ensino público, gratuito e de qualidade, e, ainda, por ter me acolhido e me formado como pessoa. Agradeço à Faculdade de Direito, ao Programa de Pós-Graduação em Direito e Inovação e a todos os professores e funcionários que possibilitam o funcionamento dessas instituições e a formação de tantas outras pessoas.

Por fim, a todos aqueles — colegas, amigos e familiares — que, de alguma forma, estiveram presentes ao longo desta jornada.

RESUMO

Sistemas de inteligência artificial (IA) estão cada vez mais presentes em atividades e processos decisórios que fazem parte do cotidiano de diversas pessoas. Com a promessa de gerar ganhos em produtividade e eficiência, tais sistemas também podem impactar o exercício de direitos humanos, especialmente diante de vieses e imprecisões. Diante desse cenário, emergem debates sobre instrumentos para identificação e mitigação de riscos no desenvolvimento ou implementação de sistemas de IA, como a Avaliação de Impacto Algorítmico (AIA). Nesse contexto, a presente dissertação tem como objetivo investigar quais são os contornos jurídicos da AIA e analisar seu papel enquanto ferramenta para proteção de direitos humanos. Trata-se de uma pesquisa de caráter exploratório, realizada mediante análise bibliográfica e documental. A partir do levantamento bibliográfico, entende-se que a AIA deve ser estruturada a partir de um tripé pautado em *accountability*, precaução e transparência. Para além desses parâmetros, o presente trabalho busca demonstrar que, para que seja efetivamente um instrumento de mitigação de riscos, é necessário adotar uma abordagem contextual no desenvolvimento da AIA, a qual reconheça que os impactos gerados por sistemas de IA podem variar consideravelmente a depender das particularidades sociais, econômicas e éticas do contexto de desenvolvimento e implementação, inclusive em relação às pessoas e grupos afetados.

Palavras-chave: inteligência artificial; Avaliação de Impacto Algorítmico; *accountability*.

ABSTRACT

Artificial intelligence (AI) systems are increasingly present in activities and decision-making processes that are part of people's daily lives. With the promise of generating gains in productivity and efficiency, such systems can also impact the exercise of human rights, especially in the face of biases and inaccuracies. Against this backdrop, debates are emerging about tools for identifying and mitigating risks in the development or implementation of AI systems, such as Algorithmic Impact Assessment (AIA). In this context, this dissertation aims to investigate the legal outlines of AIA and analyze its role as a tool for protecting human rights. This is an exploratory study, carried out through bibliographic and documentary analysis. Based on the bibliographic survey, it is understood that the AIA should be structured on the basis of a tripod based on accountability, precaution and transparency. In addition to these parameters, this paper seeks to demonstrate that, in order for it to be an effective risk mitigation instrument, it is necessary to adopt a contextual approach in the development of AIA, which recognizes that the impacts generated by AI systems can vary considerably depending on the social, economic and ethical particularities of the development and implementation context, including in relation to the people and groups affected.

Keywords: artificial intelligence; algorithmic impact assessment; accountability.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

AIA	Avaliação de Impacto Algorítmico
ANPD	Autoridade Nacional de Proteção de Dados
CONAMA	Conselho Nacional do Meio Ambiente
EIA	Estudo de Impacto Ambiental
FRAIA	Fundamental Rights and Algorithms Impact Assessment (Avaliação de Impacto Algorítmico para Direitos Fundamentais, em tradução livre)
FRIA	Fundamental Rights Impact Assessment (Avaliação de Impacto para Direitos Fundamentais, em tradução livre)
HRIA	Human Rights Impact Assessments (Avaliação de Impacto para Direitos Humanos, em tradução livre)
IA	Inteligência artificial
ICO	Information Commissioner's Office
LGPD	Lei Geral de Proteção de Dados
NIST	National Institute of Standards and Technology
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
RIPD	Relatório de Impacto à Proteção de Dados
RIMA	Relatório de Impacto ao Meio Ambiente
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

SUMÁRIO

1 INTRODUÇÃO

Sistemas de inteligência artificial (IA) estão cada vez mais presentes em atividades do nosso cotidiano. Em âmbito doméstico ou profissional, frequentemente as pessoas são atendidas por *chatbots*, consomem conteúdo recomendado e, por vezes, produzido por sistemas de IA, são avaliadas e classificadas em processos seletivos, solicitações de empréstimos e amplamente monitoradas pela tecnologia no dia a dia.

Fato é que exemplos de aplicação de sistemas de IA não se limitam a estes e diversas formas de utilização já fazem parte do dia a dia da sociedade e moldam interações e decisões diárias. O avanço e a aplicação de sistemas de IA geraram impactos que transcendem a compreensão puramente quantitativa, resultando em uma transformação na forma como as pessoas interagem com a tecnologia. Dessa maneira, os computadores deixaram de ser percebidos apenas como ferramentas para cálculos, sistematizações ou classificações, passando a ser comparados, em certa medida, a ações essencialmente humanas (Doneda *et al*., 2018).

Diante desse cenário, surge a necessidade de compreender não apenas os benefícios dessa tecnologia, mas também os desafios éticos e jurídicos que emergem de sua aplicação cada vez mais difundida.

1.1 APRESENTAÇÃO DO TEMA, PROBLEMA DE PESQUISA E OBJETIVOS

A inteligência artificial é uma tecnologia em desenvolvimento; seus novos usos e aplicações são diariamente testados. Diante dessa popularização, os impactos éticos, jurídicos e sociais decorrentes da aplicação de sistemas de IA se tornam perceptíveis. Para além de seu campo de aplicação, é importante compreender que a inteligência artificial, enquanto tecnologia emergente, é mais do que mais um campo de estudo técnico, estando diretamente interligada a um conjunto mais amplo de estruturas políticas e sociais, abrangendo instituições, dimensões políticas e aspectos culturais (Crawford, 2021).

Doneda, Mendes, Souza e An (2018) apontam que grande parte das questões que atualmente cercam o desenvolvimento e a implementação de sistemas de IA teve suas bases teóricas formuladas nas discussões iniciais sobre os impactos da automação e da inteligência artificial. No entanto, embora as bases teóricas dos componentes computacionais sejam debatidas desde o início da construção deste campo de estudo, sua efetiva utilidade e aplicabilidade são mais recentes, especialmente em razão de duas

limitações: primeiro, as restrições na capacidade de processamento dos computadores; segundo, as limitações das abordagens iniciais na implementação de sistemas de inteligência artificial (Doneda *et al*., 2018).

Com o progresso contínuo da capacidade computacional, as novas abordagens e técnicas empregadas em sistemas de IA mais recentes se mostraram decisivas, especialmente ao abandonar o uso de um conjunto de regras previamente definidas em favor de algoritmos que "aprendem" por meio da observação e análise de grandes volumes de dados. Desse modo, a maior disponibilidade de recursos computacionais e de informação, resultantes tanto do avanço tecnológico quanto do paradigma do *big data*, permitiu o surgimento de sistemas de inteligência artificial que não dependem exclusivamente de modelos baseados em regras predefinidas, mas que utilizam grandes quantidades de dados para fundamentar suas decisões e moldar seus padrões decisionais (Doneda *et al*., 2018).

No âmbito dos impactos sociais, a bibliografia existente na área de IA destaca que os modelos matemáticos responsáveis por alimentar a economia de dados foram baseados em escolhas feitas por seres humanos falíveis e, embora algumas dessas escolhas tenham sido feitas com as melhores intenções, foram codificados os preconceitos, a incompreensão e a parcialidade humana em sistemas que trazem impactos diretos para a sociedade (O'Neil, 2021).

Na mesma direção, Broussard (2023) aponta que não se deve tratar os problemas trazidos pela tecnologia como "falhas" ("*bugs*" ou "*glitches*", na língua inglesa). Para a autora, chamar algo de falha significa entender que se trata de um problema temporário, algo inesperado, mas inconsequente. No entanto, Broussard (2023) entende que os preconceitos embutidos na tecnologia são mais do que meras falhas, na verdade, são problemas estruturais e não podem ser resolvidos com uma rápida atualização do código.

Quando se rotula um problema tecnológico como uma "falha", implicitamente se minimiza a gravidade e a extensão do problema, sugerindo que ele é uma anomalia em um sistema que, de outra forma, funcionaria perfeitamente. Uma visão simplista do problema ignora a complexidade inerente à interação entre tecnologia e sociedade e a possibilidade de que desigualdades e preconceitos já existentes sejam incorporados de forma profunda e persistente em estruturas tecnológicas.

Diante desse contexto, torna-se evidente que os sistemas de inteligência artificial influenciam e moldam relações sociais e percepções do mundo (Crawford, 2021). Por isso, já não se discute se a IA terá um grande impacto na sociedade; o debate atual se

concentra em avaliar em que medida esse impacto será positivo ou negativo, para quem, de que maneira, em quais contextos e em que prazo (Floridi *et al.*, 2018).

Quando se trata do desenvolvimento e da aplicação de sistemas de IA, é importante compreender que existem diversas partes interessadas — como trabalhadores, pacientes, crianças e adolescentes, consumidores, sociedade civil como um todo, governos, investidores e empresas —, e esses agentes podem ser afetados de formas distintas. Essas diferenças nos ganhos e na vulnerabilidade aos impactos da IA surgem não somente dentro dos países, mas também entre países e partes do mundo (Coeckelbergh, 2020). Isso significa que países com economias desenvolvidas, infraestrutura digital robusta e forte governança podem colher os frutos da IA de maneira mais segura e controlada, ao passo que outros países podem enfrentar desafios adicionais, que incluem ausência de regulamentação adequada, desigualdade no acesso a tecnologias e um impacto desproporcional sobre populações já vulneráveis.

As disparidades tecnológicas entre o Norte e o Sul globais, por exemplo, podem exacerbar as desigualdades existentes, criando um cenário em que os benefícios da IA são distribuídos de forma desigual, e as vulnerabilidades são ampliadas em contextos com menos recursos e proteção legal. A compreensão dessas nuances é fundamental para o desenvolvimento de políticas que garantam que a IA seja utilizada de maneira ética e responsável, minimizando os riscos e maximizando os benefícios para todas as partes interessadas, inclusive em relação à promoção de práticas de inovação responsável e proteção de direitos humanos.

Da mesma forma, a maneira como a tecnologia é incorporada em diferentes contextos sociais e culturais reflete e reforça normas, valores e poderes existentes. A inteligência artificial não se refere apenas à tecnologia, mas também ao que os humanos fazem com ela, como a usam, a percebem, a experimentam, e como a incorporam em ambientes sociotécnicos mais amplos, de modo que é necessário incluir uma perspectiva histórica e sociocultural no debate (Coeckelbergh, 2020).

Assim como a lei, a tecnologia também atua como intermediária nas relações sociais entre seres humanos, sendo possível sustentar que o debate sobre regulação se relaciona, na verdade, com a discussão sobre (i) como as pessoas interagem com novas invenções; e (ii) como as pessoas interagem com outras pessoas usando essas novas invenções (Balkin, 2015). Cada avanço tecnológico, desde a imprensa e a rede de energia elétrica até a rede mundial de computadores e os sistemas de inteligência artificial, tem

transformado as dinâmicas sociais e criado formas de interação entre humanos e entre humanos e máquinas.

Nesse contexto, em decorrência do desenvolvimento e da crescente utilização de sistemas de IA, diversos países já iniciaram debates regulatórios, bem como incorporação de diretrizes éticas e padrões de responsabilidade. Além das iniciativas governamentais, organizações internacionais, como a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) e a Organização para Cooperação e Desenvolvimento Econômico (OCDE), também têm desempenhado um papel relevante na formulação de diretrizes éticas para a IA.

Nos Estados Unidos, por exemplo, observa-se regulamentações setoriais específicas e abordagens de co-regulamentação, isto é, arranjos regulatórios nos quais as agências federais já existentes estabelecem diretrizes a serem implementadas. Nessa direção, em novembro de 2023, a Presidência dos Estados Unidos divulgou uma Ordem Executiva (*Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, no original) que prevê princípios e ações prioritárias traduzidos em diretrizes gerais e mandamentos principiológicos, permitindo flexibilidade e adaptabilidade diante da constante evolução do campo da IA. Além disso, a norma prevê a elaboração de guias de boas práticas e diretrizes setoriais por parte das agências federais, como o *National Institute of Standards and Technology* (NIST), que, de modo geral, serão responsáveis pela elaboração de padrões, ferramentas e testes para desenvolvimento de sistemas de IA seguros e confiáveis.

A União Europeia, por sua vez, possui uma regulamentação abrangente sobre inteligência artificial, aprovada pelo Parlamento Europeu e pelo Conselho Europeu em 2024. A proposta, originalmente intitulada *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence And Amending Certain Union Legislative Act* é atualmente conhecida como *Artificial Intelligence Act* ou *AI Act* (Regulamento EU 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial).

O objetivo da norma aprovada é melhorar o funcionamento do mercado interno europeu mediante o estabelecimento de um quadro jurídico uniforme para o desenvolvimento, a colocação no mercado ou em serviço e a utilização de sistemas de IA na União Europeia, de modo a assegurar a proteção da saúde, da segurança e dos direitos fundamentais. O regulamento europeu é baseado em uma abordagem de definição de

práticas proibidas e de classificação de riscos dos sistemas de IA. Com base nessa classificação de riscos, as obrigações regulatórias são definidas de acordo com o nível de risco associado ao sistema. Em síntese, trata-se de uma estrutura regulatória na qual o nível de intervenção legislativa é modulável conforme o risco que um determinado sistema de IA representa.

No Brasil, por sua vez, foi estabelecida a Comissão de Juristas encarregada de subsidiar a elaboração de um substitutivo sobre Inteligência Artificial no país (CJSUBIA). Após análise de proposições legislativas relacionadas ao tema, incluindo o Projeto de Lei (PL) n.º 5.051, de 2019, de autoria do Senador Styvenson Valentim, o PL n.º 21, de 2020, do Deputado Federal Eduardo Bismarck (aprovado pela Câmara dos Deputados e encaminhado ao Senado Federal), e o PL n.º 872, de 2021, do Senador Veneziano Vital do Rêgo, a comissão realizou audiências públicas e recebeu contribuições sobre o assunto. Em dezembro de 2022, o relatório final e um anteprojeto de lei para a regulamentação da inteligência artificial foram publicados. Posteriormente, o Senador Rodrigo Pacheco apresentou o Projeto de Lei n.º 2.338, de 2023, que versa sobre o uso da inteligência artificial no Brasil, baseado na proposta de anteprojeto elaborada pela CJSUBIA.

Atualmente, os Projetos de Lei mencionados estão sendo discutidos no âmbito da Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA), instalada no Senado Federal em agosto de 2023. Em abril de 2024, houve apresentação de relatório preliminar de novo texto para o Projeto de Lei n.º 2.338, de 2023 e, em junho de 2024, apresentação do relatório final, com proposta de substitutivo. Posteriormente, após apresentação de emendas, houve complementação do voto por parte do relator do Projeto de Lei.

No momento em que a presente pesquisa é concluída, o relatório final e o voto do relator do Projeto de Lei foram apresentados no âmbito da CTIA, mas não houve uma votação acerca do texto e as atividades da Comissão foram prorrogadas. Sendo assim, este trabalho considera o debate regulatório brasileiro até a apresentação da Complementação de Voto, entregue pelo relator do Projeto de Lei nº. 2.338/2023 à CTIA em julho de 2024.

Em relação ao seu conteúdo, o Projeto de Lei n.º 2.338 busca estabelecer normas para a utilização de sistemas de inteligência artificial no Brasil e incorpora, em certa medida, a abordagem regulatória adotada pela União Europeia. Conforme será abordado neste trabalho, o texto da proposta brasileira também adota uma sistematização de classificação de riscos dos sistemas de IA, incluindo sistemas de risco excessivo,

considerados inaceitáveis e, portanto, banidos, e sistemas de alto risco, que atraem uma maior carga de obrigações regulatórias.

Para além das experiências nos Estados Unidos, na União Europeia e no Brasil, nota-se que diversos países também iniciaram movimentos regulatórios, inclusive no âmbito da estruturação de estratégias nacionais para desenvolvimento do campo da inteligência artificial, como é o caso de Singapura (Singapura, 2023) e Reino Unido (Reino Unido, 2023) que possuem políticas nacionais que buscam o fortalecimento de um ecossistema pró-inovação.

Em 2023, Singapura lançou uma política nacional abrangente voltada para a IA, com o objetivo de fortalecer seu ecossistema de inovação, o que inclui a promoção de parcerias público-privadas, investimentos em pesquisa e desenvolvimento, e a criação de diretrizes éticas para o uso de IA (Singapura, 2023). Também em 2023, o Reino Unido desenvolveu uma estratégia nacional para a IA, visando consolidar sua posição como líder global no campo. Entre as medidas adotadas, destacam-se investimentos substanciais em pesquisa, o incentivo à colaboração entre universidades, empresas e governo, e a educação e capacitação da força de trabalho (Reino Unido, 2023).

Evidentemente, além de Singapura e Reino Unido, outras nações também estão desenvolvendo suas próprias estratégias de IA. O Brasil, por sua vez, anunciou em julho de 2024 uma proposta de Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028, que objetiva tornar o Brasil um modelo global de eficiência e inovação no uso de IA no setor público e apresenta uma série de ações para desenvolvimento da infraestrutura nacional de dados (Brasil, 2024).

Os movimentos regulatórios e estratégias nacionais refletem um reconhecimento global da importância da inteligência artificial como motor de inovação e crescimento econômico. Desse modo, a discussão sobre a estruturação de ambientes regulatórios abrange não apenas o cenário de competição global, mas também a garantia de que os benefícios do desenvolvimento tecnológico sejam amplamente distribuídos e os riscos potenciais sejam mitigados.

Tendo em vista a emergência de debates regulatórios, Mantelero (2022) apresenta uma classificação de três possíveis abordagens regulatórias para fundamentar a regulamentação da IA na promoção de direitos humanos: (i) abordagem baseada em princípios, que inclui princípios orientadores derivados de instrumentos internacionais de direitos humanos — vinculantes e não vinculantes — e fornecem uma estrutura abrangente para a compreensão regulatória da IA; (ii) abordagem baseada nos impactos

da IA sobre os direitos individuais e de sua proteção por meio de uma avaliação de risco, com base em direitos; e (iii) abordagem baseada no risco, com ênfase no gerenciamento de sistemas de alto risco, concentrando-se na segurança do produto e na avaliação da conformidade.

Apesar de apresentarem diferenças, os três modelos compartilham uma preocupação central com a proteção dos direitos humanos, reconhecida como uma questão fundamental. No entanto, é importante notar que a ênfase no gerenciamento de riscos nem sempre é acompanhada por modelos eficazes de avaliação de impacto para os direitos humanos (Mantelero, 2022), o que pode resultar em lacunas na compreensão dos reais impactos gerados por determinado sistema de IA.

A análise dos modelos regulatórios descritos por Mantelero (2022) revela a existência de movimentos concentrados na regulação ética, por exemplo, por meio de instrumentos baseados em princípios, padrões e códigos de conduta que, embora contenham diretrizes normativas, não têm caráter de obrigatoriedade legal, sendo classificados como instrumentos de *soft law*. O termo *soft law* (direito flexível ou suave, em tradução livre) se refere a um conjunto de normas que não são obrigatórias, ou que possuem um poder vinculativo inferior, servindo como uma orientação, mas sem punições diretas para quem não a segue. Com frequência, essa abordagem é adotada em um contexto de autorregulação, no qual os próprios agentes econômicos elaboram e aplicam tais regras.

Por outro lado, é igualmente perceptível a existência de propostas que adotam abordagens regulatórias normativas e vinculantes — denominadas instrumentos de *hard law* —, manifestadas por meio da elaboração de leis específicas e da instituição de agências regulatórias com competências de regulamentação e fiscalização do tema, utilizando mecanismos regulatórios vinculantes e dotados de poder coercitivo, capazes de impor obrigações diretas ou sanções aos agentes regulados.

A partir dessas discussões, emergem debates sobre as obrigações, os mecanismos e as ferramentas necessárias para o gerenciamento de riscos associados a sistemas de IA, bem como para o fomento à transparência, responsabilidade e controle social. Nesse contexto, o presente trabalho destaca o instrumento denominado Avaliação de Impacto Algorítmico (AIA), que pode ser compreendido como uma ferramenta que visa identificar e avaliar os riscos e efeitos de um sistema de IA sobre interesses socialmente relevantes, com particular atenção a potenciais externalidades negativas que esses sistemas possam gerar.

Conforme será abordado por este trabalho, o desenvolvimento da AIA está relacionado ao entendimento de características, limites e capacidades de determinado sistema de IA, bem como à construção de confiança entre as partes interessadas e ao registro de aspectos do funcionamento do sistema para fins de *accountability*. De modo geral, este instrumento tem por objetivo analisar o impacto de sistemas de IA para pessoas e grupos afetados por meio da identificação de potenciais riscos e desenvolvimento de estratégias e medidas mitigatórias, bem como estabelecimento de mecanismos de monitoramento e avaliação contínuos.

Ademais, as avaliações de impacto são ferramentas capazes de auxiliar no processo de tomada de decisão informada, inclusive sobre a pertinência de implementar ou não determinado sistema de IA em uma atividade; assim, as avaliações de impacto se traduzem em mecanismos de proteção de interesses sociais relacionados, pois oferecem uma análise abrangente dos riscos e benefícios associados ao sistema de IA (Koshiyama; Engin, 2019).

Diante desse contexto, o presente trabalho tem como objetivo analisar os fundamentos teóricos e jurídicos da Avaliação de Impacto Algorítmico (AIA) e investigar sua efetividade como instrumento de promoção e proteção dos direitos humanos, bem como de mitigação de riscos em diferentes contextos socioeconômicos. A escolha deste tema decorre da necessidade de examinar o potencial agravamento das violações aos direitos humanos decorrentes do desenvolvimento e uso de novas tecnologias, particularmente dos sistemas de IA. Desse modo, além de analisar a AIA como uma ferramenta de governança, este estudo visa avaliar a efetividade desse instrumento na proteção e promoção de direitos humanos em variados contextos de desenvolvimento e implementação de sistemas de IA.

Ressalta-se que este trabalho se concentra, estritamente, em aspectos jurídicos, e não avalia elementos técnicos, como programação, ciência da computação, segurança da informação, entre outros. Essa delimitação ocorre devido à formação acadêmica da pesquisadora, que se restringe ao campo jurídico. Consequentemente, considerações de natureza técnica estão fora do escopo definido por esta pesquisa.

Para desenvolvimento da pesquisa, o trabalho explora, em primeiro lugar, as definições e propostas regulatórias para conceituação do termo “inteligência artificial”. A expressão foi popularizada por John McCarthy, em 1956, na conferência *Dartmouth Summer Research Project on Artificial Intelligence*, porém, atualmente não há consenso sobre uma definição para o termo. Diante desse cenário, é essencial explorar as definições

em debate, bem como questionar se, de fato, é necessário que exista uma definição específica e abrangente para o termo.

A partir do conceito adotado, será delimitado o escopo de aplicabilidade dos instrumentos regulatórios, sejam eles normas previstas em legislações nacionais, regras de governança estabelecidas por entidades privadas ou diretrizes orientativas de âmbito global. Nesse sentido, abordar o conceito de inteligência artificial é relevante para o presente trabalho porque o vocabulário utilizado para denominação e conceituação da tecnologia pode ser responsável por restringir ou ampliar o escopo de aplicabilidade de eventual regulamentação e, consequentemente, influenciar a necessidade de desenvolvimento (ou não) da AIA em uma situação concreta.

Em seguida, o trabalho investiga a fundamentação teórica da AIA, com base nas contribuições da doutrina, examinando interpretações sobre a estrutura, o escopo de aplicação desse instrumento e seus respectivos objetivos. Na sequência, tendo em vista as perspectivas regulatórias em relação à inteligência artificial, analisa-se o surgimento da AIA e de outras ferramentas semelhantes como instrumentos regulamentados na proposta legislativa atualmente em discussão no Brasil e na regulamentação aprovada pela União Europeia.

A escolha deste recorte que coloca ênfase nas discussões em âmbito brasileiro e europeu se dá em razão de dois fatores: (i) o protagonismo da União Europeia na proposição de regulações que envolvem o uso e desenvolvimento de tecnologias e serviços digitais; e (ii) o fato de que o texto aprovado na União Europeia e o texto em tramitação no Brasil se aproximam ao trazer regras abrangentes e aplicáveis para todos os tipos de sistemas de inteligência artificial, afastando-se de modelos regulatórios setoriais ou de usos específicos adotados em outros países.

Posteriormente, a pesquisa propõe-se a analisar a AIA com base em um conjunto conceitual formado pelos seguintes fundamentos: *accountability*, precaução e transparência. Procura-se demonstrar que cada um desses pilares desempenha um papel relevante na compreensão e no aprimoramento das práticas de avaliação de impacto, integrando-se como mecanismos de gestão de riscos, prestação de contas e controle social.

A compreensão da AIA a partir desta base tríplice faz com que sua lógica de elaboração considere que o agente responsável pelo desenvolvimento ou implantação de um sistema de IA tenha que agir preventivamente e justificar suas escolhas, verificar se estas impactam (positiva e/ou negativamente) direitos de pessoas e grupos afetados e,

ainda, apresentar medidas mitigatórias que podem ser julgadas e questionadas por um fórum (interno ou externo).

Consecutivamente, visa-se demonstrar, por meio da análise de duas situações envolvendo o desenvolvimento de avaliações de impacto, a importância de diferenciar esse instrumento dos métodos tradicionais de gestão de riscos. Estes, por vezes, concentram-se nos riscos para a própria organização em detrimento de uma análise contextual mais ampla, capaz de revelar os impactos sobre os direitos de pessoas e grupos afetados pelo desenvolvimento ou implementação de sistemas de IA. Argumenta-se, portanto, que é fundamental que a AIA seja desenvolvida a partir de uma abordagem contextual, considerando os impactos concretos de sistemas de IA sobre direitos humanos.

Por fim, retoma-se o conteúdo apresentado de forma objetiva e apresenta-se a conclusão do trabalho.

1.2 ASPECTOS TEÓRICO-METODOLÓGICOS

A presente dissertação busca estabelecer uma análise qualitativa sobre os contornos jurídicos da AIA enquanto instrumento de governança para gerenciamento de riscos associados ao desenvolvimento e implantação de sistemas de IA. Nesse sentido, o trabalho é desenvolvido entorno do seguinte problema de pesquisa: em que medida a Avaliação de Impacto Algorítmico (AIA) pode ser considerada um instrumento eficaz na mitigação de riscos para direitos humanos, associados ao desenvolvimento e uso de sistemas de IA em diferentes contextos socioeconômicos?

Diante da pergunta de pesquisa elaborada, este trabalho procura verificar a incidência das seguintes hipóteses: (H1) a AIA pode contribuir significativamente para a identificação e mitigação de riscos, inclusive mediante reconhecimento de que impactos gerados por sistemas de IA podem ser diferenciados, a depender de contextos sociais e econômicos; e (H2) implementação da AIA pode ser percebida como uma formalidade burocrática sem impacto substancial na mitigação de riscos para direitos humanos, especialmente em situações nas quais o contexto de desenvolvimento ou aplicação do sistema de IA não é efetivamente analisado e considerado.

Para desenvolvimento do problema de pesquisa, o presente trabalho também enfrentará as seguintes questões: (i) quais parâmetros podem ser utilizados para determinar quando a AIA deve ou não ser realizada? (ii) qual é a metodologia para

desenvolvimento da AIA? (iii) qual conteúdo a AIA deveria abranger? (iv) como a AIA pode ser realizada considerando aspectos contextuais relacionados ao desenvolvimento e/ou ao uso de sistemas de IA?

A metodologia adotada é baseada em uma abordagem exploratória, considerando o incipiente contexto regulatório ao qual os sistemas de inteligência artificial estão submetidos. Desse modo, busca-se uma proximidade com o problema de pesquisa desenvolvido, com intuito de formulação de hipóteses mais robustas posteriormente. Nessa direção, Babbie (2013) esclarece que a abordagem exploratória ocorre tipicamente quando a pesquisa se concentra em examinar um novo interesse ou quando o objeto de estudo é relativamente novo e pouco estudado.

A abordagem exploratória permite melhor compreensão do contexto geral do tema, possibilitando a identificação de lacunas e áreas não abordadas em relação ao assunto. Nesse sentido, a escolha pelo desenvolvimento da presente pesquisa a partir de uma abordagem exploratória se deve, em primeiro lugar, em razão da novidade e constante atualização do tema e, consequentemente, da ausência de um corpo substancial de conhecimento consolidado. Por isso, a aplicação da abordagem exploratória pode auxiliar a mapear o contexto e identificar lacunas e oportunidades de pesquisas mais robustas.

Além disso, a Avaliação de Impacto Algorítmico é um tema que envolve uma série de campos de estudos que estão interligados, como ética, sociedade, direito e tecnologia, de modo que o desenvolvimento da pesquisa a partir de uma abordagem exploratória permite entendimento geral da complexidade do tema e de suas implicações éticas, técnicas, sociais e jurídicas. Por fim, a escolha pela pesquisa exploratória também se deve ao fato de que, no momento de desenvolvimento deste trabalho, o Brasil não conta com regulamentação específica sobre inteligência artificial, mas tão somente com propostas sendo debatidas em âmbito legislativo.

Em relação a técnicas de pesquisa, o presente trabalho adota a revisão sistemática de bibliografia como a principal técnica, uma vez que a pesquisa científica é iniciada por meio da pesquisa bibliográfica, em que o pesquisador busca obras já publicadas relevantes para conhecer e analisar o tema problema da pesquisa a ser realizada (Cervo; Bervian, 2022). Nesse sentido, a pesquisa bibliográfica é realizada a partir do levantamento de referências teóricas já analisadas e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de websites, possibilitando ao pesquisador conhecer o que já se estudou sobre o assunto (Fonseca, 2022).

Para Severino (2007), a pesquisa bibliográfica possibilita o uso de dados de categorias teóricas já trabalhadas por outros pesquisadores e devidamente registrados, de modo que os textos se tornam fontes dos temas a serem pesquisados. Na mesma direção, Amaral (2007) aponta que esta técnica consiste no levantamento, seleção, fichamento e arquivamento de informações relacionadas à pesquisa.

Esta pesquisa adota a técnica de revisão bibliográfica com intuito de mapear o que já foi estudado e publicado sobre Avaliações de Impacto Algorítmico, bem como identificar tendências recentes e desenvolvimentos neste eixo de estudo. Além disso, com essa estratégia metodológica, visa-se demonstrar que o caminho aberto por autores que já analisaram o tema não considera amplamente o fato de que o gerenciamento de riscos deve ser adaptado em razão de circunstâncias contextuais que fazem com que a própria exposição ao risco varie consideravelmente em razão de particularidades sociais, econômicas e éticas do contexto de desenvolvimento e implementação de sistemas de IA, inclusive em relação às pessoas e grupos afetados.

Conforme apontam Lakatos e Marconi (2003), é importante ressaltar que a pesquisa bibliográfica não é mera repetição do que já foi dito ou escrito sobre certo assunto, mas propicia o exame de um tema sob novo enfoque ou abordagem, possibilitando que o pesquisador chegue a conclusões inovadoras. Desse modo, a revisão de literatura é realizada neste trabalho para fins de levantamento de subsídios sobre como o tema foi tratado na literatura científica, porém, para além do levantamento de fontes de pesquisa, realiza-se análise crítica do conteúdo levantado com o objetivo de atualizar, desenvolver o conhecimento e contribuir com a realização da pesquisa (Ruiz, 2009).

Além disso, a pesquisa adota a técnica de análise documental, o que inclui o estudo de documentos técnicos, relatórios publicados por organizações do setor privado, do setor público e do terceiro setor, bem como propostas regulatórias debatidas no contexto brasileiro e em outros países. Conforme apontam Sá-Silva, Almeida e Guindani (2009), a análise documental pode ser compreendida como um procedimento que se utiliza de métodos e técnicas para a apreensão, compreensão e análise de documentos dos mais variados tipos.

O trabalho adota a técnica de análise documental com o objetivo de possibilitar que esta pesquisa, mediante contato com documentos já produzidos, possa produzir novos conhecimentos a partir do conhecimento já existente. Para tanto, cabe destacar que se compreende por “documento”, conforme aponta Godoy (1995) os materiais escritos (como jornais, revistas, diários, obras literárias, científicas e técnicas, cartas,

memorandos, relatórios), as estatísticas (que produzem um registro ordenado e regular de aspectos da vida de determinada sociedade) e os elementos iconográficos (como sinais, grafismos, imagens, fotografias, filmes). A análise documental realizada para fins da presente pesquisa se concentra em documentos escritos e essencialmente jurídicos, como legislações, projetos de lei e relatórios produzidos no âmbito do debate regulatório sobre inteligência artificial no Brasil.

Em relação ao desenvolvimento da pesquisa e aplicação da técnica de análise documental, destaca-se a influência do debate regulatório, especialmente no Brasil e na União Europeia. As discussões regulatórias e as alterações significativas nas propostas em tramitação trouxeram desafios para a análise documental, uma vez que as informações e temas em discussão se atualizavam constantemente. No entanto, apesar das dificuldades metodológicas enfrentadas, o processo permitiu que a pesquisa se enriquecesse consideravelmente com a dinâmica do ambiente regulatório, que proporcionou uma compreensão mais profunda das diferentes perspectivas e interesses envolvidos, enriquecendo assim a análise documental com uma visão atualizada e contextualmente informada.

Por fim, o presente trabalho analisa duas situações de aplicação concreta da AIA, de modo a verificar a metodologia utilizada para desenvolvimento da avaliação, os elementos considerados pela análise para fins de identificação de riscos, o resultado obtido e as consequências da avaliação dentro do contexto particular de desenvolvimento do instrumento.

Feitas as considerações metodológicas, serão feitas agora as considerações teóricas que nortearam o desenvolvimento deste trabalho. Nesse sentido, três lentes teóricas foram fundamentais para o desenvolvimento desta pesquisa, contribuindo para que o objeto fosse melhor investigado.

Em primeiro lugar, o conceito de vulnerabilidade adotado pelo trabalho refere-se a uma dimensão mais ampla, que vai além dos conceitos técnicos presentes em legislações, como o Código de Defesa do Consumidor. Neste trabalho, a vulnerabilidade está intrinsecamente ligada às estruturas de poder e dinâmicas sociais que permeiam a sociedade, envolvendo desigualdades que se manifestam a partir de fatores como raça, gênero, classe e outras interseccionalidades que moldam as experiências dos indivíduos na sociedade (Machado; Negri; Giovanini, 2023).

Em segundo lugar, o presente trabalho guia-se pelo entendimento de que a avaliação de impacto de sistemas de IA deve englobar a avaliação ampla de direitos

humanos, indo além do direito à privacidade e do direito à proteção de dados pessoais, que são tradicionalmente avaliados em circunstâncias relacionadas ao desenvolvimento e uso de tecnologias emergentes[1]. Referido ponto de partida é profundamente abordado por Alessandro Mantelero (2022), que, embora reconheça que os dados estão no centro do funcionamento dos sistemas de IA, questiona se a legislação de privacidade e proteção de dados também poderia fornecer uma estrutura regulatória eficaz para sistemas de IA.

Nessa direção, Mantelero (2022) aponta que o desenvolvimento e o uso de sistemas de IA podem impactar uma variedade de direitos e liberdades fundamentais muito mais ampla do que a esfera coberta pela proteção de dados pessoais. Para o autor, isso deve necessariamente ser refletido nas metodologias de Avaliação de Impacto Algorítmico, que devem ir além da perspectiva limitada adotada nos modelos atuais de avaliação do impacto para proteção de dados (por exemplo, no âmbito do desenvolvimento dos relatórios de impacto à proteção de dados), que se concentram principalmente no tratamento de dados pessoais, na atribuição de responsabilidades e em parâmetros como qualidade e segurança.

O impacto de sistemas de IA envolve questões éticas e sociais que, por vezes, não são consideradas durante a elaboração de instrumentos próprios da gramática de proteção de dados. Por tais razões, Mantelero (2022) sugere que uma abordagem holística dos problemas apresentados pela IA deve ir além da ênfase tradicional da proteção de dados na transparência, na informação e na autodeterminação. Para o autor, é necessário analisar, em uma primeira camada, o impacto para direitos humanos e, em segunda camada, os valores sociais e éticos que desempenham um papel importante na abordagem de questões não jurídicas associadas ao desenvolvimento e uso soluções de IA e sua aceitabilidade, bem como acerca do equilíbrio entre os direitos e liberdades, em diferentes contextos.

---

[1] Acerca da relação entre o direito à privacidade e à proteção de dados, ainda que não seja tema central de discussão nesta pesquisa, é importante esclarecer que, inicialmente, a privacidade era compreendida por meio de noções de isolamento, reserva e reclusão, sendo pautada na ideia de um direito de estar só ("*right to be let alone*", expressão cunhada pelo magistrado Thomas McIntyre Cooleyem, em 1888, em seu "*Treatise of the law of torts"*). Com a evolução e complexidade da forma como indivíduos se relacionam, foi possível verificar um aumento do fluxo de informações e, em razão da formação de uma sociedade cada vez mais conectada, a noção de reclusão e isolamento da vida privada se tornou, em certa medida, insuficiente para assegurar a tutela da pessoa em todas as dimensões de sua personalidade. Conforme ensina Doneda (2019), os avanços tecnológicos permitem mais possibilidades de escolhas capazes de "influir diretamente em nossa esfera privada", por isso, a tutela da privacidade deixa de se basear em torno do eixo "pessoa-informação-segredo" e passa a ser sendo fundamentada no eixo "pessoa-circulação-controle", que envolve o direito de manter controle sobre as próprias informações pessoais (Rodotà, 2008).

Desse modo, no modelo proposto por Mantelero (2022), os valores éticos e sociais são observados sob a ótica dos direitos humanos, o que é fundamental para uma interpretação baseada nas particularidades contextuais e regionais, como no caso no Sul Global. Diferentes contextos culturais, sociais e econômicos podem influenciar significativamente o impacto associado a sistemas de IA, de modo que a AIA deve considerar elementos contextuais, como ambiente, público-alvo, aspectos demográficos e socioeconômicos, dentre outros.

Por fim, em relação à terceira lente teórica que ampara este trabalho, trata-se do conceito de "*accountability* algorítmica". Para ser compreendida como um instrumento eficaz de *accountability*, a AIA deve abordar o ator (quem), os fóruns (quando e onde) e o conteúdo (o quê). Essa perspectiva é adotada por Metcalf, Moss, Watkins, Singh e Elish (2023), que apontam que a AIA somente será um mecanismo eficiente se houver uma relação entre os atores e os fóruns, possibilitando que o fórum possa julgar ou exigir mudanças ao ator responsável pelo desenvolvimento ou pela implementação de determinado sistema de IA.

Os atores (por exemplo, tomadores de decisão, desenvolvedores ou responsáveis pela implementação de um sistema de IA) têm a obrigação de explicar e justificar o uso, desenvolvimento e/ou demais decisões acerca do sistema de IA, bem como seus efeitos subsequentes. Como variados tipos de atores participam em diferentes etapas do ciclo de vida de um sistema de IA, eles podem ser responsabilizados por várias espécies de fóruns — interno ou externo à organização, formal ou informal, entre outros —, seja em razão de aspectos específicos, seja em razão da totalidade do sistema.

Desse modo, o presente trabalho apoia-se na ideia de que a AIA deve possibilitar a co-construção dos impactos associados ao desenvolvimento e uso de sistemas de IA por meio das interações, prestações de conta e assunção de responsabilidades entre atores e fóruns. Metcalf, Moss, Watkins, Singh e Elish (2023) esclarecem que, ao concordar com as categorias de "impactos", as partes interessadas os estabilizam como objetos de avaliação sobre os quais podem agir.

Em síntese, sem um "fórum" de *accountability*, uma organização pode elaborar uma AIA e, ainda assim, não realizar nenhuma mitigação que realmente reduza eventuais riscos associados a sistemas de IA. Em última análise, trata-se de evitar a prática de *washing* (expressão é derivada do termo *whitewashing*, que, no inglês, refere-se à tinta cal — uma tinta branca de baixo custo, aplicada em fachadas para encobrir imperfeições), por meio da qual há a construção de uma imagem marcada pela adoção de controles e

prestação de contas, mas que, em concreto, não se verifica implementação de medidas e procedimentos efetivos na mitigação de riscos. Trata-se, em síntese, de um processo que busca mascarar falhas e omitir a falta de medidas reais por meio de uma imagem artificial de responsabilidade. Desse modo, a AIA não deve se tornar um instrumento utilizado para maquiar os riscos decorrentes do desenvolvimento e uso de sistemas de IA, mas deve ter o objetivo de identificá-los e, a partir do processo de relação ator-fórum, construir as medidas mitigatórias aplicáveis a cada situação concreta.

Sendo assim, para além da aplicação das técnicas de metodologia de pesquisa delineadas anteriormente, o trabalho se guiará por estes referenciais teóricos.

2 DEFINIÇÕES EM DISPUTA: OS CONTORNOS CONCEITUAIS DA INTELIGÊNCIA ARTIFICIAL

O termo “inteligência artificial” é comumente associado a John McCarthy (1956), que o cunhou na conferência *Dartmouth Summer Research Project on Artificial Intelligence*, porém, atualmente, não há um consenso sobre uma definição para o termo. Nesse contexto, o presente trabalho busca mapear e explorar as definições em debate, pois, o conceito adotado por eventuais legislações e regras de governança pode impactar diretamente os gatilhos que ensejam o desenvolvimento de uma Avaliação de Impacto Algorítmico.

Em primeiro lugar, uma importante distinção é a compreensão dos termos “inteligência artificial” e “sistemas de inteligência artificial”. O primeiro refere-se a uma área de estudo, composta por subcampos que incluem, por exemplo, linguagem natural, aprendizagem de máquinas, redes neurais e robótica. Trata-se, portanto, de um termo “guarda-chuva”, que abarca diferentes técnicas em estudo (Cerka *et al*., 2015). Por outro lado, o segundo termo faz referência específica aos sistemas que utilizam as abordagens técnicas de inteligência artificial, isto é, aos artefatos e instrumentos que incorporam modelos e são aplicados em diversos programas, *softwares* e funcionalidades.

Silva (2022) aponta que, entre os anos 1950 e o início da década de 1990, predominou a perspectiva simbólico-dedutiva da inteligência artificial, que buscava emular sistemas físicos de símbolos processados por cérebros humanos com a hipótese de que “a mente acessa diretamente o mundo, mas consiste em representações internas do mundo que podem ser descritas e organizadas na forma de símbolos inseridos nos programas” (Cardon; Cointet; Mazieres, 2018 apud Silva, 2022). Nessa abordagem, o ambiente computacional processa dados e segue um conjunto de instruções, como pontuações e cálculos, para gerar resultados – *outputs* – que sejam classificados ou operacionalizados de acordo com os objetivos predefinidos.

Silva (2022) destaca que, posteriormente, a partir dos anos 1990, a perspectiva conexionista-indutiva passou a ter destaque no campo da IA. Em tal perspectiva, os algoritmos recebem dados de treinamento que representam um grupo de instâncias (*input*) e exemplos de resultados (*output*) para correlacionar *inputs* e *outputs* de forma complexa, a fim de gerar decisões preditivas sobre novos dados. Desse modo, com base em um alto volume de dados de entrada (*inputs*) e de dados de saída (*outputs*) já conhecidos, o

objetivo da aplicação é construir e atualizar constantemente o “programa”, para uma otimização contínua.

Nesse contexto, na área de estudo sobre inteligência artificial, por vezes, destaca-se a técnica de aprendizado de máquina (*machine learning*), a qual é definida como qualquer metodologia e conjunto de técnicas que emprega dados, a fim de criar padrões e conhecimentos, e gera modelos a serem usados para previsões eficazes sobre esses dados, com capacidade de definir ou modificar regras de tomada de decisão de forma autônoma (Otterlo, 2013).

Cortiz (2020) aponta que as técnicas podem se dividir entre as seguintes abordagens: o aprendizado supervisionado, baseado em alto volume de dados rotulados; o aprendizado não supervisionado, baseado em alto volume de dados não rotulados; e o aprendizado por reforço, baseado em modelos de tentativa e erro. Tanto o aprendizado supervisionado quanto o não supervisionado dependem de um grande volume de dados para que a máquina aprenda. No primeiro caso, os dados precisam ter sido previamente “rotulados”, como uma espécie de etiqueta sobre o que representam, enquanto no segundo caso a rotulação prévia não é essencial.

Silva (2022) esclarece que, nesses modelos, o alvo do cálculo se desloca para o mundo externo ao modelo que lhe fornece exemplos “etiquetados” ou “classificados” de pequenos traços ou sinais em prol do objetivo do sistema algorítmico. O autor explica que, a partir das bases de treinamento, os sistemas de inteligência artificial identificam o aspecto correlacional dos dados e realizam os cálculos do programa, por vezes, com o objetivo de tomada de decisão e desenhos preditivos, como *ranking* de currículos, *score* de risco, identificação de características biométricas, classificações etc.

Em outra perspectiva, Oswald (2020), por sua vez, aponta que não é a primeira vez que uma tecnologia se torna parte da vida cultural popular e, ao mesmo tempo, é promovida como uma solução para problemas sociais. De acordo com o autor, esse movimento também pôde ser observado na primeira parte do século XX, com o polígrafo (ou "detector de mentiras").

A partir dessa constatação, Oswald (2020) traça paralelos entre as duas tecnologias e os objetivos daqueles que promoviam seu uso: ambas as tecnologias tentam prever ou categorizar o comportamento humano com base em suposições e comparação com o comportamento de outras pessoas; buscam minimizar o papel do humano como operador e intérprete da saída e como sujeito da análise; possuem a opacidade como característica

comum; e se apoiam no "realismo jurídico reformista", definido como uma abordagem que visa aumentar a produtividade, a eficiência e a condição humana.

Oswald (2020) conclui que, da mesma forma que o polígrafo não detecta mentiras — apenas registra mudanças corporais, sendo a interpretação humana de seus resultados fundamental para qualquer diagnóstico —, o aprendizado de máquina também não pode "prever", mas tão somente reduzir a experiência em dados e treinar um algoritmo para detectar padrões ou semelhanças baseadas em probabilidades.

Acerca deste ponto, Negri (2016) ressalta que a armadilha das equiparações é fazer com que as diferenças sejam mascaradas, concretizando o processo denominado por Stefano Rodotà de "expropriação da subjetividade", no qual, sob o pretexto de proteção de um sujeito abstrato, usurpam-se, no plano concreto, direitos inerentes ao ser humano. Por tal razão, o afastamento de definições pautadas em generalizações abstratas e reduções unitárias, indiferentes às distintas abordagens técnicas presentes no campo da inteligência artificial, é essencial.

Nessa direção, Rodotà (2015) destaca a importância de uma reflexão sobre o domínio da tecnologia até mesmo no vocabulário; por exemplo, os "*smartphones*" contém, em sua formação, a palavra "inteligente" (*smart*). Para o autor, isso não é um detalhe ou uma indicação trivial. Na verdade, trata-se da passagem de uma situação em que a inteligência era reconhecida apenas pelos humanos para uma situação na qual passa a ser apresentada como um atributo, também, das coisas abstratas.

Ressalta-se que, inicialmente o debate regulatório sobre IA na União Europeia enfrentou a discussão sobre atribuição de elementos essencialmente humanos a entes abstratos, uma vez que, em 2017, o Parlamento Europeu apresentou uma resolução com orientações sobre robótica, com uma proposta da criação de uma personalidade eletrônica para artefatos robóticos "inteligentes" (Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, com recomendações à Comissão Direito Civil sobre Robótica).

Naquele contexto, foi debatida a sugestão de criação de um estatuto jurídico de robôs para os artefatos mais complexos e a personalidade jurídica eletrônica foi apresentada como uma promessa de solução para os problemas de responsabilidade civil em decorrência dos danos causados por artefatos robóticos. Posteriormente, conforme será abordado por este trabalho, a proposta apresentada em abril de 2021, atualmente já aprovada e conhecida como *AI Act*, se afastou da criação de uma personalidade jurídica eletrônica e optou por uma abordagem baseada na classificação do risco de sistemas de IA.

De todo modo, é importante notar que, por vezes, sistemas de IA são compreendidos a partir de lentes antropomórficas e, consequentemente, estimulam comparações entre características humanas e atribuições de máquinas (Negri, 2020). Essa reflexão é importante porque determinados conceitos podem equiparar, indistintamente, pessoas e sistemas de IA. Isso pode ser percebido, por exemplo, no caso em que a Suprema Corte Americana reconheceu a liberdade religiosa de uma pessoa jurídica, autorizando que deixasse de fornecer aos empregados um seguro saúde que assegurava acesso a métodos anticoncepcionais (Burwell v. Hobby Lobby Stores. Inc). Da mesma forma, caso não haja uma reflexão sobre as definições que aproximam sistemas de inteligência artificial a elementos inerentemente humanos, é possível vislumbrar um cenário concreto de potencial restrição de direitos para seres humanos. Em relação ao contexto de populações historicamente marginalizadas, esse processo pode, inclusive, se manifestar de forma acentuada na usurpação de direitos inerentes ao ser humano, sob o pretexto de proteção de meras ferramentas abstratas.

Dessa forma, dois conceitos relevantes para o presente debate são “IA Geral” e “IA Estreita”, geralmente utilizados para caracterizar o alcance e a profundidade da atuação de um sistema de IA. A IA Geral ou “IA Forte” representaria um sistema de inteligência artificial que possui a capacidade de executar uma ampla gama de tarefas cognitivas com o mesmo nível de habilidade que um ser humano (Bostrom, 2018). Por outro lado, a IA Estreita ou “IA Fraca” representa sistemas especializados em tarefas específicas e limitados em termos de capacidade de adaptação a outras tarefas ou domínios, pois, em geral, não são capazes de aprender de forma independente ou adequar-se a novos contextos; dessa forma, dependem de treinamento e programação para desempenhar outras funções.

Russel e Norvig (2013) adotam o entendimento de que a inteligência artificial está relacionada a uma ação racional, ou seja, um sistema de IA é um agente inteligente que adota a melhor ação possível em uma determinada situação. Para os autores, os “agentes” caracterizam-se pela capacidade de perceber o ambiente por meio de sensores e agir por intermédio de atuadores. Nesse sentido, Russel e Norvig (2013) apontam que o conceito de racionalidade pode ser aplicado a uma série de agentes que operam em diversos espaços. Esse entendimento pode ser ilustrado, por exemplo, pelo funcionamento do robô aspirador de pó, que percebe em qual quadrado está e verifica se há sujeira nesse quadrado; uma vez que esteja sujo, a sujeira é aspirada, caso contrário, o aspirador se move para outro quadrado.

Russel e Norvig (2013) esclarecem que a definição do que é racional depende da análise de quatro fatores: a medida de desempenho definidora do critério de sucesso; o conhecimento prévio que o agente tem do ambiente; as ações que o agente pode executar; e a sequência de percepções do agente até o momento. Esses fatores levam à compreensão do termo "agente racional", que significa que para cada sequência de percepções possível, um agente racional deve selecionar uma ação que maximize sua medida de desempenho, dada a evidência fornecida pela sequência de percepções e por qualquer conhecimento interno do agente.

Os autores entendem que, quando um agente se baseia no conhecimento anterior de seu projetista e não em suas próprias percepções, esse não possui autonomia. Dessa forma, um agente racional deve ser autônomo, e isso é possível caso adquira experiência suficiente sobre o seu ambiente, tornando-se efetivamente independente do conhecimento anterior, ou seja, a incorporação do aprendizado possibilita a projeção de um agente racional.

Acerca da autonomia, Powers e Ganascia (2020) observam que esse conceito tem sido comumente vinculado aos sistemas que se comportam sem intervenção humana. Os autores apontam que há dificuldade em diferenciar autonomia de automaticidade (isto é, funcionamento "automatizado"), pois em ambos os casos há entidades que agem por si mesmas. No entanto, a distinção é relevante porque um autônomo obedece às próprias regras e reflete sobre essas, enquanto um artefato meramente automatizado atua obedecendo às regras que lhe são impostas, sem qualquer reflexão, apenas seguindo instruções previamente fornecidas.

Rodotà (2015) já questionava a autonomia atribuída a sistemas, veículos e demais artefatos dotados de inteligência artificial. Para o jurista, é necessário questionar o referencial comparativo na atribuição de autonomia, pois a autonomia parece abandonar o humano e se aproximar das coisas. Nessa direção, Negri (2020) aponta que, diante da incerteza sobre o significado do termo "autonomia", por vezes, a expressão é confundida com a imprevisibilidade do resultado.

Essa incerteza contribui para a naturalização da atribuição de autonomia a sistemas de inteligência artificial. Nesse contexto, Negri (2020) esclarece que, em termos filosóficos, a autonomia relaciona-se com a ideia de que a responsabilidade só pode ser atribuída a um agente moral, por isso, a naturalização da autonomia gera o falso entendimento de que todo robô dotado de modelo ou técnica de inteligência artificial toma decisões de forma autônoma e independente.

Ocorre que, até o momento, um sistema de IA com capacidade cognitiva igual ou superior à de um ser humano ainda não foi alcançado, de modo que a IA Geral pode ser encarada como um desafio de longo prazo, não havendo consenso sobre quando (ou se) esse tipo de sistema será alcançado. Teixeira (2015) aponta que a ideia de singularidade surge no campo da física e significa a divisão por zero, que ocorre em determinadas equações nas quais o tempo seria nulo. Por outro lado, no campo da inteligência artificial, a singularidade seria o momento no qual a inteligência das máquinas se equipararia à inteligência humana ou, até mesmo, a superaria. Para o autor, a ideia de singularidade, quando transposta para a filosofia da mente, retoma debates tradicionais, como o problema mente-cérebro e a questão da consciência.

Nesse contexto, é amplamente reconhecido o argumento do Quarto Chinês, desenvolvido por John Searle. Searle (2000) imaginou uma pessoa trancada em um quarto sem portas e sem janelas, com apenas duas portinholas em paredes opostas. Essa pessoa só conhecia um determinado idioma, mas lhe era fornecido um texto em chinês e uma espécie de tabela com regras escritas em seu idioma para que ela, a partir de sentenças escritas em chinês, gerasse novas sentenças, também em chinês. Desse modo, o ocupante do quarto gerava um terceiro texto, com base no texto inicial e nos novos textos — todos em chinês —, usando as regras de transformação da tabela.

O Quarto Chinês retrata o que aconteceria no interior de um computador, isto é, o texto inicial corresponde ao *input* fornecido ao computador. O ponto central é: ainda que a pessoa produza sentenças em chinês, ela não compreende chinês; do mesmo modo, um computador apenas manipula símbolos, os quais não possuem qualquer significado para o computador. Seguindo esse raciocínio, Searle (2000) esclarece que não há mente sem consciência, e a intencionalidade e consciência tem como base o cérebro vivo. Para o autor, um sistema de inteligência artificial é, no máximo, a simulação de um processo cognitivo — e não um processo cognitivo por si só, o que significa que "parecer" consciente não é o mesmo que "ser" consciente.

Um exemplo deste debate se dá entorno da alegação feita por Blake Lemoine, engenheiro do Google, de que o sistema de inteligência artificial LaMDA teria desenvolvido consciência. Segundo o engenheiro, se comportava de forma tão sofisticada em suas interações que parecia capaz de sentir e raciocinar como um humano, expressando sentimentos, preocupações com direitos e até medo da morte. Nesse contexto, Souza (2022) esclarece que, embora sistemas como o LaMDA sejam impressionantes em sua capacidade de gerar respostas coerentes e contextualmente

apropriadas, isso não significa que possuam consciência, pois foram projetados para gerar texto com base em padrões linguísticos, sem uma compreensão real do conteúdo envolvido.

Desse modo, apesar dos constantes avanços no campo da inteligência artificial, a replicação da consciência em máquinas envolve desafios complexos e ainda é tema de especulações. A própria natureza da consciência humana ainda não é completamente compreendida pela ciência, por isso, conceituações devem alcançar diferentes usos e aplicações de inteligência artificial, evitando equiparações antropomórficas que podem vir a afetar o exercício de direitos por parte de seres humanos.

Conceitos e definições que atribuem características humanas para objetos de uso comum devem ser questionados. Nessa direção, Crawford (2021) aponta que sistemas de IA não são autônomos, racionais ou capazes de discernir nada sem um treinamento extenso e computacionalmente intensivo, com conjuntos de dados, regras e recompensas pré-definidas. Por tal razão, o presente trabalho não trata sobre a viabilidade e potenciais definições para IA Geral, focando na aplicação de IA Estreita.

A antropomorfização ou supervalorização das capacidades de IA pode distorcer o debate público e levar a decisões éticas e regulatórias inadequadas. Nessa direção, Silva (2022) aponta que a tendência a hipervisibilizar os debates filosóficos sobre robôs autômatos e seus possíveis direitos no futuro, no âmbito do desenvolvimento de sistemas de IA Geral, pode tirar o foco de debates voltados para a realidade material do impacto da IA Estreita na vida contemporânea.

Ainda que alcançar um consenso não seja tarefa fácil, o debate é essencial, considerando a natureza transversal da inteligência artificial, aplicável a diferentes atividades, setores e países, o que levanta também uma futura necessidade de interoperabilidade entre normas de diferentes Estados. Nessa direção, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) definia a inteligência artificial como

> um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos por humanos, fazer previsões, recomendações ou tomar decisões que influenciam ambientes reais ou virtuais e que os sistemas de IA são projetados para operar com níveis variados de autonomia (OCDE, 2019).

Posteriormente, a definição adotada pela OCDE foi atualizada para definir sistema de IA como

> um sistema baseado em máquina que pode, para objetivos explícitos ou implícitos, inferir, a partir da entrada que recebe, como gerar resultados, tais como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após a implantação (OCDE, 2023).

Observa-se que a atualização visa destacar os objetivos de um sistema de IA como explícitos — quando são diretamente programados no sistema por um desenvolvedor humano — ou implícitos — quando são estabelecidos por um conjunto de regras especificadas por um ser humano ou quando o sistema é capaz de aprender novos objetivos, como ocorre com carros autônomos e *large language models* (modelos de inteligência artificial que foram treinados em grandes conjuntos de dados linguísticos para realizar tarefas relacionadas ao processamento de linguagem natural).

Além disso, a inclusão da frase "inferir, a partir da entrada que recebe" enfatiza a função dos *inputs*, que podem ser fornecidos por humanos ou máquinas, na operação de sistemas de IA. Por fim, destaca-se a adição do trecho "adaptabilidade após a implantação", o qual reflete que alguns sistemas de IA podem continuar evoluindo após serem projetados e implantados — por exemplo, sistemas de recomendação que se ajustam às preferências individuais ou sistemas de reconhecimento de voz que se adaptam à voz do usuário (Russel, 2023).

Por sua vez, a UNESCO aponta que não há uma única definição para o termo e, caso houvesse, necessitaria ser alterada com o tempo. A fim de endereçar as funcionalidades de sistemas de IA que possuem relevância ética, a UNESCO define sistemas de IA como sistemas que têm a capacidade de processar dados e informações de uma forma que se assemelha a um comportamento inteligente e, normalmente, inclui aspectos de raciocínio, aprendizado, percepção, previsão, planejamento ou controle.

Para a UNESCO, os sistemas de IA são tecnologias de processamento de informações que integram modelos e algoritmos, os quais produzem uma capacidade de aprender e executar tarefas cognitivas, alcançando resultados como previsão e tomada de decisões em ambientes materiais e virtuais. Os sistemas de IA são projetados para operar com vários graus de autonomia, por meio da modelagem, da representação do conhecimento, da exploração de dados e do cálculo de correlações; além disso, eles podem incluir vários métodos, como aprendizagem de máquina, incluindo aprendizagem profunda e por reforço, e raciocínio de máquina, incluindo planejamento, programação, representação de conhecimento e de pesquisa de raciocínio e otimização.

Nos Estados Unidos, diferentes definições emergem em documentos e propostas regulatórias. A proposta regulatória denominada *FUTURE of Artificial Intelligence Act of 2020* (*Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act of 2020*) define que sistemas de IA são (i) quaisquer sistemas artificiais que executem tarefas em circunstâncias variáveis e imprevisíveis, sem supervisão humana significativa, ou que possam aprender com sua experiência e melhorar seu desempenho. Esses sistemas podem ser desenvolvidos em *software* de computador, *hardware* físico ou outros contextos ainda não contemplados. Eles podem resolver tarefas que exijam percepção, cognição, planejamento, aprendizado, comunicação ou ação física semelhantes às humanas; (ii) sistemas que pensam como seres humanos, como arquiteturas cognitivas e redes neurais; (iii) sistemas que agem como seres humanos, como os que podem passar no teste de Turing ou outro teste comparável por meio de processamento de linguagem natural, representação de conhecimento, raciocínio automatizado e aprendizado; e (iv) sistemas que agem racionalmente, como agentes de *software* inteligentes e robôs incorporados que atingem objetivos por meio de percepção, planejamento, raciocínio, aprendizado, comunicação, tomada de decisões e ação.

Já na *National Artificial Intelligence Research Resource Task Force*, também nos Estados Unidos, adota-se a definição de que um sistema de IA é um sistema baseado em uma máquina que pode, para um determinado conjunto de objetivos definidos por humanos, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas (IA) usam entradas baseadas em máquinas e em humanos para perceber ambientes reais e virtuais; para abstrair essas percepções em modelos por meio de análise de forma automatizada; e para usar a inferência de modelos para formular opções de informações ou ações.

Ainda no âmbito dos Estados Unidos, o *White House Office of Science and Technology* traz uma definição com ênfase em tecnologias que impactam direitos individuais e coletivos, apontando que sistemas de IA são sistemas automatizados com o potencial de afetar significativamente os direitos, as oportunidades ou o acesso do público americano a recursos ou serviços essenciais. Um "sistema automatizado" é qualquer sistema, *software* ou processo que use a computação — como parte ou todo de um sistema — para determinar resultados, tomar decisões, informar a implementação de políticas, coletar dados ou interagir de outra forma com indivíduos e/ou comunidades. Os sistemas automatizados incluem, entre outros, sistemas derivados de aprendizado de máquina, estatística ou outras técnicas de processamento de dados e inteligência artificial, e

excluem a infraestrutura de computação passiva. Em toda essa estrutura, os sistemas automatizados considerados no escopo da definição são apenas aqueles que têm o potencial de impactar significativamente os direitos, as oportunidades ou o acesso de indivíduos ou comunidades.

O Canadá, por meio da proposta de *Artificial Intelligence and Data Act* define o sistema de IA como um sistema tecnológico que, de forma autônoma ou parcialmente autônoma, processa dados relacionados a atividades humanas por meio do uso de um algoritmo genético, uma rede neural, aprendizado de máquina ou outra técnica para gerar conteúdo, tomar decisões, fazer recomendações ou previsões.

No Japão, as *Governance Guidelines for Implementation of AI Principles*, publicadas pelo *AI Governance Guidelines Working Group*, estabelecem que um sistema de IA é desenvolvido com uma abordagem de aprendizado de máquina, incluindo aprendizado supervisionado, não supervisionado e por reforço, usando uma ampla variedade de métodos, inclusive aprendizado profundo, e que pode, para um determinado conjunto de objetivos definidos por humanos, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. As *Guidelines* apontam que os sistemas de IA são projetados para operar com níveis variados de autonomia e que a definição inclui não apenas *software*, mas também uma máquina que contém *software* como um elemento.

No âmbito da União Europeia, nota-se que existem diferentes definições em circulação. O Conselho da Europa, em seu Glossário de Inteligência Artificial, aponta que a inteligência artificial (enquanto área de estudo) é um conjunto de ciências, teorias e técnicas, cujo objetivo é reproduzir, em uma máquina, as habilidades cognitivas de um ser humano, de modo que os desenvolvimentos atuais visam confiar a uma máquina tarefas complexas, anteriormente delegadas a seres humanos.

Por outro lado, a Comissão Europeia, por meio do *High-Level Expert Group on AI*, já definiu que sistemas de IA são sistemas de *software* (e, possivelmente, também de *hardware*) projetados por seres humanos que, tendo em vista um objetivo complexo, atuam na dimensão física ou digital, percebendo seu ambiente por meio da aquisição de dados, interpretando os dados — estruturados ou não estruturados — coletados, raciocinando sobre o conhecimento ou processando as informações derivadas desses dados e decidindo as melhores ações a serem tomadas para atingir o objetivo determinado.

O *High-Level Expert Group on AI* aponta que os sistemas de IA podem usar regras simbólicas ou aprender um modelo numérico, e podem adaptar seu comportamento

analisando como o ambiente é afetado por suas ações anteriores. Além disso, o grupo esclarece que, como disciplina científica, a IA inclui várias abordagens e técnicas, tal qual aprendizagem de máquina — da qual a aprendizagem profunda e a aprendizagem por reforço são exemplos específicos —, raciocínio de máquina — que inclui planejamento, programação, representação e raciocínio de conhecimento, pesquisa e otimização — e robótica — que inclui controle, percepção, sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas ciberfísicos.

Em relação à definição adotada por propostas regulatórias, nota-se que, em 2021, a Comissão Europeia publicou a primeira versão da Proposta de Regulamento do Parlamento Europeu e do Conselho: Estabelecimento de regras harmonizadas sobre IA e alteração de determinados atos legislativos da União (*Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on AI and Amending Certain Union Legislative Acts*). O referido documento definia que um sistema de IA é um *software* desenvolvido com uma ou mais das técnicas e abordagens listadas no Anexo I e que pode, para um determinado conjunto de objetivos definidos por humanos, gerar resultados como conteúdo, previsões, recomendações ou decisões que influenciam os ambientes com os quais interagem.

O Anexo I previa as seguintes técnicas e abordagens de IA: (i) abordagens de aprendizado de máquina, inclusive aprendizado supervisionado, não supervisionado e por reforço, usando uma ampla variedade de métodos, inclusive aprendizado profundo; (ii) abordagens baseadas em lógica e conhecimento, inclusive representação de conhecimento, programação indutiva (lógica), bases de conhecimento, mecanismos de inferência e dedução, raciocínio (simbólico) e sistemas especializados; (iii) abordagens estatísticas, estimativa bayesiana, métodos de busca e otimização.

Posteriormente, em razão do processo de negociação institucional entre o Parlamento Europeu, a Comissão Europeia e o Conselho Europeu — no qual um dos temas mais debatidos foi a definição de sistema de inteligência artificial — houve alteração da abordagem de definição. A redação adotada pelo texto final do *AI Act* é baseada na definição da OCDE, mas conta com pequenas variações gramaticais: sistema baseado em máquina projetado para operar com vários níveis de autonomia e que pode apresentar adaptabilidade após a implantação e que, para objetivos explícitos ou implícitos, infere, a partir das informações recebidas, como gerar resultados que podem influenciar ambientes físicos ou virtuais.

No Brasil, a redação original do Projeto de Lei n.° 2.338 aborda a definição de sistemas de inteligência artificial, se aproximando da definição adotada pela OCDE e conceituando, em seu art. 4º, inciso I, um sistema de inteligência artificial como

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

Além disso, a redação do Projeto de Lei n.° 2.338/2023 apresentava as figuras dos agentes de inteligência artificial, isto é, fornecedores e operadores de sistemas de IA. Nos termos do art. 4º, II da proposta original, o fornecedor era definido como

> [...] a pessoa natural ou jurídica, de natureza pública ou privada, que desenvolva um sistema de inteligência artificial, diretamente ou por encomenda, com vistas a sua colocação no mercado ou a sua aplicação em serviço por ela fornecido, sob seu próprio nome ou marca, a título oneroso ou gratuito (Brasil, 2023).

Por outro lado, o operador era conceituado como

> [...] a 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).

Após as discussões e debates realizados no âmbito da Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA), o relatório final apresentado pelo Relator trouxe alterações em relação ao texto inicialmente apresentado. Em primeiro lugar, a definição de sistema de inteligência artificial passou por alterações que seguem o racional adotado pela conceituação proposta pela OCDE:

> sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial, previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real (Brasil, 2024).

Além disso, o relatório final introduz a definição de sistema de inteligência artificial de propósito geral (SIAPG)

> sistema de IA baseado em um modelo de IA treinado com bases de dados em grande escala, capaz de realizar uma ampla variedade de tarefas distintas e servir diferentes finalidades, incluindo aquelas para as quais não foram especificamente desenvolvidos e treinados, podendo ser integrado em diversos sistemas ou aplicações (Brasil, 2024).

E, ainda, de inteligência artificial generativa (“IA generativa”)

> modelo de IA especificamente destinado a gerar ou modificar significativamente, com diferentes graus de autonomia, texto, imagens, áudio, vídeo ou código de software (Brasil, 2024).

Acerca dos agentes de inteligência artificial, estes passam a ser compreendidos como os desenvolvedores, fornecedores e aplicadores que atuem na cadeia de valor e na governança interna de sistemas de IA, o que será especificamente definido em futuro regulamento. De toda forma, o relatório final apresenta a respectiva definição de cada agente de inteligência artificial

> desenvolvedor de sistema de inteligência artificial: pessoa natural ou jurídica, de natureza pública ou privada, que desenvolva um sistema de inteligência artificial, diretamente ou por encomenda, com vistas a sua colocação no mercado ou a sua aplicação em serviço por ela fornecido, sob seu próprio nome ou marca, a título oneroso ou gratuito;
>
> fornecedor: pessoa natural ou jurídica, de natureza pública ou privada, que disponibiliza e distribui sistema de IA para que terceiro o opere a título oneroso ou gratuito (Brasil, 2024).
>
> aplicador: pessoa natural ou jurídica, de natureza pública ou privada, que empregue ou utilize, em seu nome ou benefício, sistema de inteligência artificial, inclusive configurando, gerenciando, mantendo ou apoiando com o fornecimento de dados para sua operação e monitoramento

Por fim, cabe ressaltar que, no âmbito dos *frameworks*, códigos e padrões para gerenciamento de sistemas de IA, o *framework* desenvolvido pela BSA – The Software Alliance define “inteligência artificial” como sistemas que usam algoritmos de aprendizado de máquina, passíveis de analisar grandes volumes de dados de treinamento para identificar correlações, padrões e outros metadados, os quais podem ser utilizados para desenvolver um modelo capaz de fazer previsões ou recomendações baseadas em futuras entradas de dados (BSA, 2021).

Por sua vez, o *NIST AI Risk Management Framework* (AI RMF), desenvolvido pelo *National Institute of Standards and Technology* (NIST), ligado ao Departamento de Comércio dos Estados Unidos, define “sistema de inteligência artificial” como um

sistema de engenharia, ou baseado em máquina, que pode, para um determinado conjunto de objetivos, gerar resultados como previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais, pontuando que tais sistemas são projetados para operar com vários níveis de autonomia (NIST, 2023).

A partir do levantamento das discussões em âmbito nacional e internacional, é possível notar que não há um consenso sobre uma definição para sistemas de inteligência artificial. As definições em debate variam significativamente, abarcando desde abordagens com ênfase na capacidade de aprendizado e adaptação das máquinas até perspectivas que realçam a imitação do comportamento humano e a execução de tarefas complexas.

Schuett (2023) elenca requisitos que uma definição legal para o termo inteligência artificial" deveria observar: (i) não deve ser excessivamente inclusiva, isto é, incluir situações que não necessitam de regulamentação; (ii) não deve ser subinclusiva, ou seja, não incluir casos que deveriam ter sido incluídos em uma regulamentação; (iii) deve ser precisa, de modo que seja possível determinar claramente se um sistema específico se enquadra ou não na definição; (iv) deve ser compreensível para que pessoas sem conhecimento técnico altamente especializado sejam capazes de aplicar a definição; (v) deve ser prática, de modo que seja possível determinar com pouco esforço se um caso concreto se enquadra ou não na definição; e (vi) deve ser flexível para acomodar o progresso técnico, contendo apenas elementos que provavelmente não mudarão em um futuro previsível.

Conforme aponta Schuett (2023), definir o escopo das regulamentações de IA é particularmente desafiador porque o termo "IA" é utilizado para conceituar uma série de sistemas diferentes, aplicáveis em diferentes atividades e setores e que, portanto, devem ser tratados de forma diferente. Além disso, a expressão "inteligência artificial" é altamente ambígua e seu significado muda com o tempo. Em síntese, a natureza multifacetada e em constante evolução dos sistemas de IA contribui para essa dificuldade na formulação de uma definição unificada.

Além disso, nota-se que as propostas de definição veiculadas no âmbito regulatório — por exemplo, em propostas legislativas — são, geralmente, mais amplas, além de abrangerem grande parte dos sistemas de IA. Trata-se de uma tentativa de extensão do escopo de aplicabilidade da regulação. Por outro lado, verifica-se que determinadas definições, como a proposta pela BSA, possuem escopo de aplicação mais restrito.

Desse modo, a ausência de consenso não apenas suscita desafios teóricos, mas também impacta diretamente questões práticas, como o escopo de aplicabilidade das regulamentações e, consequentemente, de direitos e obrigações eventualmente previstos em tais instrumentos. Por isso, Schuett (2023) defende que o escopo material das regulamentações sobre IA não deve se basear na definição do termo “inteligência artificial”, sendo preferível se concentrar nos riscos específicos que tais regulamentações desejam reduzir, por exemplo, por meio da definição das principais fontes de riscos relevantes.

Por conseguinte, embora a presente pesquisa não se concentre primariamente na análise de definições e conceitos pertinentes ao termo "sistemas de inteligência artificial", reconhece-se a essencialidade de abordar este debate. Isso se justifica pelo fato de que, a depender das escolhas realizadas para fins de construção de uma definição, o âmbito de aplicação das Avaliações de Impacto Algorítmico poderá ser substancialmente afetado.

Nesse contexto, entende-se que as estratégias de regulamentação e a própria condução de Avaliações de Impacto Algorítmico devem ser orientadas principalmente pelos objetivos e pelos efeitos do uso da tecnologia ao invés de focar exclusivamente em abordagens técnicas ou jurídicas de definição do termo “inteligência artificial”. Trata-se, em síntese, de uma estratégia de priorização de uma compreensão abrangente e holística dos impactos sociais, éticos e jurídicos gerados por sistemas de IA e não apenas de aspectos técnicos.

Sem dúvida, a incerteza em torno das definições impacta diretamente a determinação de quais sistemas seriam abrangidos por futuras regulações específicas, desafiando a formulação de práticas regulatórias que atendam adequadamente a diversidade e a complexidade das aplicações de inteligência artificial em diversos setores da sociedade, mas que não sejam excessivamente amplas e aplicáveis a uma infinidade de sistemas, programas e *softwares*.

A clareza e precisão na definição conceitual serão, portanto, importantes para estabelecer critérios eficazes sobre quando será necessário desenvolver Avaliações de Impacto Algorítmico. No entanto, é importante ressaltar que a dificuldade técnica na definição não deve ser vista como um obstáculo para o desenvolvimento da AIA e de outros instrumentos que buscam mitigar riscos associados a sistemas de IA. Em síntese, a decisão sobre desenvolver ou não uma AIA pode ser baseada nas finalidades e nos efeitos da aplicação da tecnologia e não exclusivamente em um conceito técnico ou jurídico sobre sistemas de IA. A falta de um conceito amplamente aceito para "sistemas

de IA" e a complexidade associada à sua definição não devem ser justificativas para os agentes envolvidos no uso, comercialização e implantação de sistemas de IA não realizem a avaliação dos impactos gerados por tais sistemas.

Desse modo, em vez de se fixarem em definições rígidas e precisas de "sistemas de IA", os agentes envolvidos podem avaliar a necessidade de condução de uma avaliação de impacto quando – pelas finalidades e potenciais efeitos da tecnologia – for identificado potencial risco elevado para as pessoas afetadas. Portanto, a ausência de um conceito globalmente aceito para a expressão "sistemas de IA" não deve ser uma justificativa para a inação em relação à identificação de riscos associados ao desenvolvimento e o uso de sistemas de IA, ao desenvolvimento de avaliações de impacto e à implementação de eventuais medidas mitigatórias aplicáveis.

3 FUNDAMENTAÇÃO TEÓRICA ACERCA DA AVALIAÇÃO DE IMPACTO ALGORÍTMICO

No contexto do crescente desenvolvimento e implementação de sistemas de inteligência artificial, especialmente em atividades relacionadas à tomada de decisões e à geração de pontuações, classificações e previsões, surgem preocupações sobre a necessidade de promover práticas eficazes de governança e prestação de contas dos agentes envolvidos em várias etapas do ciclo de vida de um sistema de IA. Em síntese, à medida que a IA se torna cada vez mais integrada em processos críticos, como aqueles que influenciam decisões humanas, surgem questionamentos sobre como garantir a transparência e a responsabilidade no uso desses sistemas.

Dentre as práticas de governança adotadas por entes que desenvolvem ou implementam sistemas de IA, emerge a Avaliação de Impacto Algorítmico (AIA), como um instrumento que busca avaliar o nível de segurança, exatidão, solidez e desempenho de sistemas de IA, bem como avaliar impactos e mitigar riscos às pessoas e grupos afetados. Além disso, a AIA carrega o potencial de promover uma abordagem proativa e preventiva de gestão de riscos, o que pode fortalecer a confiança do público nos sistemas de IA e nas instituições que os empregam.

Dessa forma, a Avaliação de Impacto Algorítmico busca avaliar os efeitos e as consequências do desenvolvimento e/ou implementação de sistemas de IA, não apenas considerando suas finalidades e benefícios, mas também identificando eventuais riscos e efeitos adversos. O objetivo é garantir que o desenvolvimento e uso desses sistemas estejam em conformidade com princípios éticos e jurídicos, a fim de assegurar uma abordagem responsável e alinhada com o respeito e promoção aos direitos humanos.

De todo modo, é importante ressaltar que o que constitui um "impacto" não é sempre algo claro e evidente, o que levanta o desafio central de determinar o alcance do impacto a ser avaliado por uma AIA. O termo "impacto" denota uma relação causal, na qual uma ação realizada por um ente (ou por um sistema operado por ele) gera uma mudança no mundo, afetando algum aspecto do ambiente e tornando-o diferente (Watkins *et al*., 2021). Nesse sentido, Watkins *et al.* (2021) apresentam um desafio central para as avaliações de impacto: “O que pode ser identificado como um impacto resultante de uma causa específica?” e “Como essa causa pode ser adequadamente identificada como decorrente de algo controlado pela organização?”.

Nesse sentido, é importante notar que o processo de identificação, medição, formalização e contabilização dos "impactos" é intrinsecamente marcado por dinâmicas de poder. Trata-se de processo que carece de neutralidade na definição do que constitui um "impacto" e, por conseguinte, será ou não passível de análise no âmbito da AIA. A definição do “impacto” é, portanto, influenciada pelo poder social, econômico e político, uma vez que os impactos a serem avaliados em uma AIA são, em última instância, determinados por decisões sobre a inclusão ou exclusão de determinados efeitos para avaliação (Watkins *et al*., 2021).

Em síntese, a noção de "impacto" em uma AIA não é apenas uma medida objetiva e neutra. Quando os impactos de um sistema de inteligência artificial (IA) são considerados pelo ente que o desenvolve ou o aplica, ele está, na verdade, realizando um exercício de atribuição de valor e importância a determinados resultados e consequências. Essa atribuição de valor não é uma questão puramente técnica, pois as decisões sobre quais impactos serão considerados relevantes e merecedores de avaliação são influenciadas pelas estruturas sociais, econômicas e políticas que moldam as interações e relações de poder na sociedade.

Tendo em vista essas considerações iniciais sobre a Avaliação de Impacto Algorítmico (AIA) como um processo marcado pelo exercício de poder e pela influência de diversos fatores sociais, econômicos e políticos, parte-se da premissa de que, embora seja concebida como uma ferramenta de *accountability*, a AIA não deve ser compreendida como um instrumento neutro. A determinação de quais impactos são priorizados e como são ponderados frequentemente reflete as agendas e interesses das partes envolvidas. A partir dessa compreensão, o presente trabalho se propõe a explorar e analisar, neste capítulo, as diferentes concepções da AIA na literatura especializada, com o objetivo de fornecer uma compreensão abrangente e crítica deste instrumento.

Nesse sentido, para o *Ada Lovelace Institute* (2022), a AIA é apresentada como uma ferramenta para avaliar os possíveis impactos sociais de um sistema de IA antes mesmo de sua implementação. A AIA visa estabelecer confiança pública no desenvolvimento e uso de sistemas de IA, ao mesmo tempo em que busca mitigar o potencial de danos a indivíduos e grupos, e maximizar os benefícios gerados pela tecnologia. Trata-se de um instrumento que pode ser inserido na estrutura de governança de uma organização, oferecendo suporte à tomada de decisões por meio da visualização e monitoramento dos resultados dos sistemas. Além disso, fornecem suporte na

disponibilização de informações sobre as razões por trás de uma decisão e ajudam a equilibrar conflitos de interesses (Mökander; Floridi, 2021).

A AIA também pode ser compreendida a partir da metodologia mais ampla das avaliações de impacto em geral, caracterizadas como um tipo de avaliação com histórico de uso em outros domínios, por exemplo, no setor financeiro, na segurança cibernética, na proteção de dados e no âmbito ambiental e climático. Desse modo, verifica-se que a realização de uma avaliação de impacto oferece aos agentes responsáveis pelo uso ou desenvolvimento de sistemas de IA uma maneira de avaliar os possíveis impactos econômicos, sociais e ambientais de uma política, tecnologia ou intervenção proposta em determinado contexto.

Atualmente, o campo de estudo acerca da AIA está em desenvolvimento, passando por constantes debates, de modo que não há uma única regra ou uma metodologia definida para elaboração desse instrumento. Os processos conduzidos podem variar a depender do contexto de desenvolvimento ou implementação de determinado sistema de IA. Como resultado, pesquisadores e profissionais estão constantemente explorando novas abordagens, técnicas e práticas para melhorar a eficácia e a aplicabilidade deste instrumento.

Nesse contexto, é importante esclarecer as principais diferenças entre os instrumentos comumente associados às práticas de *accountability* e governança em IA: auditoria, garantia (*assurance*) e avaliação (*assessment*). Uma auditoria é uma avaliação independente, que segue regras claras e visa atender à sociedade, ao público, aos usuários ou a algum outro órgão independente. As auditorias envolvem três partes distintas: o auditor, a organização avaliada e os interessados na auditoria; e seus critérios são estabelecidos para produzir um resultado binário, indicando se o sistema está em conformidade ou não com as regras utilizadas. As garantias, por sua vez, são baseadas em mecanismos de *soft law*, tais como princípios, regras ou padrões que não possuem obrigatoriedade legal ou não estão codificados em legislação (Hasan *et al*., 2022).

Por fim, as avaliações podem ser realizadas internamente ou por terceiros e, assim como as garantias, não têm uma natureza binária de conformidade ou não conformidade. Seu propósito é fornecer *feedback* e recomendações relacionadas a medidas mitigatórias ou áreas de melhoria para um melhor desempenho em relação a padrões legais e/ou éticos (Hasan *et al*., 2022). Essas avaliações podem incluir componentes técnicos e não técnicos (Mökander; Floridi, 2021), sendo exemplos comuns as avaliações técnicas de viés (*bias*) e as avaliações éticas de risco ou impacto (Brown; Davidovic; Hasan, 2021).

Selbst (2021) aponta que, em geral, a AIA pode ser encaixada em três categorias, a depender do modelo utilizado para sua elaboração: (i) modelos baseados na NEPA (*National Environmental Policy Act Assessment*), uma avaliação conduzida na área ambiental — semelhante à Avaliação de Impacto Ambiental presente no contexto brasileiro em razão da Política Nacional do Meio Ambiente (Lei nº 6.938, de 31 de agosto de 1981); (ii) modelos baseados no *Data Protection Impact Assessment* (DPIA), uma avaliação existente no âmbito do *General Data Protection Regulation* (GDPR), semelhante ao Relatório de Impacto à Proteção de Dados (RIPD), previsto pela Lei Geral de Proteção de Dados (Lei Federal n.º 13.708/2019); e (iii) modelo de questionário pré-formulado, nos quais o agente responsável pela condução da AIA realiza o preenchimento de campos já parametrizados.

O modelo de avaliação ambiental é altamente detalhado e envolve transparência e participação pública. No ordenamento jurídico brasileiro, a avaliação de impacto ambiental é compreendida como instrumento de planejamento e gestão e, ainda, como um procedimento associado a processos decisórios, como é o caso do licenciamento ambiental, tendo por objetivo final a análise da viabilidade ambiental de um projeto, programa ou plano (Milaré, 2000).

No Brasil, há dois instrumentos que fazem parte desse arranjo de avaliação de impactos: o Estudo de Impacto Ambiental (EIA) e o Relatório de Impacto Ambiental (RIMA), ambos regulados pelo Conselho Nacional do Meio Ambiente (CONAMA). Em síntese, o RIMA é o relatório que apresenta as conclusões obtidas no EIA e deverá conter informações como as finalidades e justificativas do projeto e sua relação com políticas setoriais e planos governamentais, a descrição e alternativas tecnológicas do projeto, um resumo dos diagnósticos ambientais e uma descrição dos prováveis impactos ambientais da implementação da atividade. O texto do RIMA deve ser publicizado, possibilitando acesso ao público, sendo, ainda, instruído por mapas, quadros, gráficos e demais recursos necessários ao entendimento do projeto e de suas consequências.

Por outro lado, as avaliações de risco à privacidade — embora também possuam conteúdo expansivo — não envolvem a obrigatoriedade de tornar o processo avaliativo público. No contexto brasileiro, a Lei Geral de Proteção de Dados (LGPD) estabelece que o relatório de impacto à proteção de dados pessoais é documentação que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco.

A Autoridade Nacional de Proteção de Dados (ANPD), em suas orientações sobre a elaboração de relatórios de impacto à proteção de dados, esclarece que não há uma obrigação de divulgação pública do relatório, mas que permitir o acesso ao público em geral pode ser uma medida que demonstra a preocupação em relação aos dados pessoais tratados (ANPD, 2023). No entanto, a ANPD esclarece que a versão pública do RIPD pode ser distinta da versão interna, no intuito de resguardar segredos comercial e industrial e outras informações protegidas por lei. Por outro lado, no caso de entidades e órgãos públicos, o relatório de impacto à proteção de dados deverá ser publicado: (i) por determinação da ANPD, nos termos do art. 32 da LGPD; ou (ii) pelo próprio controlador, quando não identificada hipótese de sigilo aplicável ao caso, em conformidade com a Lei n.º 12.527, de 18 de novembro de 2011.

No âmbito da proteção de dados, é interessante notar que parte da doutrina entende a necessidade de conduzir uma AIA ao lidar com tratamento de dados pessoais em processos de tomada de decisão automatizada (Kaminski; Malgiei, 2020). Isso se deve ao fato de que, em geral, um dos fatores considerados para fins de avaliação do risco de uma atividade de tratamento de dados pessoais é a presença de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais. Esse entendimento foi adotado, inclusive, pela Eslovênia na regulamentação interna do GDPR e pela *Information Commissioner's Office* (ICO) — autoridade britânica de proteção de dados — em proposta de *framework* para auditoria de IA (*Guidance on the AI auditing framework*), ainda em fase de consulta.

Como o uso de dados pessoais é essencial para o funcionamento de diversos sistemas de IA, perspectivas iniciais podem ser extraídas da regulação de privacidade e proteção de dados, como parâmetros de segurança, qualidade e governança. Por exemplo, em uma AIA pode ser relevante verificar, dentre outros parâmetros, se o sistema de IA captura e interpreta corretamente as relações que existem nos dados de treinamento, produzindo resultados de qualidade e livres de enviesamento, se este sistema é sensível a ataques adversários ou, de forma mais geral, a mudanças inesperadas em seu ambiente ou uso, afetando sua segurança.

Acerca da compreensão da AIA como um modelo de avaliação de riscos de proteção de dados, é importante observar que as consequências do desenvolvimento e uso de sistemas de IA não se limitam às questões de privacidade e tratamento de dados pessoais, abrangendo impactos para direitos humanos em geral (Mantelero, 2022). Assim, é fundamental que a AIA seja compreendida a partir de uma lente de proteção aos direitos

e liberdades fundamentais, abordando, também aspectos éticos e sociais, e não somente de proteção de dados pessoais.

Além disso, o avanço dos sistemas de inteligência artificial, especialmente de ferramentas de IA generativa, por vezes, tensiona e desafia a disciplina de proteção de dados, incluindo direitos, regras e princípios tradicionalmente previstos em normas que tratam da matéria. Por exemplo, em relação aos princípios de proteção de dados, é possível notar pontos de tensão, especialmente no que diz respeito aos princípios da finalidade, adequação e necessidade e transparência (art. 6º, I, II e III, da LGPD). Isso porque a IA generativa frequentemente utiliza dados para finalidades novas e imprevistas, além do escopo original para o qual foram coletados, desafiando a tradicional máxima de limitação da finalidade, que exige que os dados pessoais sejam utilizados apenas para finalidades claramente definidas no momento da coleta.

Na mesma direção, o princípio da necessidade exige que o tratamento de dados seja limitado ao mínimo necessário, abrangendo somente dados pertinentes, proporcionais e não excessivos em relação às finalidades. Ocorre que, em geral, sistemas de IA dependem de grandes volumes de dados para evitar vieses e melhorar sua precisão. Inclusive, excluir ou mascarar dados pessoais pode, em alguns casos, prejudicar a qualidade do modelo. Nesse ponto, há uma nítida tensão entre os princípios de minimização de dados e a necessidade de ampla diversidade de dados para evitar vieses, como aponta o *Centre for Information Policy Leadership* (CIPL, 2024). A limitação do prazo de retenção de dados pessoais, que determina que tais informações não devem ser mantidas por mais tempo do que o necessário também entra em conflito com as práticas relacionadas ao desenvolvimento e uso de sistemas de IA, uma vez que a retenção prolongada de dados é, por vezes, essencial para o treinamento contínuo, rastreabilidade, auditoria e supervisão dos modelos.

Evidentemente, os pontos de tensão não se limitam a estes, sendo possível citar, ainda, os desafios de atendimento de direitos dos titulares de dados (previstos pelo art. 18 da LGPD), enquadramento das atividades de tratamento de dados pessoais em uma das bases legais previstas pela legislação (arts. 7º e 11, da LGPD) e, até mesmo, legitimidade de operações de transferência internacional de dados (art. 33, da LGPD), dada a eventual necessidade de utilizar dados diversificados e geograficamente dispersos para treinar e operar modelos de IA.

Desse modo, nota-se que, embora o relatório de impacto à proteção de dados possa ser um alicerce inicial para avaliação de sistemas de IA, é necessário ir além dos

tradicionais controles presentes na disciplina de privacidade e proteção de dados pessoais que podem ser insuficientes e, por vezes, incoerentes em relação ao funcionamento de tais sistemas.

Por fim, a terceira abordagem é baseada em questionários para preenchimento por parte dos agentes responsáveis pela elaboração da AIA. Nesse caso, os agentes realizam o preenchimento de campos previamente formulados e parametrizados, participando apenas da inserção de informações, mas não da construção do modelo de avaliação de impacto. Esse modelo já é adotado, por exemplo, pelo Governo do Canadá, no âmbito da *Directive on Automated Decision-Making*, que exige que órgãos e agências governamentais preencham o modelo de AIA antes de realizar projetos envolvendo decisões automatizadas (Canadá, 2019).

Por outro lado, existem abordagens que, embora forneçam previamente os tópicos que devem ser avaliados e preenchidos, não estabelecem parametrizações e cálculos de risco previamente. Por exemplo, o Governo da Holanda disponibiliza a *Fundamental Rights and Algorithms Impact Assessment* (FRAIA), que é apresentada como uma ferramenta para auxiliar na tomada de decisões, por parte de organizações governamentais, em situações envolvendo desenvolvimento ou uso de sistemas de IA (Holanda, 2021).

A metodologia proposta pela FRAIA é baseada em três etapas. Na primeira etapa, que representa o momento de preparação, é necessário decidir os motivos pelos quais um sistema de IA será utilizado e quais serão seus possíveis efeitos. Posteriormente, na segunda etapa, é necessário avaliar quais características são esperadas do sistema de IA em questão e quais dados serão utilizados para alimentá-lo. Por fim, a terceira etapa está relacionada ao monitoramento e supervisionamento dos resultados produzidos pelo sistema de IA.

Em relação ao conteúdo, a FRAIA envolve informações sobre o sistema de IA, seu contexto de desenvolvimento ou uso e, ainda, questões mais amplas sobre direitos fundamentais. Desse modo, ainda que existam campos de avaliação previamente definidos para que o agente faça o preenchimento, são campos abertos e sem qualquer tipo de parametrização ou cálculo de risco a partir das respostas inseridas.

Além das três abordagens categorizadas por Selbst (2021), é importante notar a existência de uma categoria adicional, composta por instrumentos autorregulatórios de responsabilidade social corporativa. Entre eles, destaca-se a Avaliação de Impacto Social (*Social Impact Assessment* – SAI) e a Avaliação de Impacto para Direitos Humanos

(*Human Rights Impact Assessments* – HRIA), que têm aplicabilidade, também, no setor privado.

Selbst (2021) elenca, ainda, três aspectos essenciais para a AIA. Em primeiro lugar, ela deve ser compreendida como um instrumento de intervenção antecipada, sendo conduzida em estágios iniciais de projetos envolvendo sistemas de IA. Além disso, o autor entende que uma AIA eficaz deve ser construída por meio de perguntas abertas, ou seja, em vez de questionar se foram feitas verificações específicas, como em uma auditoria, os agentes envolvidos devem explicar suas decisões, apresentar vantagens e desvantagens, identificar riscos e apresentar medidas mitigatórias aplicáveis. Por fim, Selbst (2021) aponta que a AIA deve ser um mecanismo de *accountability*, como, por exemplo, por meio de incentivos regulatórios inteligentes que a tornem útil aos próprios agentes responsáveis pelo desenvolvimento e implementação de sistemas de IA, de modo que desejem cooperar e investir no sucesso da elaboração de uma AIA.

Para Mökander e Floridi (2021), tais avaliações devem ser compreendidas como um processo contínuo, holístico, dialético, estratégico e orientado para o *design*. Isso significa dizer que as avaliações devem monitorar e avaliar continuamente os resultados do sistema e documentar as características de desempenho. A IA não é uma tecnologia isolada, mas faz parte de sistemas sociotécnicos mais amplos; isso exige uma abordagem holística e um processo dialético o qual garante a escolha das perguntas certas.

Na mesma direção, Metcalf, Moss, Watkins, Singh e Elish (2021) apontam que as Avaliações de Impacto Algorítmico são práticas emergentes de governança para delinear a responsabilidade, tornando visíveis os potenciais danos causados pelos sistemas de IA e assegurando que medidas práticas sejam tomadas para amenizar tais possíveis danos. Nesse contexto, os autores ressaltam que os "danos" só se tornam "impactos" em uma relação de *accountability* que obriga os desenvolvedores e implementadores de sistemas a identificar, justificar ou mitigar os danos reais ou potenciais de tais sistemas.

Para Selbst (2021), a avaliação de impacto tem dois objetivos principais. O primeiro é fazer com que os desenvolvedores de sistemas de IA reflitam metodicamente sobre os detalhes e os possíveis impactos de um projeto complexo antes de sua implementação, de modo a prevenir a materialização de riscos e evitar que a correção tenha de ser feita quando o sistema já está totalmente desenvolvido e em uso, o que pode gerar custos e onerosidades. Trata-se de um entendimento derivado da abordagem de *values-in-design*, isto é, quanto mais cedo os valores sociais forem considerados no desenvolvimento do projeto, maior será a probabilidade de que o resultado reflita esses

ideais. O segundo objetivo da AIA é criar e fornecer documentação das decisões tomadas durante o desenvolvimento e implementação de um sistema de IA, bem como suas justificativas, a fim de gerar responsabilização por essas decisões e alcançar informações úteis para futuras intervenções políticas.

Selbst (2021) aponta que as avaliações de impacto são instrumentos úteis diante de projetos com impactos desconhecidos e difíceis de medir. Os responsáveis pelo desenvolvimento de tais projetos são os únicos agentes que possuem visibilidade, conhecimento e experiência para estimar seus impactos. Para o autor, o desenvolvimento e a implementação de sistemas de IA encontra-se nessa mesma situação: sabe-se que há possíveis impactos associados aos sistemas de IA, mas o público não tem as informações ou o conhecimento necessário para se aprofundar e descobrir quais tipos de decisões no projeto ou na utilização do sistema levam a tipos específicos de problemas.

Dada a disparidade de informações entre os agentes responsáveis pelo desenvolvimento ou pela implementação de sistemas de IA e o público afetado por tais sistemas, conforme identificado por Selbst (2021), surge um desafio prático quando se considera que as avaliações serão conduzidas pelas próprias organizações que desenvolvem ou utilizam os sistemas de IA. Isso ocorre porque a experiência e as informações do próprio setor são essenciais para uma avaliação completa dos impactos, o que significa que o setor privado desempenhará um papel fundamental em sua própria governança. Diante desse cenário, o autor argumenta que, para garantir a eficácia da AIA como instrumento de *accountability*, também é importante considerar o ambiente institucional do setor privado.

Tendo em vista os objetivos da AIA, os quais visam considerar os impactos desde o início de projetos envolvendo sistemas de IA, trabalhar para mitigá-los antes do momento de desenvolvimento ou implementação e criar documentação de decisões e testes que possam apoiar o aprendizado futuro, Selbst (2021) observa que as lógicas institucionais do setor privado, essencialmente baseadas na promoção do lucro, podem dificultar a plena realização da primeira meta no curto prazo. No entanto, a AIA ainda pode ser um instrumento benéfico, pois o segundo objetivo — produzir as informações necessárias para uma melhor compreensão dos sistemas de IA — independe da motivação privada, embora possa encontrar obstáculos na proteção do segredo comercial e industrial. Além disso, Selbst (2021) sugere que, com o tempo, a AIA pode fazer parte de uma mudança cultural mais ampla em direção à responsabilidade dentro do setor técnico-privado, aumentando a adesão aos instrumentos de *accountability*.

Considerando que a AIA busca promover a responsabilidade e a prestação de contas, é possível compreendê-la por meio dos elementos “ator” e “fórum”. Neste cenário, um “ator” (desenvolvedor ou responsável pela implementação do sistema de IA) é responsável perante um “fórum” com poderes para emitir julgamentos, exigir mudanças do ator e aplicar penalidades ou medidas equivalentes. Desse modo, para criar regimes eficazes de *accountability* algorítmica, as AIAs precisariam abordar o ator (quem), os fóruns (quando e onde) e o conteúdo (o que), ambos relacionados ao sistema de IA em análise (Metcalf *et al.*, 2021).

Diante da revisão de bibliografia realizada, é possível compreender que a AIA funciona como um método formal, por meio do qual desenvolvedores e responsáveis pela implementação de sistemas de IA demonstram os riscos identificados, bem como as medidas e alterações realizadas para proteção do público que utiliza ou é afetado por sistemas de IA. A *accountability* e a prestação de contas residem, portanto, nas relações entre os agentes que desenvolvem e implementam sistemas de IA, as autoridades regulatórias e o público (Metcalf *et al*., 2023).

As práticas de documentação são um componente-chave de uma estrutura de *accountability*, sendo essenciais para garantir a transparência, rastreabilidade e responsabilização em todas as fases do ciclo de vida de um sistema de IA. No entanto, há uma vulnerabilidade inerente a essas práticas, conforme observado por Selbst (2021) e por Metcalf, Singh, Moss, Tafesse e Watkins (2023): o agente responsável pelo desenvolvimento ou pela implementação do sistema de IA deve, necessariamente, fazer a maior parte ou toda a documentação, permitindo-lhe escolher quais características ou consequências do sistema serão documentadas ou não.

Desse modo, o agente possui o poder de decidir quais aspectos do sistema serão documentados com mais detalhe e quais poderão ser omitidos ou tratados de forma superficial. Isso significa que o responsável pela documentação pode escolher não registrar certas características ou consequências do sistema que possam ser potencialmente negativas ou controvérsias, o que pode resultar em uma avaliação incompleta.

Diante desse cenário, as exigências regulamentares de documentação se tornam particularmente relevantes. Ainda que tais obrigações regulamentares possam não ser necessariamente cumpridas perfeitamente pelos agentes que desenvolvem ou implementam sistemas de IA, há uma mudança na forma como a ignorância e o

conhecimento são incentivados em quadros de governança e mudam a cultura organizacional para práticas compatíveis com *accountability* (Selbst, 2021).

Nesse contexto, também é importante analisar o nível de autonomia que os agentes responsáveis pelo desenvolvimento ou pelo uso de sistemas de IA devem possuir para adaptar o conteúdo da AIA às particularidades de cada caso concreto, de modo a evitar modelos excessivamente engessados e que não produzam os resultados esperados. Tendo em vista a aplicação de sistemas de IA em uma ampla gama de contextos, cada um com suas especificidades e desafios únicos, entende-se que a flexibilidade na elaboração da AIA é relevante para possibilitar adaptações, por parte dos agentes responsáveis, e refletir as nuances de cada contexto.

É relevante que a AIA possa ser conduzida de acordo com práticas-padrão do setor. Essas práticas podem ser desenvolvidas através de mecanismos de autorregulação, onde as próprias organizações ou associações do setor estabelecem diretrizes e padrões de boas práticas. Alternativamente, as práticas-padrão podem ser impostas por uma autoridade regulatória setorial competente, que define critérios e procedimentos específicos para a realização da AIA conforme o caso concreto. Essa abordagem dual – permitindo autonomia, mas dentro de um quadro de práticas-padrão estabelecidas – pode assegurar que a AIA seja conduzida de maneira consistente, ao mesmo tempo em que se mantém suficientemente flexível para ser aplicada de modo efetivo em diferentes contextos.

Além disso, considerando que os agentes de IA são quem efetivamente conhecem os aspectos relacionados às funcionalidades do sistema e ao seu contexto de aplicação, entende-se que é razoável que tenham flexibilidade para escolher os métodos e processos mais adequados e alinhados. Por esse motivo, compreende-se que eventual regulamentação sobre o tema não deve estabelecer regras rígidas, gerais e abstratas sobre o desenvolvimento da AIA, mas indicar diretrizes gerais e objetivas sobre seu desenvolvimento e conteúdo. Desse modo, reconhece-se a importância de que uma futura regulamentação tenha a capacidade de produzir mudanças significativas nas práticas internas da organização, e não apenas rituais regulatórios que sejam insatisfatórios ao objetivo primordial da regulamentação (Metcalf *et al*., 2023).

Por fim, a partir da revisão de literatura acerca das bases teóricas da Avaliação de Impacto Algorítmico, sintetiza-se a seguir as principais conclusões acerca de sua fundamentação:

1. A AIA é um instrumento voltado à identificação, avaliação, mitigação e gestão de riscos associados ao desenvolvimento e uso de sistemas de IA.
2. A AIA deve ser compreendida como um instrumento pautado na prevenção e, por conseguinte, é preferível que seja conduzido desde as etapas iniciais de planejamento de desenvolvimento ou implementação de sistemas de IA.
3. Para que seja efetiva, o desenvolvimento da AIA deve ser acompanhado de um espaço interno ou externo ("fórum") com poderes para emitir julgamentos, exigir mudanças ou aplicar sanções ou medidas equivalentes.
4. Eventual regulamentação sobre metodologias e regras para elaboração da AIA deve fornecer diretrizes gerais que permitam flexibilidade e adaptação às particularidades de cada caso, garantindo práticas eficazes e relevantes, ao invés de estabelecer regras rígidas que possam limitar sua aplicação e impacto.

4 AVALIAÇÃO DE IMPACTO ALGORÍTMICO NO CONTEXTO DE DEBATES REGULATÓRIOS

O debate em torno da inteligência artificial tem sido dominado por uma ênfase na dimensão ética do desenvolvimento e na aplicação de sistemas de IA, com especial atenção para a presença de vieses e o risco de discriminação. Para Mantelero (2022), o debate no campo da IA passou por uma construção que retirou o tema da regulação e focou na ética, sendo importante considerar dois estágios diferentes e consecutivos nesse contexto: o debate acadêmico e as iniciativas institucionais.

O autor aponta que o debate acadêmico sobre a ética das máquinas faz parte de uma reflexão mais ampla e antiga sobre ética e tecnologia; por outro lado, as iniciativas institucionais são mais recentes, têm natureza não acadêmica e visam avançar o debate regulatório. Mantelero (2022) ressalta que a passagem da análise teórica para a arena política representa uma grande mudança, de modo que a mensagem dos reguladores para o ambiente de tecnologia explicitava a importância não só da lei, mas também da ética.

Nesse contexto, Mantelero (2022) destaca que os documentos produzidos por especialistas nomeados por reguladores e formuladores de políticas, geralmente, são minimalistas em termos de estrutura teórica, concentrando-se na mensagem política sobre a relevância da dimensão ética. Desse modo, as questões éticas são adicionadas em tais instrumentos (caracterizados, em geral, como *soft law*), mas a tarefa de investigação aprofundada e aplicação desses parâmetros são deixadas aos destinatários dos documentos. Ao considerar o impacto de sistemas de IA para direitos humanos, a abordagem dominante em muitos documentos de *soft law* se concentra na listagem de direitos e liberdades potencialmente afetados, mas não propõe modelos de avaliação de tais impactos (Mantelero, 2022).

Nos últimos anos, no entanto, observou-se um aumento de propostas baseadas em *hard law*, refletindo a crescente preocupação global em estabelecer diretrizes regulatórias sólidas para lidar com os desafios éticos, legais e sociais associados aos sistemas de IA. Tais propostas buscam criar um arcabouço jurídico vinculativo e, por vezes, abrangem a obrigação de elaboração de instrumentos como a Avaliação de Impacto Algorítmico.

4.1 A EVOLUÇÃO DA AVALIAÇÃO DE IMPACTO ALGORÍTMICO NO DEBATE REGULATÓRIO BRASILEIRO

No contexto do debate regulatório brasileiro, verifica-se que a avaliação de impacto regulatório foi tratada com maior nível de profundidade e detalhe no PL n.º 2.338/23, que apresenta seção específica acerca do tema. A tabela abaixo consolida o desenvolvimento das disposições sobre avaliação de impacto em projetos de lei sobre inteligência artificial no Brasil.

Quadro 1 - Disposições acerca de avaliação de impacto de sistemas de IA em projetos de lei em tramitação no Brasil

Projeto de Lei	Disposições acerca de avaliação de impacto de sistemas de IA
Projeto de Lei n.º 5051, de 2019	O PL n.º 5051/2019 estabelece princípios para o uso da inteligência artificial no Brasil. Trata-se de proposta legislativa com teor essencialmente principiológico, de modo que não há menção a uma obrigação de realização de avaliação de impacto de sistemas de IA.
Projeto de Lei n.º 21, de 2020	O texto original do PL n.º 21/2020, que estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, determina, em seu art. 13, que a União, os Estados, o Distrito Federal e os Municípios poderão solicitar aos agentes dos sistemas de inteligência artificial, observadas as suas funções e justificada a necessidade, a publicação de relatórios de impacto de inteligência artificial e recomendar a adoção de padrões e de boas práticas para implantação e operação dos sistemas. Nesse sentido, o relatório de impacto de inteligência artificial é definido, pelo texto original, como a documentação dos agentes de inteligência artificial que contém a descrição do ciclo de vida do sistema de inteligência artificial, bem como medidas, salvaguardas e mecanismos de gerenciamento e mitigação dos riscos relacionados a cada fase do sistema, incluindo segurança e privacidade. Posteriormente, após apresentação do Parecer Preliminar de Plenário e de texto substitutivo ao PL, pela relatora Deputada Luisa Canziani (PTB/PR), as disposições acerca da realização de relatório de impacto de inteligência artificial foram suprimidas.
Projeto de Lei n.º 872, de 2021	O PL n.º 872/2021, que dispõe sobre o uso da inteligência artificial, apresenta diretrizes essencialmente principiológicas, não havendo menção a uma obrigação de realização de avaliação de impacto de sistemas de IA.
Projeto de Lei n.º 2.338, de 2023	O PL n.º 2.338/2023 conta com seção específica acerca da realização de Avaliação de Impacto Algorítmico, dispondo acerca do tema de forma ampla e detalhada. Além disso, a Complementação de Voto apresentada pelo relator do Projeto de Lei à Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA), em junho de 2024, manteve a Avaliação de Impacto Algorítmico em sua proposta de texto substitutivo.

Fonte: Elaborado pela autora (2024).

Conforme redação original do PL n.º 2.338/2023, sempre que o sistema for considerado como de alto risco pela avaliação preliminar, a realização da Avaliação de Impacto Algorítmico seria uma obrigação dos agentes de inteligência artificial. Acerca desse tema, é necessário esclarecer que a proposta brasileira prevê duas classificações: risco excessivo e alto risco.

Sistemas de risco excessivo são aqueles de desenvolvimento, implementação e uso vedado e incluem, conforme relatório final que apresenta texto substitutivo ao PL n.º 2.338/2023, os sistemas com propósito de (i) induzir o comportamento da pessoa natural ou de grupos de maneira que cause danos à saúde, segurança ou outros direitos fundamentais próprios ou de terceiros; (ii) explorar quaisquer vulnerabilidades de pessoa natural ou de grupos com o objetivo ou o efeito de induzir o seu comportamento de maneira que cause danos à saúde, segurança ou outros direitos fundamentais próprios ou de terceiros; (iii) possibilitar a produção, disseminação ou facilitar a criação de material que caracterize ou represente abuso ou exploração sexual de crianças e adolescentes; e (iv) avaliar os traços de personalidade, as características ou o comportamento passado, criminal ou não, de pessoas singulares ou grupos, para avaliação de risco de cometimento de crime, infrações ou de reincidência.

Além disso, texto substitutivo ao PL n.º 2.338/2023 propõe que sejam vedados o desenvolvimento, a implementação e o uso de sistemas de IA, pelo poder público, para avaliar, classificar ou ranquear as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, por meio de pontuação universal, para o acesso a bens e serviços e políticas públicas, de forma ilegítima ou desproporcional. Ainda, seriam vedados sistemas de armas autônomas (SAA) e sistemas de identificação biométrica à distância, em tempo real e em espaços acessíveis ao público, com exceção das hipóteses de (i) instrução de inquérito ou processo criminal, mediante autorização judicial prévia e motivada, quando houver indícios razoáveis da autoria ou participação em infração penal, a prova não puder ser feita por outros meios disponíveis e o fato investigado não constitua infração penal de menor potencial ofensivo; (ii) busca de vítimas de crimes, de pessoas desaparecidas ou em circunstâncias que envolvam ameaça grave e iminente à vida ou à integridade física de pessoas naturais; (iii) flagrante delito de crimes punidos com pena privativa de liberdade superior a 2 (dois) anos, com imediata comunicação à autoridade judicial; e (iv) recaptura de réus evadidos, cumprimento de mandados de prisão e de medidas restritivas ordenadas pelo Poder Judiciário.

Por outro lado, de acordo com o texto substitutivo apresentado, os sistemas classificados como de alto risco são determinados a partir de uma lista de finalidades e contextos, levando-se em conta a probabilidade e a gravidade dos impactos adversos sobre pessoa ou grupos afetados, que inclui: (i) aplicação como dispositivos de segurança na gestão e no funcionamento de infraestruturas críticas, tais como controle de trânsito e redes de abastecimento de água e de eletricidade, quando houver risco relevante à integridade física das pessoas e à interrupção de serviços essenciais, de forma ilícita ou abusiva; (ii) educação, formação profissional para a determinação de acesso a instituições de ensino ou de formação profissional ou para avaliação e monitoramento de estudantes; (iii) recrutamento, triagem, filtragem, avaliação de candidatos, tomada de decisões sobre promoções ou cessações de relações contratuais de trabalho, repartição de tarefas e controle e avaliação do desempenho e do comportamento das pessoas afetadas por tais aplicações de IA nas áreas de emprego, gestão de trabalhadores e acesso ao emprego por conta própria; (iv) avaliação de critérios de acesso, elegibilidade, concessão, revisão, redução ou revogação de serviços privados e públicos que sejam considerados essenciais, incluindo sistemas utilizados para avaliar a elegibilidade de pessoas naturais quanto a prestações de serviços públicos de assistência e de seguridade; (v) avaliação e classificação de chamadas, ou determinação de prioridades para serviços públicos essenciais, tais como de bombeiros e assistência médica; (vi) administração da justiça, no que toca o uso sistemas que auxiliem autoridades judiciárias em investigação dos fatos e na aplicação da lei quando houver risco às liberdades individuais e ao Estado democrático de direito, excluindo-se os sistemas que auxiliem atos e atividades administrativas; (vii) veículos autônomos em espaços públicos, quando seu uso puder gerar risco relevante à integridade física de pessoas; (viii) aplicações na área da saúde para auxiliar diagnósticos e procedimentos médicos, quando houver risco relevante à integridade física e mental das pessoas; (ix) estudo analítico de crimes relativos a pessoas naturais, permitindo às autoridades policiais pesquisar grandes conjuntos de dados, disponíveis em diferentes fontes de dados ou em diferentes formatos, no intuito de identificar padrões e perfis comportamentais; (x) investigação por autoridades administrativas para avaliar a credibilidade dos elementos de prova no decurso da investigação ou repressão de infrações, para prever a ocorrência ou a recorrência de uma infração real ou potencial com base na definição de perfis de pessoas singulares; (xi) sistemas de identificação e autenticação biométrica para o reconhecimento de emoções, excluindo-se os sistemas de autenticação biométrica cujo único objetivo seja a confirmação de uma pessoa singular

específica; (xii) gestão da imigração e controle de fronteiras para avaliar o ingresso de pessoa ou grupo de pessoas em território nacional; e (xiii) produção, curadoria, difusão, recomendação e distribuição, em grande escala e significativamente automatizada, de conteúdo por provedores de aplicação, com objetivo de maximização do tempo de uso e engajamento das pessoas ou grupos afetados, quando o funcionamento desses sistemas puder representar riscos relevantes aos fundamentos.

É importante notar que a proposta de substitutivo estabelece que não serão considerados de alto risco, quando utilizadas para as finalidades listadas anteriormente, as tecnologias intermediárias que não influenciam ou determinem resultado ou decisão. Além disso, o desenvolvedor e o aplicador que considerar que o sistema de IA não se enquadra na classificação de alto risco apresentada poderá apresentar petição fundamentada às autoridades competentes, juntamente com sua avaliação preliminar, conforme será detalhado em regulamento.

A proposta de substitutivo apresentada estabelece que caberia ao Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA) regulamentar a classificação da lista dos sistemas de IA de alto risco. O SIA é a proposta de criação de um ecossistema regulatório coordenado pela autoridade competente, que teria o objetivo de promover e garantir a cooperação e a harmonização com as demais autoridades setoriais e órgãos reguladores, sem vínculo de subordinação hierárquica entre eles, e outros sistemas nacionais para a plena implementação e fiscalização do cumprimento da norma. Por sua vez, a autoridade competente é definida como uma entidade da administração pública federal, dotada de autonomia técnica e decisória.

Nesse sentido, caberia ao SIA identificar novas hipóteses de sistemas de IA de alto risco, levando em consideração a probabilidade e a gravidade dos impactos adversos sobre pessoa ou grupos afetados e ao menos um dos seguintes critérios: (i) a implementação ser em larga escala, levando-se em consideração o número estimado de pessoas afetadas e a extensão geográfica, bem como a sua duração e frequência do uso; (ii) o sistema produzir, de forma ilícita ou abusiva, efeitos jurídicos relevantes e impactar negativamente o acesso a serviços públicos ou essenciais; (iii) alto potencial danoso de ordem material ou moral, bem como viés discriminatório ilegal; (iv) o sistema afetar significativamente pessoas de um grupo vulnerável; (v) o nível de irreversibilidade dos danos; (vi) histórico danoso, de ordem material ou moral; (vii) grau de transparência, explicabilidade e auditabilidade do sistema de inteligência artificial, que dificulte significativamente o seu controle ou supervisão; (viii) alto potencial danoso sistêmico,

tais como à segurança cibernética, higidez do processo eleitoral e violência contra grupos vulneráveis; (ix) extensão e probabilidade dos benefícios do sistema de inteligência artificial, incluindo as medidas de mitigação dos riscos adotadas e possíveis melhorias, de acordo com os princípios e fundamentos desta lei; (x) riscos significativos à saúde humana integral – física, mental e social – nas dimensões individual e coletiva; (xi) risco à integridade da informação, o processo democrático e pluralismo, como, por exemplo, através da disseminação de desinformação e de discursos que promovam o ódio ou a violência; e (xii) possibilidade de impactar negativamente o desenvolvimento e a integridade física, psíquica ou moral de crianças e adolescentes.

Além disso, de acordo com a proposta de substitutivo, cabe à autoridade competente expedir orientações normativas gerais em relação aos impactos dos sistemas de inteligência artificial sobre os direitos e liberdades fundamentais ou que produzam efeitos jurídicos relevantes e às autoridades setoriais, no âmbito de suas atribuições e em caráter prevalente aos órgãos do SIA, dispor sobre os aspectos técnicos e específicos de aplicações de IA no mercado regulado, inclusive podendo estabelecer listas sobre hipóteses classificadas ou não classificadas como de alto risco e precisar o rol exemplificativo de sistemas de alto risco.

Especificamente acerca da vinculação da obrigatoriedade de elaboração da AIA à classificação do sistema de IA em questão como de alto risco, entende-se que se trata de uma vinculação acertada. Primeiramente, ao estabelecer essa conexão, o legislador reconhece a necessidade de direcionar recursos e esforços regulatórios para os sistemas de IA que apresentam maior potencial de causar danos significativos ou impactos adversos na sociedade. Ademais, ao limitar a obrigatoriedade da AIA aos sistemas de IA classificados como de alto risco, evita-se ônus excessivo para sistemas de baixo risco ou de menor complexidade. Dessa forma, é possível promover um ambiente regulatório equilibrado, que não sobrecarregue desenvolvedores e aplicadores de sistemas de IA com exigências desproporcionais.

Em síntese, ao estabelecer a obrigatoriedade somente em casos de alto risco, é possível garantir que os recursos e esforços sejam direcionados de forma mais eficaz e proporcional aos sistemas que apresentam os maiores desafios e impactos potenciais. Isso não impediria que, a título de boa prática, um agente tenha autonomia para conduzir uma AIA em casos cujo sistema não seja de risco alto, mas desperte preocupações éticas, jurídicas ou técnicas. Essa prática, mesmo não sendo obrigatória, pode ser adotada como

uma medida proativa de boa governança, permitindo que o agente demonstre diligência e responsabilidade ao abordar potenciais riscos de maneira preventiva.

Desse modo, ainda que o Projeto de Lei nº. 2.338/2023 não avance na esfera legislativa e o arranjo regulatório de classificação de riscos de sistemas de IA apresentado anteriormente não se torne uma realidade no ordenamento jurídico brasileiro, ressalta-se que a AIA pode ainda ser adotada como uma boa prática de mercado. Nesse sentido, mesmo em um cenário onde não exista uma obrigação legal específica, entende-se que a AIA deve ser elaborada por desenvolvedores e implementadores de sistemas de IA sempre que se verificar, na prática, que o desenvolvimento ou implementação daquele sistema apresenta alto risco. Isso se justifica pelo fato de que, mesmo sem uma classificação legal definida para sistemas de IA de alto risco, os potenciais impactos sobre indivíduos e grupos afetados devem ser considerados, de modo a antecipar e mitigar eventuais impactos que possam representar violações à direitos humanos.

Em relação ao debate sobre a AIA enquanto instrumento regulado, a tabela comparativa a seguir busca apresentar as diferentes redações do Projeto de Lei n. 2.338/2023 durante sua tramitação na Comissão Temporária de Inteligência Artificial (CTIA) do Senado Federal.

Quadro 2 – Comparativo entre a redação original do Projeto de Lei n.º 2.338/2023, o Texto Substitutivo Preliminar apresentado em abril de 2024 e o Texto Substitutivo Final apresentado em julho de 2024

Projeto de Lei n.º 2.338/2023 (redação original)	Relatório Preliminar da CTIA (texto apresentado em 24.04.2024)	Substitutivo Final da CTIA (texto apresentado em 04.07.2024)
Art. 22. A avaliação de impacto algorítmico de sistemas de inteligência artificial é obrigação dos agentes de inteligência artificial, sempre que o sistema for considerado como de alto risco pela avaliação preliminar. Parágrafo único. A autoridade competente será notificada sobre o sistema de alto risco, mediante o compartilhamento das avaliações preliminar e de impacto algorítmico.	Art. 22. A avaliação de impacto algorítmico de sistemas de inteligência artificial é obrigação dos agentes de inteligência artificial, sempre que o sistema for considerado de alto risco pela avaliação preliminar, nos termos do art. 12 desta Lei. Parágrafo único. Os agentes de inteligência artificial deverão compartilhar com a autoridade competente as avaliações preliminares e	Art. 25. A avaliação de impacto algorítmico de sistemas de IA é obrigação do desenvolvedor e aplicador, sempre que o sistema for considerado de alto risco pela avaliação preliminar, nos termos do art. 12 desta Lei. § 1º Os desenvolvedores de sistemas de IA deverão compartilhar com as autoridades competentes as avaliações preliminares e de impacto algorítmico, nos termos do

	de impacto algorítmico, nos termos do regulamento.	regulamento, cuja metodologia considerará e registrará, ao menos, avaliação dos riscos e benefícios aos direitos fundamentais, medidas de atenuação e efetividade destas medidas de gerenciamento. § 2º Caberá às autoridades setoriais definir as hipóteses em que avaliação de impacto algorítmico será simplificada, observado o papel de cada um dos agentes de IA e as normas gerais da autoridade competente. § 3º Quando da utilização de sistemas IA que possam gerar impactos irreversíveis ou de difícil reversão, a avaliação de impacto algorítmico levará em consideração também as evidências incipientes. § 4º A autoridade competente, a partir das diretrizes do Conselho Permanente de Cooperação Regulatória (CRIA), estabelecerá critérios gerais e elementos para a elaboração de avaliação de impacto e a periodicidade de atualização das avaliações de impacto; § 5º Caberá às autoridades setoriais, a partir do estado da arte do desenvolvimento tecnológico e melhores práticas, a regulamentação dos critérios e da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de IA de alto risco. § 6º Os agentes de IA que, posteriormente à sua

		introdução no mercado ou utilização em serviço, tiverem conhecimento de risco ou impacto inesperado e relevante que apresentem a direitos de pessoas naturais, comunicará o fato imediatamente às autoridades competentes e às pessoas afetadas pelo sistema de IA.
Art. 23. A avaliação de impacto algorítmico será realizada por profissional ou equipe de profissionais com conhecimentos técnicos, científicos e jurídicos necessários para realização do relatório e com independência funcional. Parágrafo único. Caberá à autoridade competente regulamentar os casos em que a realização ou auditoria da avaliação de impacto será necessariamente conduzida por profissional ou equipe de profissionais externos ao fornecedor;	Art. 23. A avaliação de impacto algorítmico será realizada por profissional ou equipe de profissionais com independência funcional, bem como com conhecimentos técnicos, científicos, regulatórios e jurídicos necessários e considerando as boas práticas setoriais e internacionais. § 1º Caberá à autoridade competente regulamentar: a) os critérios quanto à independência funcional referida no caput; b) os casos em que a realização ou auditoria da avaliação de impacto será necessariamente conduzida por profissional ou equipe de profissionais externos ao fornecedor.	Não aplicável (dispositivo removido da proposta de redação final).
Art. 24. A metodologia da avaliação de impacto conterá, ao menos, as seguintes etapas: I – preparação; II – cognição do risco; III – mitigação dos riscos encontrados; IV – monitoramento. § 1º A avaliação de impacto considerará e registrará, ao menos: a) riscos conhecidos e previsíveis associados ao	Art. 24. A metodologia da avaliação de impacto conterá, ao menos, as seguintes etapas: I – cognição do risco; II – a mitigação dos riscos encontrados; III − monitoramento. § 1º A avaliação de impacto considerará e registrará, ao menos: a) riscos a direitos fundamentais individuais e sociais conhecidos e	Não aplicável (dispositivo removido da proposta de redação final).

sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar; b) benefícios associados ao sistema de inteligência artificial; c) probabilidade de consequências adversas, incluindo o número de pessoas potencialmente impactadas; d) gravidade das consequências adversas, incluindo o esforço necessário para mitigá-las; e) lógica de funcionamento do sistema de inteligência artificial; f) processo e resultado de testes e avaliações e medidas de mitigação realizadas para verificação de possíveis impactos a direitos, com especial destaque para potenciais impactos discriminatórios; g) treinamento e ações de conscientização dos riscos associados ao sistema de inteligência artificial; h) medidas de mitigação e indicação e justificação do risco residual do sistema de inteligência artificial, acompanhado de testes de controle de qualidade frequentes; e i) medidas de transparência ao público, especialmente aos potenciais usuários do sistema, a respeito dos riscos residuais, principalmente quando envolver alto grau de nocividade ou periculosidade à saúde ou segurança dos usuários,	previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar; b) benefícios associados ao sistema de inteligência artificial; c) probabilidade e gravidade de consequências adversas, incluindo o número de pessoas potencialmente impactadas e o esforço necessário para mitigá-las; d) lógica de funcionamento do sistema de inteligência artificial; e) treinamento e ações de conscientização dos riscos associados ao sistema de inteligência artificial; e f) medidas de transparência ao público, especialmente aos potenciais usuários do sistema, a respeito dos riscos residuais, principalmente quando envolver alto grau de nocividade ou periculosidade à saúde ou segurança dos usuários, nos termos dos artigos 9º e 10 da Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor). § 1º Em atenção ao princípio da precaução, quando da utilização de sistemas de inteligência artificial que possam gerar impactos irreversíveis ou de difícil reversão, a avaliação de impacto algorítmico levará em consideração também as evidências incipientes,	

nos termos dos artigos 9º e 10 da Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor). § 2º Em atenção ao princípio da precaução, quando da utilização de sistemas de inteligência artificial que possam gerar impactos irreversíveis ou de difícil reversão, a avaliação de impacto algorítmico levará em consideração também as evidências incipientes, incompletas ou especulativas. § 3º A autoridade competente poderá estabelecer outros critérios e elementos para a elaboração de avaliação de impacto, incluindo a participação dos diferentes segmentos sociais afetados, conforme risco e porte econômico da organização. § 4º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de inteligência artificial de alto risco e os campos de aplicação, podendo incorporar melhores práticas setoriais. § 5º Os agentes de inteligência artificial que, posteriormente à sua introdução no mercado ou utilização em serviço, tiverem conhecimento de risco inesperado que apresentem a direitos de pessoas naturais, comunicará o fato	incompletas ou especulativas. § 2º à autoridade competente, em colaboração com as demais entidades do SIA, poderá estabelecer outros critérios e elementos para a elaboração de avaliação de impacto e a periodicidade de atualização das avaliações de impacto; § 3º Considerando eventual regulamentação setorial existente, caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de inteligência artificial de alto risco e podendo incorporar melhores práticas setoriais. § 5º Os agentes de inteligência artificial que, posteriormente à sua introdução no mercado ou utilização em serviço, tiverem conhecimento de risco inesperado e relevante que apresentem a direitos de pessoas naturais, comunicará o fato imediatamente às autoridades competentes e às pessoas afetadas pelo sistema de inteligência artificial.	

imediatamente às autoridades competente e às pessoas afetadas pelo sistema de inteligência artificial.		
Art. 25. A avaliação de impacto algorítmico consistirá em processo iterativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de inteligência artificial de alto risco, requeridas atualizações periódicas. § 1º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto. § 2º A atualização da avaliação de impacto algorítmico contará também com participação pública, a partir de procedimento de consulta a partes interessadas, ainda que de maneira simplificada.	Art. 25. A elaboração da avaliação de impacto deve, sempre que possível, conforme risco e porte econômico da organização, incluir a participação pública efetiva dos diferentes segmentos sociais afetados, especialmente de grupos vulneráveis potencialmente afetados pelos sistemas. Parágrafo único. Caberá à autoridade competente estabelecer as hipóteses em que a participação pública referida no caput será dispensada, assim como as hipóteses em que poderá ser realizada de maneira simplificada, indicando os critérios para esta participação.	Art. 26. A elaboração da avaliação de impacto incluirá, conforme risco e porte econômico da organização, a participação pública dos diferentes segmentos sociais afetados, especialmente de grupos vulneráveis potencialmente afetados pelos sistemas, nos termos do regulamento Parágrafo único. Caberá às autoridades competentes estabelecer as hipóteses em que a participação pública referida no caput será dispensada, assim como as hipóteses em que poderá ser realizada de maneira simplificada, indicando os critérios para esta participação.
Não aplicável.	Art. 26. A avaliação de impacto algorítmico consistirá em processo iterativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de inteligência artificial de alto risco, requeridas atualizações periódicas. § 1º Considerando eventual regulamentação setorial existente, caberá à autoridade competente a regulamentação, em colaboração com as demais entidades do SIA, definir parâmetros gerais acerca da periodicidade de atualização das avaliações de impacto que deve, ao menos, ser realizada	Art. 27. A avaliação de impacto algorítmico consistirá em processo interativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de IA de alto risco, requeridas atualizações periódicas. Parágrafo Único Considerando eventual regulamentação setorial existente, caberá à autoridade competente, em colaboração com as demais entidades do SIA, definir: I - parâmetros gerais acerca da periodicidade de atualização das avaliações de impacto que deve, ao menos, ser realizada

	quando da existência de alterações significativas nos sistemas. § 2º A atualização da avaliação de impacto algorítmico contará também com participação pública, a partir de procedimento de consulta a partes interessadas, ainda que de maneira simplificada.	quando da existência de alterações significativas nos sistemas; e II - definir as hipóteses em que a avaliação de impacto algorítmico será simplificada, considerando o tipo de agentes de sistemas de IA.
Não aplicável.	Art. 27. Caso o agente de IA tenha que elaborar relatório de impacto à proteção de dados pessoais, nos termos da Lei nº 13.709, de 14 de agosto de 2018, a avaliação de impacto algorítmico poderá ser realizada em conjunto com o referido documento, que pode ser publicado sob a forma de anexo.	Art. 28. Caso o agente de IA tenha que elaborar relatório de impacto à proteção de dados pessoais, nos termos da Lei nº 13.709, de 14 de agosto de 2018, a avaliação de impacto algorítmico poderá ser realizada em conjunto com o referido documento.
Art. 26. Garantidos os segredos industrial e comercial, as conclusões da avaliação de impacto serão públicas, contendo ao menos as seguintes informações: I – descrição da finalidade pretendida para a qual o sistema será utilizado, assim como de seu contexto de uso e escopo territorial e temporal; II – medidas de mitigação dos riscos, bem como o seu patamar residual, uma vez implementada tais medidas; e III – descrição da participação de diferentes segmentos afetados, caso tenha ocorrido, nos termos do § 3º do art. 24 desta Lei.	Art. 28. As conclusões da avaliação de impacto serão públicas, observados os segredos industrial e comercial, nos termos do regulamento.	Art. 29. As conclusões da avaliação de impacto serão públicas, observados os segredos industrial e comercial, nos termos do regulamento.

Fonte: Elaborado pela autora (2024).

Nota-se, portanto, que a redação original do PL n.º 2.338/2023 estabelecia, de modo amplo, que os "agentes de inteligência artificial" deveriam compartilhar com a autoridade competente as avaliações preliminares e de impacto algorítmico. O texto substitutivo final, por sua vez, estabelece que essa é uma obrigação específica dos desenvolvedores de sistema de IA. Entende-se que referida alteração pode ter sido motivada por considerações práticas e operacionais, uma vez que os desenvolvedores de sistemas de IA estão em melhor posição para fornecer informações detalhadas sobre o funcionamento antes mesmo de sua colocação em mercado ou em serviço.

Além disso, a redação original estabelecia que a AIA deve ser realizada por profissional ou equipe de profissionais com os conhecimentos técnicos, científicos e jurídicos necessários para realização do relatório e com independência funcional, cabendo à autoridade competente regulamentar os casos em que a realização ou auditoria da avaliação de impacto será necessariamente conduzida por profissional ou equipe de profissionais externos ao fornecedor. O texto substitutivo apresentado retira referida obrigação, possivelmente porque tal exigência poderia representar um desafio em termos de disponibilidade de recursos humanos qualificados e especializados, bem como aumentar os custos e a complexidade do processo de realização da avaliação.

A proposta legislativa original também apresentava regras detalhadas acerca da metodologia para realização da AIA, estabelecendo etapas, critérios e parâmetros a serem considerados. Ocorre que, embora a redação original do projeto de lei estabelecesse que a AIA deveria registrar os riscos associados ao sistema de IA, não havia uma previsão expressa para avaliação dos impactos do sistema de IA para os direitos fundamentais das pessoas e grupos afetados. A ausência de menção expressa acerca da necessidade de avaliação do sistema de IA sob a perspectiva de proteção de direitos fundamentais pode fazer com que o escopo de abrangência da AIA seja significativamente reduzido aos riscos técnicos do sistema de IA, deixando de lado a ênfase nos direitos das pessoas e grupos afetados.

Nessa direção, destaca-se que o relatório preliminar apresentado para discussão na Comissão Temporária Interna sobre Inteligência Artificial no Brasil propôs expressamente que a Avaliação de Impacto Algorítmico registre os riscos aos direitos fundamentais conhecidos e previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar.

Ocorre que, em razão da apresentação de emendas no âmbito da CTIA, as disposições que estabeleciam regras específicas acerca da metodologia de desenvolvimento da AIA foram removidas. Consequentemente, o texto final do substitutivo não contempla o dispositivo que estabelecia expressamente que a AIA deveria considerar e registrar os riscos a direitos fundamentais individuais e sociais conhecidos e previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar.

Entende-se que a remoção de dispositivos específicos sobre metodologia na legislação é acertada, uma vez que possibilita que os próprios agentes desenvolvam metodologias adequadas às suas particularidades, contexto e recursos disponíveis. No entanto, considera-se importante que eventual legislação sobre o tema defina o escopo geral da Avaliação de Impacto Algorítmico (AIA), o que implica estabelecer que a AIA se destina à avaliação dos impactos sobre direitos humanos. Isso garantiria que, independentemente da metodologia adotada, a AIA se mantivesse como um instrumento voltado à proteção e à promoção de direitos humanos.

De toda forma, nota-se que o texto substitutivo define a Avaliação de Impacto Algorítmico como uma análise do impacto sobre os direitos fundamentais, que apresenta medidas preventivas, mitigadoras e de reversão dos impactos negativos, bem como medidas potencializadoras dos impactos positivos de um sistema de IA. Desse modo, a partir da definição adotada, ainda é possível vincular a AIA ao objetivo primordial de avaliação dos impactos sobre direitos fundamentais.

Por fim, destaca-se que ao redação original do PL n.º 2.338/2023 determinava que as conclusões da avaliação de impacto serão públicas, de modo que, no mínimo, as seguintes informações deverão ser divulgadas, garantidos os segredos industrial e comercial: (i) descrição da finalidade pretendida, assim como de seu contexto de uso e escopo territorial e temporal; (ii) descrição das medidas de mitigação dos riscos, bem como o seu patamar residual, uma vez implementada tais medidas; e (iii) descrição da participação de diferentes segmentos afetados, caso tenha ocorrido.

Desde o relatório preliminar divulgado no âmbito da CTIA, referida disposição foi reduzida, de modo que, atualmente, o substitutivo proposto prevê a obrigatoriedade de publicização das conclusões da avaliação de impacto, observados os segredos industrial e comercial e conforme regulamento futuro. De toda forma, permanece a previsão de que caberá à autoridade competente a criação e manutenção de base de dados de inteligência artificial de alto risco, acessível ao público, que disponibilizará os

documentos públicos das avaliações de impacto, respeitados os segredos comercial e industrial e legislação pertinente, conforme regulamento a ser desenvolvido.

De modo geral, nota-se que o avanço do PL n.º 2.338/2023 deixou certas questões polêmicas para regulamentação futura por parte da autoridade competente. Essa escolha reflete a intenção de conferir maior adaptabilidade à norma, permitindo que questões complexas sejam tratadas em maior detalhe conforme se tornem mais bem compreendidas. A delegação de poder normativo à autoridade competente cria, assim, uma estrutura aberta que permite ajustes contínuos e uma resposta rápida às mudanças, o que é fundamental para a eficácia e longevidade da norma.

Contudo, a transferência de temas para regulamentação futura pela autoridade competente também levanta questões relacionadas à falta de um debate parlamentar mais amplo sobre certos temas, o que também pode limitar o próprio papel das autoridades competentes, que, sem orientações claras e consistentes, podem enfrentar dificuldades ao tomar decisões de impacto amplo ou que exigem um alto grau de aceitação social. Temas complexos podem ser beneficiar de um debate plural, envolvendo não apenas especialistas e reguladores, mas também o Poder Legislativo e a sociedade civil, de modo a garantir que a regulamentação reflita os valores e interesses coletivos.

Desse modo, entende-se que temas de grande impacto social e econômico demandam o envolvimento de uma esfera regulatória mais ampla e representativa, capaz de assegurar a correspondência entre o avanço regulatório e as demandas da sociedade. Por isso, aproveitar o momento atual de debate regulatório é essencial para examinar com profundidade e cuidado as questões que foram minimizadas ao longo das diferentes versões do projeto de lei, mas que são essenciais, especialmente, (i) os parâmetros para participação pública no processo de desenvolvimento da AIA; e (ii) os critérios para transparência e publicização das conclusões da AIA.

Portanto, até o presente momento, infere-se que as alterações propostas no texto substitutivo do PL n.º 2.338/2023 refletem uma tentativa de equilibrar a necessidade de uma regulamentação robusta e eficaz com a realidade prática do desenvolvimento e implementação de sistemas de IA. A inclusão explícita da necessidade de registrar os riscos aos direitos fundamentais no texto substitutivo destaca uma evolução significativa em direção a uma regulamentação mais sensível aos direitos humanos, o que pode garantir uma abordagem mais efetiva na promoção e proteção de direitos de pessoas e grupos afetados.

4.2 A AVALIAÇÃO DE IMPACTO E CONFORMIDADE DE SISTEMAS DE IA NO CONTEXTO EUROPEU

Feitas as considerações acerca da proposta regulatória brasileira, é importante explorar, também, o debate regulatório europeu, que ainda pode impactar a discussão sobre regulamentação de inteligência artificial no Brasil. O *AI Act* estabelece regras para definir sistemas de IA de alto risco e determina que tais sistemas terão que cumprir um conjunto de requisitos obrigatórios, bem como seguir procedimentos de avaliação de conformidade (*conformity assessment*) antes de serem colocados no mercado da União Europeia.

Acerca da classificação de riscos de sistemas de IA, vale ressaltar que, no âmbito da regulação europeia, existem práticas de IA proibidas, sistemas de risco elevado e modelos de IA de finalidade geral que podem carregar "risco sistêmico". Os modelos de IA de finalidade geral são caracterizados pela generalidade e pela capacidade de desempenhar, com competência, uma vasta gama de funções, podendo ser colocados no mercado de formas distintas, como por meio de bibliotecas e interfaces de programação de aplicações.

Vale destacar que, de acordo com o texto do *AI Act*, os modelos de IA não constituem, por si só, sistemas de IA, pois exigem a adição de outros componentes, como, por exemplo, uma interface de utilizador, para se tornarem sistemas de IA. Assim, um sistema de IA de finalidade geral é entendido como um sistema de IA baseado num modelo de IA de finalidade geral, com a capacidade de servir para diversas finalidades, tanto para utilização direta como para integração em outros sistemas de IA.

Nesse sentido, os modelos de IA de finalidade geral são classificados em modelos com risco sistêmico e sem risco sistêmico. O risco sistêmico é compreendido como um risco específico, gerado pelo potencial elevado de impacto do modelo, devido ao seu alcance ou a efeitos negativos reais e razoavelmente previsíveis na saúde e segurança pública, nos direitos fundamentais e na sociedade, o qual pode propagar em escala ao longo da cadeia de valor.

É importante notar que os modelos de IA de finalidade geral impactaram significativamente a estratégia de classificação de riscos anteriormente adotada, dada sua natureza transversal e capacidade de adaptação para diversas finalidades. Ao contrário de sistemas especializados, que possuem aplicações mais restritas e previsíveis, a IA generativa pode ser empregada em uma variedade de contextos. Nessa dinâmica, verifica-

se que o usuário desempenha um papel de destaque, já que o uso dos modelos de IA de finalidade geral influencia diretamente o nível de risco envolvido, colocando maior ênfase na governança responsável e na necessidade de mecanismos de controle adequados ao longo da cadeia de utilização.

Por sua vez, são práticas proibidas aquelas que empreguem técnicas subliminares, alheias à consciência de uma pessoa, ou técnicas propositalmente manipuladoras e enganosas, que têm como objetivo distorcer materialmente o comportamento de uma pessoa, prejudicando a capacidade de tomar decisões informadas e induzindo a uma decisão que não teria tomado sob outras condições, de forma a causar danos significativos a essa pessoa, a outra pessoa ou a um grupo de pessoas; as que explorem quaisquer vulnerabilidades — devido à idade, deficiência ou situação socioeconômica —, com o objetivo de distorcer materialmente o comportamento de uma pessoa, de forma que cause danos significativos.

Além disso, também são proibidas as práticas que utilizem sistemas de categorização biométrica, os quais categorizem, individualmente, pessoas físicas, com base em seus dados biométricos, para deduzir ou inferir sua raça, opiniões políticas, filiação sindical, crenças religiosas ou filosóficas, vida sexual ou orientação sexual — essa proibição não abrange a rotulagem ou filtragem de conjuntos de dados biométricos adquiridos legalmente —; as que tenham por finalidade a avaliação ou classificação de pessoas naturais ou de grupos de pessoas naturais durante um determinado período de tempo, com base em seu comportamento social ou em características pessoais ou de personalidade conhecidas, inferidas ou previstas, com pontuação social; as que utilizem sistemas de identificação biométrica remota "em tempo real", em espaços acessíveis ao público, para fins de segurança pública, exceto quando uso seja estritamente necessário, em casos como: para uma busca direcionada de vítimas específicas de sequestro, tráfico humano e exploração sexual, bem como de pessoas desaparecidas; para a prevenção de uma ameaça específica, substancial e iminente à vida ou à segurança física, ou uma ameaça genuína, previsível, de um ataque terrorista; para localização ou identificação de suspeitos de uma infração penal; para fins de investigação criminal, ação penal ou execução de uma sanção penal, por infrações referidas em anexo, e puníveis com pena ou medida de segurança privativa de liberdade por um período máximo de quatro anos.

Ademais, é proibido o uso e sistemas de predição em âmbito penal, com base exclusivamente no perfil de uma pessoa física ou na avaliação de seus traços e características de personalidade — no entanto, essa proibição não se aplica aos sistemas

de IA usados para apoiar a avaliação humana do envolvimento em atividade criminosa, que já se baseia em fatos objetivos e verificáveis, diretamente ligados à atividade criminosa —; a utilização de sistemas que criem ou expandam bancos de dados de reconhecimento facial por meio da extração não direcionada de imagens faciais da Internet ou de filmagens de CCTV; e, por fim, a utilização de sistemas para inferir emoções de uma pessoa física nas áreas de local de trabalho e instituições de ensino, exceto nos casos em que o uso do sistema de IA se destine a ser colocado em prática por razões médicas ou de segurança.

Por sua vez, os sistemas de IA classificados como de risco elevado são aqueles utilizados como um componente de segurança de um produto, ou são propriamente um produto, listados nos anexos II e III da regulação, que incluem sistemas aplicados em áreas como biometria, infraestrutura crítica, educação e treinamento, emprego, acesso a serviços e benefícios, migração, asilo e controle de fronteiras, e administração da justiça.

É possível extrair, ainda, a classificação de sistemas de risco limitado, como *deepfakes* e sistemas que interagem diretamente com pessoas naturais (como é o caso dos *chatbots*), que estão sujeitos a obrigações específicas de transparência, incluindo informar usuários sobre interações com o sistema e realizar marcação de conteúdo sintético (áudio, vídeo, imagens e texto). Por fim, por efeito residual, existem sistemas considerados como de risco mínimo, que não são abrangidos pelas demais classificações.

Os sistemas de IA classificados como de risco elevado estão submetidos a obrigações que envolvem sistema de gerenciamento de riscos, governança de dados, documentação técnica, manutenção de registros, transparência e fornecimento de informações, supervisão humana, precisão, robustez e segurança cibernética. É importante notar que, no contexto europeu, a obrigação de conduzir uma avaliação de conformidade não é nova e faz parte de diversas normas que tratam da segurança de produtos. Inclusive, a própria regulamentação sobre inteligência artificial pode ser compreendida como uma norma que trata sobre segurança de produtos.

Uma nova avaliação de conformidade deve ser realizada quando um sistema de IA de risco elevado for substancialmente modificado, ou seja, quando uma alteração afeta a conformidade de um sistema com os requisitos para sistemas de IA de risco elevado ou resulta em uma modificação da finalidade pretendida do sistema de IA. No entanto, não há necessidade de nova avaliação quando o sistema continua a se desenvolver depois de ser colocado no mercado, desde que essas alterações sejam predeterminadas no momento

da avaliação de conformidade inicial e estejam descritas na documentação técnica do sistema, como, por exemplo, no caso de sistemas que envolvem aprendizado de máquina.

A avaliação de conformidade deve ser realizada pelo fornecedor de um sistema de IA de risco elevado, mas também pode ser realizada, em situações específicas, pelo fabricante do produto, pelo distribuidor ou pelo importador de um sistema de IA de alto risco elevado, bem como por um terceiro. Na regulação europeia, o fornecedor é uma pessoa física ou jurídica, autoridade pública, agência ou outro órgão que desenvolva ou mande desenvolver um sistema de IA ou um modelo de IA de finalidade geral e o coloque no mercado ou em serviço sob o seu próprio nome ou marca, a título oneroso ou gratuito.

Nesse contexto, é importante ressaltar que, na regulação europeia, existe, ainda, a figura do "responsável pela implantação", que é definido como a pessoa física ou jurídica, autoridade pública, agência ou outro órgão que utilize um sistema de IA sob a sua própria autoridade, salvo se o sistema de IA for utilizado no âmbito de uma atividade pessoal de caráter não profissional. Há, também, a figura do importador, que é quem coloca no mercado europeu um sistema de IA que ostenta o nome ou a marca de uma pessoa natural ou jurídica estabelecida num país terceiro, e do distribuidor, que está inserido na cadeia de abastecimento e disponibiliza um sistema de IA no mercado europeu, mas possui papel distinto do fornecedor e do distribuidor.

Há dois casos em que o fornecedor não é o agente responsável pelo desenvolvimento da avaliação de conformidade. Essa será uma obrigação do fabricante, se, cumulativamente, o sistema de IA de alto risco estiver relacionado a produtos aos quais se aplicam as leis da seção A do Anexo II da regulação; e se o sistema for colocado no mercado, ou em serviço, junto com o produto, sob o nome do fabricante.

A realização da avaliação de conformidade também poderá ser uma responsabilidade dos distribuidores, importadores ou quaisquer terceiros quando esses colocarem no mercado, ou em serviço, um sistema de IA de risco elevado com seu nome ou marca registrada; quando modificarem a finalidade pretendida, conforme determinado pelo fornecedor, de um sistema de IA de risco elevado já colocado no mercado, ou em serviço, caso em que o fornecedor inicial não será mais considerado o fornecedor para os fins da regulação; ou quando realizarem uma modificação substancial no sistema, caso em que o fornecedor inicial também não é mais considerado fornecedor para fins de aplicação da regulação.

Além disso, o *AI Act* prevê a realização de uma avaliação de impacto para direitos fundamentais (*Fundamental Rights Impact Assessment* – FRIA) por parte de organismos

regidos pelo direito público, entidades privadas que prestam serviços públicos e prestadores de serviços bancários e de seguros, que usam sistemas de IA listados como de alto risco no Anexo III, ponto 5, alíneas (b) e (c), isto é, sistemas de IA concebidos para serem utilizados para avaliar a capacidade de solvência de pessoas naturais ou estabelecer a sua classificação de crédito, salvo sistemas de IA utilizados para efeitos de deteção de fraude financeira, e sistemas de IA concebidos para serem utilizados nas avaliações de risco e na fixação de preços em relação a pessoas naturais no caso de seguros de vida e de saúde.

Nesse contexto, entende-se "organismos de direito público" quaisquer autoridades e órgãos públicos, conforme definido na Diretiva 2014/24. Por sua vez, as entidades privadas que prestam serviços públicos podem ser compreendidas como agentes privados que prestam serviços ligados a funções de interesse público, designadamente no domínio da educação, dos cuidados de saúde, dos serviços sociais, da habitação e da administração da justiça, conforme se extrai do Considerando 96 do *AI Act*. Destaca-se, ainda, que o Quadro de Qualidade para Serviços de Interesse Geral na Europa ("*Quality Framework*") enumera os prestadores de serviços de interesse geral: serviços postais, serviços bancários, serviços de transporte, serviços de energia e serviços de comunicações eletrônicas. Desse modo, entende-se que uma ampla gama de agentes privados que atuam em setores-chave estarão submetidos à obrigação de condução de uma avaliação de impacto sobre direitos fundamentais.

Nota-se, portanto, que enquanto a avaliação de conformidade é realizada pelo fornecedor do sistema de IA, a FRIA é elaborada pelo responsável pela implantação do sistema, quando aplicável, pois nem sempre os fornecedores poderão identificar e avaliar todos os potenciais cenários de implantação de um sistema e os potenciais riscos gerados por estes.

De acordo com a regulação europeia, a FRIA deverá abranger descrição dos processos em que o sistema de IA de risco elevado será utilizado; do período e da frequência de uso; das categorias de pessoas físicas e grupos que provavelmente serão afetados por seu uso no contexto específico; dos riscos e danos específicos que podem afetar as categorias identificadas; da implementação de medidas de supervisão humana; das medidas a serem tomadas em caso de materialização dos riscos identificados, incluindo seus arranjos para governança interna e mecanismos de reclamação.

A referida obrigação é aplicável somente ao primeiro uso do sistema de IA de risco elevado. Desse modo, o agente responsável por implementar determinado sistema de IA

pode, em casos semelhantes, basear-se nas avaliações de impacto realizadas pelo provedor do sistema. Contudo, se durante o uso o agente perceber a mudança ou falta de atualização dos fatores listados na avaliação, deverá tomar as medidas necessárias para atualizar as informações.

Vale destacar que a avaliação do impacto sobre os direitos fundamentais precisará ser realizada somente para os aspectos não cobertos por outras obrigações legais e deverá ser alinhada com processos de gerenciamento de riscos já existentes, a fim de eliminar quaisquer sobreposições e ônus adicional. Por exemplo, se o conteúdo desse instrumento já estiver sendo analisado no âmbito da avaliação de impactos à proteção de dados (*Data Protection Impact Assessment*), a avaliação de impacto para direitos humanos deverá ser realizada em conjunto com a avaliação de impacto para a proteção de dados.

Além disso, é importante ressaltar que, no âmbito da União Europeia, a conformidade com a obrigação de realização de avaliações de impacto para direitos fundamentos será facilitada pelo *AI Office*, encarregado de desenvolver um modelo de questionário a ser utilizado pelos agentes, para atender aos requisitos previstos.

O quadro comparativo abaixo elenca o conteúdo mínimo estabelecido pela proposta de *AI Act* para elaboração da avaliação de impacto para direitos fundamentais, bem como o conteúdo proposto pela redação do texto substitutivo ao Projeto de Lei n.º 2.338/23 (ainda pendente de votação na CTIA) para a Avaliação de Impacto Algorítmico:

Quadro 3 – Comparativo entre o conteúdo mínimo elencado pelo *AI Act* para elaboração da avaliação de impacto para direitos fundamentais e o conteúdo mínimo proposto pelo Projeto de Lei n.º 2.338/23 para a Avaliação de Impacto Algorítmico

Avaliação de impacto algorítmico Texto substitutivo - Projeto de Lei n.º 2.338/2023 - Brasil	**Fundamental rights impact assessment for high-risk AI systems** *Artificial Intelligence Act Proposal* - União Europeia
a) Riscos conhecidos e previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar; b) Benefícios associados ao sistema de inteligência artificial; c) Probabilidade de consequências adversas, incluindo o número de pessoas potencialmente impactadas;	a) Descrição dos processos do responsável pela implantação em que o sistema de IA de risco elevado será utilizado, conforme sua finalidade prevista; b) Descrição do período em que o sistema de IA de risco elevado se destina a ser utilizado e com que frequência de uso; c) Categorias de pessoas físicas e grupos que provavelmente serão afetados por seu uso no contexto específico;

d) Gravidade das consequências adversas, incluindo o esforço necessário para mitigá-las; e) Lógica de funcionamento do sistema de inteligência artificial; f) Processo e resultado de testes e avaliações e medidas de mitigação realizadas para verificação de possíveis impactos a direitos, com especial destaque para potenciais impactos discriminatórios; g) Treinamento e ações de conscientização dos riscos associados ao sistema de inteligência artificial; h) Medidas de mitigação e indicação e justificação do risco residual do sistema de inteligência artificial, acompanhado de testes de controle de qualidade frequentes; e i) Medidas de transparência ao público, especialmente aos potenciais usuários do sistema, a respeito dos riscos residuais, principalmente quando envolver alto grau de nocividade ou periculosidade à saúde ou segurança dos usuários, nos termos dos artigos 9º e 10 da Lei n.º 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor).	d) Descrição dos riscos e danos específicos que podem afetar as categorias de pessoas ou grupos de pessoas identificados; e) Descrição da implementação de medidas de supervisão humana; f) Descrição das medidas a serem tomadas em caso de materialização dos riscos identificados, incluindo seus arranjos para governança interna e mecanismos de reclamação.

Fonte: Elaborado pela autora (2024).

Desse modo, nota-se que, embora as abordagens regulatórias possam variar em termos de especificidade e detalhamento, as avaliações de conformidade e de impacto surgem como instrumentos de *accountability* no desenvolvimento e implementação de sistemas de IA, de modo a assegurar o uso responsável e a promoção de construção de confiança perante usuários e partes interessadas. Especificamente em relação à comparação entre a AIA no contexto brasileiro e a FRIA no contexto europeu, embora ambas sejam ferramentas desenvolvidas para monitorar e mitigar riscos associados a sistemas de IA, não se pode tratá-las como instrumentos equivalentes.

A AIA, ao focar em uma análise predominantemente técnica, prioriza a avaliação de aspectos operacionais e tende a considerar os direitos fundamentais como um elemento secundário, avaliando-os de forma limitada e, possivelmente, indireta. Por outro lado, a FRIA adota uma perspectiva essencialmente centrada nos direitos fundamentais, posicionando-os como o ponto focal da análise. Inclusive, entende-se que esta abordagem se mostra mais adequada quando o objetivo é compreender os efeitos que a tecnologia

pode ter sobre pessoas e grupos afetados, considerando, para além da técnica, questões contextuais, como efeitos sociais, culturais e econômicos.

No contexto europeu, Mantelero (2024) esclarece que a FRIA foi um resultado alcançado pelo Parlamento Europeu em relação a uma proposta da Comissão Europeia que enfatizava uma abordagem centrada no ser humano e na proteção dos direitos fundamentais. Nesse sentido, o autor aponta que a FRIA não é uma invenção totalmente nova do legislador europeu no âmbito do *AI Act* e se baseia na experiência da *Human Rights Impact Assessment* (HRIA)[2], estabelecida em nível internacional e implementada por organizações e agentes privados em vários contextos. No entanto, Mantelero (2024) observa que a HRIA no âmbito da IA não deve ser compreendida como uma HRIA tradicional.

Por exemplo, uma diferença elencada por Mantelero (2024) diz respeito ao escopo operacional da HRIA e da FRIA, pois a HRIA tradicional é compreendida, principalmente, como uma ferramenta política que fornece às organizações uma avaliação dos possíveis impactos e uma lista de possíveis soluções para evitá-los ou mitigá-los, deixando a cargo do agente decidir quais soluções adotar e até que ponto reduzir esses impactos. Por outro lado, de acordo com a *AI Act*, a FRIA é uma avaliação obrigatória, cujos resultados necessariamente devem ser usados para evitar a materialização dos riscos identificados.

Além disso, Mantelero (2024) aponta que a FRIA realizada no âmbito dos sistemas de IA difere da HRIA tradicional em razão da própria natureza das situações avaliadas. Enquanto a HRIA é normalmente aplicada no contexto de atividades industriais localizadas em um território específico e que impactam uma ampla gama de direitos humanos, para Mantelero (2024), os sistemas de IA são frequentemente soluções distribuídas globalmente que normalmente impactam em uma gama limitada de direitos fundamentais. No entanto, o próprio autor reconhece que existem exceções a essa distinção, como é o caso, por exemplo, dos projetos de cidades inteligentes, em que os sistemas de IA são implantados em um contexto territorial específico, e dos *large language models*, que podem ser usados para uma ampla gama de finalidades diferentes, com possível impacto em todos os direitos fundamentais.

---

[2] Vale ressaltar que, em 2024, o Parlamento Europeu aprovou Diretiva (Directive on Corporate Sustainability Due Diligence – Directive 2024/1760) que exige que empresas e seus parceiros e fornecedores previnam ou mitiguem seus impactos adversos sobre os direitos humanos e o meio ambiente, incluindo aspectos como escravidão, trabalho infantil, exploração trabalhista, perda de biodiversidade, poluição ou destruição do patrimônio natural.

Para Mantelero (2024), a criação de modelos para identificar e avaliar os possíveis riscos aos direitos fundamentais continua sendo o grande desafio quando se fala em Avaliação de Impacto Algorítmico. Por isso, o autor entende que a FRIA não pode ser desenvolvida simplesmente levando em conta as práticas já consolidadas da HRIA ou apenas copiando os modelos da HRIA.

No contexto regulatório europeu, nota-se, portanto, que existem dois instrumentos baseados na identificação de risco e que estão diretamente relacionados: a avaliação de conformidade e a FRIA. A avaliação sob a perspectiva de direitos fundamentais está presente em ambos os modelos, como parte de uma análise mais ampla e geral na avaliação de conformidade e como um objetivo específico no caso da FRIA. A FRIA é uma obrigação direcionada apenas responsáveis pela implantação de sistemas de IA, mas a avaliação de conformidade é direcionada aos fornecedores de sistemas de IA, de modo que ambos os agentes devem avaliar impactos para direitos fundamentais.

É importante notar que a sobreposição de diferentes escopos no desenvolvimento de instrumentos de avaliação, combinada com o fato de que o responsável pela implantação de um sistema de IA não tem acesso ao mesmo nível de informações que o provedor do sistema, e vice-versa, representa um desafio significativo para a metodologia de desenvolvimento desses instrumentos de governança no contexto europeu.

Para Selbst (2021), uma regulamentação eficaz da AIA dependerá consideravelmente da cooperação e da boa-fé dos agentes regulados, especialmente das empresas de tecnologia. Por tal razão, é preferível que esse instrumento esteja em sintonia com a forma de operação do setor privado, em vez de entrar em conflito com ele. O autor argumenta que é crucial a compreensão, por parte dos reguladores, do próprio setor técnico, incluindo a tecnologia, a cultura organizacional e os padrões de autorregulação emergentes.

Os objetivos de desenvolvimento da AIA serão mais eficazmente alcançados por meio de uma parceria entre o órgão regulador e o setor privado, estabelecendo uma governança colaborativa e evitando abordagens adversárias. Selbst (2021) aponta que uma regulamentação totalmente prescritiva não é realista, uma vez que ignora a dinamicidade e o desenvolvimento da inteligência artificial, além de, por vezes, gerar lacunas de conhecimento e experiência. Além disso, ao adotar uma abordagem regulatória totalmente prescritiva, a regulação corre o risco de tornar-se obsoleta rapidamente, à medida que a tecnologia avança.

Entende-se, no entanto, que a abordagem defendida por Selbst apresenta limitações. Em primeiro lugar, é importante considerar que a proximidade excessiva entre reguladores e empresas de tecnologia pode resultar no fenômeno de captura regulatória, no qual as organizações reguladas influenciam o processo regulatório a seu favor. Referido fenômeno pode comprometer a eficácia da regulamentação, priorizando interesses comerciais do mercado em detrimento da proteção e promoção a direitos humanos.

Além disso, é necessário levar em consideração o fato de que, em geral, as empresas possuem maior nível de conhecimento técnico acerca das tecnologias que desenvolvem do que os reguladores. Esse cenário cria uma assimetria de poder e conhecimento que pode ser explorada em uma abordagem de "parceria" e resultar em regulamentações menos eficazes ou superficiais, que não enfrentam de forma adequada os riscos gerados por novas tecnologias.

Evidentemente, o objetivo de uma regulamentação eficaz da Avaliação de Impacto Algorítmico (AIA) não deve ser promover uma abordagem adversarial entre reguladores e regulados, mas promover o equilíbrio adequado entre cooperação e exigência de conformidade. Esse equilíbrio envolve a criação de um ambiente regulatório onde ao mercado tenha incentivos claros para aderir às melhores práticas e aos padrões éticos, ao mesmo tempo em que existem mecanismos de fiscalização e supervisão.

Desse modo, ainda que a AIA possa ser um instrumento regulado, é importante que esteja baseado em uma abordagem flexível e adaptável, possibilitando a evolução, juntamente com o campo, e incorporação de novos conhecimentos e experiências à medida que surgem. A flexibilidade não deve ser sinônimo de ausência de responsabilidade, mas uma chave para que a AIA seja um instrumento vivo, baseado em uma abordagem adaptativa, a qual permita ajustes contínuos para acompanhar o rápido desenvolvimento tecnológico.

Diante da análise bibliográfica realizada no capítulo anterior e do levantamento iniciativas regulatórias no Brasil e na União Europeia neste capítulo, retorna-se às primeiras três questões que o presente trabalho busca investigar: (i) quais parâmetros podem ser utilizados para determinar quando a AIA deve ou não ser realizada? (ii) qual é a metodologia para desenvolvimento da AIA? (iii) qual conteúdo a AIA deveria abranger?

Para endereçamento destes questionamentos apresentados como norteadores da pesquisa, apresenta-se a seguir as conclusões obtidas por meio das técnicas de revisão bibliográfica e análise documental de propostas regulatórias:

1) Os parâmetros utilizados para determinação de quando a AIA deve ou não ser realizada são aqueles relacionados à classificação de risco do sistema de IA em questão. Desse modo, caso um sistema de IA seja considerado como de alto risco por eventual legislação aplicável ou por metodologia adotada pelo próprio agente responsável pelo desenvolvimento ou pela implantação do sistema de IA (por exemplo, uma matriz de risco própria ou setorial), a AIA deverá ser conduzida. Sem prejuízo, caso o agente identifique fatores de risco ou potenciais impactos negativos relevantes, a AIA poderá ser conduzida a título de boa prática.
2) Entende-se que, ainda que a AIA possa ser um instrumento regulado, a forma de desenvolvimento e procedimentalização deve ser baseada uma abordagem flexível e adaptável por parte dos agentes responsáveis por sua condução. Conforme o presente trabalho procurou demonstrar, uma abordagem regulatória totalmente prescritiva pode gerar lacunas de conhecimento e experiência, além de não serem plenamente aplicáveis a depender do setor.
3) A AIA deverá abordar o ator (quem), os fóruns (quando e onde) e o conteúdo (o que). Em relação ao conteúdo, para além da descrição do funcionamento do sistema de IA, de seus aspectos técnicos pertinentes e das particularidades dos contextos e atividades nos quais tal sistema será aplicado, é importante que a AIA avalie como o sistema de IA pode afetar diferentes grupos sociais e indivíduos, o que inclui a identificação de riscos e danos específicos, bem como avaliação de impactos sobre os direitos fundamentais das pessoas afetadas.

5 ANÁLISE JURÍDICA DA AVALIAÇÃO DE IMPACTO ALGORÍTMICO A PARTIR DOS PRINCÍPIOS DE *ACCOUNTABILITY*, PRECAUÇÃO E TRANSPARÊNCIA

Conforme demonstrado, em um cenário de constante evolução dos sistemas de IA, a Avaliação de Impacto Algorítmico assume papel central na garantia da coexistência harmoniosa entre promoção da inovação e do desenvolvimento tecnológico e proteção de direitos e garantias individuais e coletivos. Nesse contexto, a AIA pode ser compreendida a partir de uma base tríplice fundamentada nos princípios de *accountability*, precaução e transparência.

A *accountability*, como primeiro pilar, estabelece a responsabilidade das partes envolvidas no desenvolvimento, implementação e uso de algoritmos, implicando que tais entes devem prestar contas acerca de suas ações. Por sua vez, a precaução reconhece a natureza intrinsecamente dinâmica dos sistemas de IA, o que impõe a necessidade de identificar e gerenciar possíveis riscos antes que se materializem. Por fim, como terceiro pilar, a transparência fomenta o diálogo informado entre agentes de inteligência artificial, pessoas afetadas pelos sistemas de IA e a sociedade como um todo.

5.1 *ACCOUNTABILITY*

A *accountability* é um conceito que se popularizou no âmbito da governança pública e, posteriormente, passou a ser aplicado na governança corporativa, tornando-se um indicativo de boas práticas. O termo não possui uma tradução direta para o português, sendo frequentemente traduzido como “responsabilidade” ou “prestação de contas”. Além disso, nota-se que, com o passar do tempo, o termo passou a ser utilizado de forma retórica, como uma expressão genérica que pode evocar diferentes imagens vagas sobre práticas de boa governança, como confiabilidade, legitimidade ou justiça.

Como aponta Bovens (2007), no discurso político e acadêmico contemporâneo, o termo “*accountability*” funciona como um guarda-chuva conceitual, que abrange vários outros conceitos distintos, como transparência, equidade, democracia, eficiência, capacidade de resposta, responsabilidade e integridade. Tais definições, excessivamente amplas e vazias, fazem com que o uso do conceito seja questionado, uma vez que a compreensão prática do que é *accountability* é alterada conforme o contexto de aplicação.

O presente trabalho compreende a *accountability* a partir das lentes de Bovens (2007), que esclarece que o conceito não se trata uma palavra de ordem meramente

política, mas de práticas concretas. Para o autor, a descrição mais concisa de *accountability* seria "a obrigação de explicar e justificar a conduta", o que implica em um relacionamento entre um ator e um fórum, no qual o ator tem a obrigação de explicar e justificar sua conduta perante o fórum, havendo possibilidade de que o fórum faça perguntas e avalie a conduta do ator, que, por sua vez, pode vir a sofrer consequências.

Nesse sentido, Bovens (2007) elenca três elementos essenciais do conceito de *accountability*. Em primeiro lugar, é fundamental que o ator seja obrigado a informar o fórum sobre sua conduta, fornecendo várias informações e dados sobre determinada atividade. Em segundo lugar, é necessário que haja a possibilidade de o fórum interrogar o ator e questionar a adequação das informações ou a legitimidade da conduta, inclusive, desse ponto decorre a estreita conexão semântica entre "*accountability*" e "responsabilidade" ou "prestação de contas". Por fim, o fórum deve ter possibilidade de julgar a conduta do ator, por exemplo, aprovando sua conduta ou condenando publicamente determinado comportamento. Ao emitir um julgamento negativo, o fórum frequentemente impõe algum tipo de consequência ao ator.

Transportando essa compreensão para o campo de análise do presente trabalho, verifica-se que, para que a Avaliação de Impacto Algorítmico possa ser categorizada como uma prática de *accountability*, é importante que o instrumento, ao ser desenvolvido pelo ator (agente responsável pelas de etapas de desenvolvimento ou implementação de sistemas de IA), seja submetido a um fórum, que possua a capacidade de efetivamente avaliar o instrumento e questionar as justificativas e conclusões expostas, bem como sugerir modificações.

Nesse ponto, um questionamento central é entender por qual razão um ator se submeteria à prestação de contas a um fórum. Uma primeira possibilidade seria porque é obrigado a fazê-lo, por exemplo, em razão de normas aplicáveis ao caso concreto. Por outro lado, o autor pode fazê-lo voluntariamente, como no caso de regras ou procedimentos de governança internos que exijam tal processo. Bovens (2007) denomina o primeiro cenário de *accountability* vertical, na qual o fórum exerce formalmente poder sobre o ator, e o último cenário, de *accountability* horizontal, na qual, em geral, não há uma relação hierárquica entre o ator e o fórum, assim como não há obrigações formais de prestação de contas. Como diferentes tipos de agentes estão envolvidos em diferentes etapas do ciclo de vida de sistemas de IA, eles podem ser responsabilizados por tipos de fóruns distintos, por exemplo, interno ou externo à organização, com estrutura formal ou informal.

Conforme examinado anteriormente, as avaliações de impacto não são instrumentos neutros e sua elaboração, por si só, não gera responsabilidade e prestação de contas, exigindo que os métodos usados para determinar os impactos sejam submetidos a um fórum que tenha a capacidade de exigir mudanças no desenvolvimento ou na implementação dos sistemas de IA ou na forma de mitigar os riscos identificados. Assim, é importante notar que impactos somente são reconhecidos em uma relação de *accountability* que obriga os agentes a identificarem, explicarem e justificarem, ou melhorarem, as características de determinado sistema de IA sob avaliação, o que pode ser realizado por meio de estruturas verticais ou horizontais.

Por vezes, nota-se uma percepção amplamente defendida de que os fóruns devem ser obrigatórios e públicos, nos moldes da *accountability* vertical. Um exemplo de aplicação do conceito de *accountability* vertical é o Projeto de Lei n.º 2.338/2023, o qual, conforme mencionado no capítulo anterior, prevê que a autoridade competente deve ser notificada sobre um sistema de IA classificado como de alto risco, mediante o compartilhamento da avaliação preliminar de classificação do sistema e da avaliação impacto algorítmico, sendo possível identificar, portanto, que há uma obrigação de que o ator preste contas a um fórum.

Entende-se, contudo, que a natureza dos fóruns pode variar, a depender do ciclo de vida do sistema de IA, dos agentes envolvidos e do próprio contexto de desenvolvimento ou implementação do sistema. Nesse sentido, é possível conduzir a Avaliação de Impacto Algorítmico em fóruns internos, acessíveis apenas aos membros da organização responsável pelo desenvolvimento ou aplicação de um sistema de IA, de modo a fornecer um ambiente controlado para discussão. Tais fóruns podem, inclusive, agregar membros externos, como especialistas, consultorias e público afetado, mas são essencialmente conduzidos voluntariamente, em uma perspectiva de *accountability* horizontal.

Um exemplo de *accountability* horizontal são os "*red teams*", instituídos internamente para testar e desafiar os sistemas de IA desenvolvidos ou sob implantação em uma organização. Trata-se de um conceito inspirado nas práticas militares, onde um "*red team*" representa um adversário em exercícios simulados. Em geral, tais equipes são compostas por profissionais com diferentes especializações técnicas e operam de maneira independente das equipes de desenvolvimento e operações de sistemas de IA. O principal objetivo é a identificação de falhas, vulnerabilidades e inconsistências de um sistema de

IA, bem como avaliação de riscos e sugestões de mitigação e modificação do sistema. O "*red team*" é, portanto, um exemplo de fórum interno.

Sob a perspectiva da *accountability*, a AIA pode ser vista como um instrumento de governança que atua tanto para evitar proativamente impactos negativos decorrentes do desenvolvimento ou implementação de um sistema de IA, quanto para sancionar reativamente o ator responsável caso um risco gerado pelo sistema se materialize. Sob uma perspectiva organizacional, os agentes envolvidos no desenvolvimento e implementação de sistemas de IA podem adotar diversos instrumentos de governança, integrando a *accountability* a uma estratégia de responsabilidade digital corporativa que enfatiza a assunção de responsabilidade pelos sistemas de IA desenvolvidos ou utilizados pela organização (Horneber; Laumer, 2023).

Estabelecida a relação entre a AIA e o conceito de *accountability*, é importante delinear o que não deve ser considerado um requisito essencial para entender a AIA como um instrumento de *accountability*. Primeiramente, a participação pública em um processo de avaliação não deve ser confundida com *accountability*. Embora a participação pública seja um elemento crucial da governança democrática, diferentes tipos de avaliação de impacto mobilizam distintas formas de representação e participação, como consultas públicas, grupos de discussão e espaços de deliberação (Watkins *et al*., 2021).

Em segundo lugar, a transparência também é frequentemente utilizada como um sinônimo de *accountability*. Todavia, a transparência em si não é suficiente para constituir uma forma de *accountabiliy* — nos termos definidos anteriormente — porque não envolve necessariamente a avaliação de um fórum (Bovens, 2007). No limite, a transparência pode ser considerada como uma ferramenta para a concretização de *accountability*, uma vez que possibilita que dados e informações estejam ao alcance dos fóruns.

Desse modo, a Avaliação de Impacto Algorítmico (AIA) desempenha um papel fundamental como instrumento de governança dentro da estrutura de *accountability*, tanto ao prevenir impactos negativos potenciais quanto ao sancionar reativamente os responsáveis quando esses riscos se materializam. No contexto organizacional, a AIA deve ser integrada a uma estratégia mais ampla de responsabilidade digital corporativa, em que a *accountability* é um princípio central. No entanto, é essencial reconhecer que nem a participação pública nem a transparência, embora importantes, são suficientes para definir a AIA como um instrumento de *accountability*. A participação pública contribui para a governança democrática, e a transparência facilita o acesso a informações, mas

ambos devem ser entendidos como componentes que apoiam, mas não substituem, a *accountability* propriamente dita.

5.2 PRECAUÇÃO

O princípio da precaução pode ser amplamente observado no ordenamento jurídico brasileiro, tendo surgido na década de 1970, a partir de iniciativas de proteção ambiental que buscavam evitar danos ambientais marcados pela incerteza e indeterminação do tipo de dano. Nesse sentido, Bioni e Luciano (2019) apontam que o princípio da precaução surge em decorrência da insuficiência dos métodos tradicionais de regulação de risco diante de incertezas.

Nessa direção, Costa (2012) aponta que, pelo princípio da precaução, em situações nas quais existam ameaças de danos graves ou irreversíveis, mesmo que não haja plena certeza científica, é necessário tomar medidas de proteção sem esperar que esses riscos se tornem plenamente aparentes. O autor destaca que a avaliação de risco e o princípio da precaução andam juntos, pois são instrumentos que determinam conjuntamente a atribuição da avaliação dos riscos e do custo dos danos.

Costa (2012) esclarece, ainda, que o princípio da precaução coloca em evidência os valores normativos de prudência e transparência, criando um dever de cuidado. Conjuntamente, tais valores implicam que as atividades devem ser realizadas de forma a evitar que efeitos prejudiciais atinjam outras pessoas, possibilitando o desenvolvimento com segurança. Assim, o princípio da precaução funciona como uma salvaguarda, promovendo uma abordagem proativa na gestão de riscos e reforçando a responsabilidade de proteger o bem-estar coletivo diante de incertezas geradas por tecnologias emergentes.

Vale ressaltar que um dos campos de aplicação do princípio da precaução é a proteção de dados pessoais, uma vez que tal princípio contribui para a consolidação de uma abordagem baseada no risco (*risk based approach*). A partir dessa abordagem, os agentes de tratamento devem implementar rotinas de avaliação de riscos em atividades de tratamento de dados pessoais e endereçar as medidas para mitigação dos riscos identificados.

Desse modo, o princípio da precaução apresenta-se como uma garantia contra riscos potenciais que, no atual momento do tratamento de dados pessoais, podem não ser identificados. Trata-se de uma abordagem que visa evitar ou minimizar os riscos associados a uma determinada atividade, mesmo na ausência de evidências científicas

conclusivas sobre esses riscos. Desse modo, verifica-se que este princípio é frequentemente aplicado situações em que existe incerteza significativa sobre os impactos de uma tecnologia ou atividade. Assim como nos campos de conhecimento ambiental e de proteção de dados, entende-se que pode ser aplicado no contexto de desenvolvimento e aplicação de sistemas de IA.

Por vezes, sistemas de IA podem se mostrar complexos e marcados por imprevisibilidade, o que gera incerteza sobre as consequências a longo prazo e os possíveis efeitos colaterais não intencionais. Considerando que sistemas de IA podem ser aplicados em processos de tomada de decisões em diferentes contextos (como carros autônomos, diagnósticos e cirurgias no setor de saúde, avaliações de crédito e concessões de empréstimos no setor financeiro etc.), a incerteza sobre os impactos de tais processos decisórios pode ensejar aplicação do princípio da precaução, de modo a evitar a materialização de potenciais danos e impactos negativos.

Por outro lado, existem doutrinadores que entendem que, ao aplicar o princípio da precaução, haverá limitação e desestímulo à inovação, prejudicando o crescimento econômico, a vantagem competitiva e o progresso social (Castro; McLaughlin, 2019). Para Castro e McLaughlin (2019), em vez de impor preventivamente regulamentações prescritivas para evitar danos hipotéticos, reguladores e formuladores de políticas deveriam aguardar e criar soluções direcionadas para problemas específicos, caso ocorram.

Castro e McLaughlin (2019) reconhecem que, em determinadas situações, como no caso de uso de energia nuclear, a aplicação do princípio da precaução faz sentido porque o risco de errar pode ser catastrófico. Contudo, os autores entendem que, para a maioria das áreas de inovação, o princípio da precaução causa mais danos do que benefícios, porque gera cenários hipotéticos que sugerem incorretamente que o avanço tecnológico apresenta ameaças graves e irreversíveis. Nessa direção, Castro e McLaughlin (2019) apontam que regulações já existentes, bem como intervenções regulatórias superficiais e direcionadas, poderiam gerenciar os riscos gerados por novas tecnologias, evitando-se proibições gerais.

O *trade-off* entre precaução e inovação é frequentemente discutido no contexto de tomada de decisões, especialmente quando se trata de introduzir novas tecnologias que podem ter impactos desconhecidos. No entanto, entende-se que esta relação não se trata de uma dicotomia inevitável, mas de um desafio de harmonização entre valores igualmente relevantes para a sociedade.

Conforme aponta Mantelero (2022), existem duas possíveis abordagens jurídicas para os desafios postos pelo desenvolvimento e aplicação de sistemas de IA: a abordagem de precaução e a de riscos. O autor aponta que tais abordagens são alternativas, mas não incompatíveis, e que tecnologias complexas com uma pluralidade de impactos diferentes podem ser melhor abordadas por meio de uma combinação entre esses dois remédios.

A precaução está intimamente ligada à noção de incerteza. No campo das novas tecnologias, quando uma aplicação tem o potencial de gerar riscos graves para indivíduos e para a sociedade, e esses riscos não podem ser antecipadamente calculados ou quantificados com precisão, é necessário adotar uma abordagem precaucionária. Nesse contexto, a incerteza associada a essas tecnologias torna inviável uma avaliação de risco concreta, que exige conhecimento específico sobre a extensão das possíveis consequências negativas, mesmo que dentro de classes específicas de riscos (Mantelero, 2022).

Em síntese, quando as possíveis consequências de um sistema de IA não podem ser amplamente previstas, será difícil realizar uma avaliação de impacto adequada, o que poderá justificar a adoção de medidas de precaução. Nessa direção, Mantelero (2022) aponta que não se trata de limitar a inovação, mas de investigar, de modo aprofundado, suas possíveis consequências, bem como orientar o processo de inovação e pesquisa, incluindo as medidas de mitigação, como estratégias de contenção, padrões e rotulagem e esquemas de compensação.

Para Mantelero (2022), quando uma abordagem pautada no princípio da precaução sugere que uma tecnologia não deve ser usada em um determinado contexto, isso não implica necessariamente na interrupção de seu desenvolvimento. Em vez disso, o desenvolvimento deve continuar até que a tecnologia atinja um nível de maturidade suficiente para demonstrar uma compreensão clara dos riscos envolvidos e apresentar soluções eficazes para mitigá-los.

A avaliação de impacto pode desempenhar um papel importante no desenvolvimento de sistemas de IA, possibilitando que seja alcançado nível adequado de compreensão acerca das possíveis consequências do sistema, reduzindo a incerteza (Mantelero, 2022). Desse modo, uma possível alternativa é a adoção de uma abordagem que busque introduzir a precaução no *design* de sistemas de IA, desde o momento de ideação e concepção de tais tecnologias. Essa estratégia pode ser mais eficiente do que tentar mitigar riscos retroativamente, após o sistema já estar em funcionamento.

Adotar essa abordagem preventiva no design dos sistemas incentiva práticas de inovação responsável, que levam em consideração os impactos éticos e sociais desde o início. Além disso, como muitas inovações tecnológicas são desenvolvidas de forma iterativa, com ajustes e melhorias contínuas, a precaução pode ser integrada ao processo de inovação à medida que novas informações e dados sobre riscos e impactos se tornam disponíveis. Isso assegura que a tecnologia evolua de maneira segura e responsável, com uma atenção constante às possíveis consequências para os direitos humanos.

Sendo assim, a precaução e o gerenciamento de riscos podem ser considerados ferramentas complementares para o desenvolvimento de tecnologias centradas nos direitos humanos (Mantelero, 2022). Nesse contexto, a Avaliação de Impacto Algorítmico desempenha um papel essencial na harmonização entre precaução e inovação, permitindo a identificação proativa de riscos, inclusive de impactos a longo prazo, e a implementação de medidas mitigatórias. Além disso, essa avaliação leva em conta aspectos sociais e éticos, contribuindo para um desenvolvimento tecnológico mais responsável e alinhado com a proteção dos direitos humanos.

5.3 TRANSPARÊNCIA

À medida que sistemas de IA desempenham atividades significativas em atividades cotidianas, a transparência vem assumindo papel de destaque nos debates sobre a governança de sistemas de IA, uma vez que a capacidade de explicar o raciocínio e os resultados obtidos por tais sistemas poderia impulsionar a compreensão e confiança em seu uso.

Nessa direção, a doutrina aponta que, por vezes, algoritmos e sistemas complexos operam de maneira opaca e sem transparência, sem que os indivíduos afetados saibam como as decisões são tomadas. A ideia de enxergar os algoritmos como “caixas-pretas”, desenvolvida principalmente por Frank Pasquale (2015), sugere que a opacidade pode levar a práticas discriminatórias e perpetuação de vieses. Assim, é importante compreender o funcionamento desses sistemas e como seus resultados são gerados, a fim de proteger direitos e prevenir abusos.

De modo geral, a busca pela transparência se apoia na ideia de que a compreensão acerca do funcionamento do sistema de IA é essencial para garantir que os resultados produzidos por esse sistema sejam justificáveis e alinhados com padrões éticos, técnicos e regulatórios. Diante desse contexto, a transparência possibilitaria maior clareza para

determinar responsabilidades e implementar medidas corretivas diante de eventual funcionamento inadequado.

Conforme apontado anteriormente, a transparência é, frequentemente, colocada como um sinônimo da *accountability*, ou compreendida como uma "forma" de *accountability*. Contudo, Bovens (2007) esclarece que a transparência, por si só, é um instrumento de *accountability*, mas não se confunde com o conceito em si. Em síntese, o autor esclarece que dimensões, como a transparência, são instrumentais para a *accountability*, mas não são constitutivas da *accountability.*

Ananny e Crawford (2016) apontam que as preocupações com a transparência são comumente impulsionadas por uma cadeia de lógica de entendimento, na qual a observação produz percepções, as quais criam o conhecimento necessário para governar e responsabilizar os sistemas. A partir dessa concepção, quanto mais fatos forem revelados, mais a verdade poderá ser conhecida por meio de uma lógica de acumulação de conhecimento. A observação é compreendida como um diagnóstico para a ação, pois os "observadores" com maior acesso aos "fatos" terão melhores condições para julgar se o sistema está funcionando como se pretende e quais mudanças são necessárias.

Assim, a lógica subjacente é que quanto mais se conhece sobre o funcionamento interno de um sistema, maior é a capacidade de governá-lo e responsabilizá-lo. No entanto, Ananny e Crawford (2016) argumentam que a transparência não é um estado final e absoluto em que tudo é claro e visível. Na realidade, a transparência possui dimensões sociais complexas, incluindo a preocupação com segredos e a percepção de que a capacidade de observar algo pode levar a um maior controle e, consequentemente, a uma sensação aumentada de segurança.

Com o avanço acadêmico na relação entre transparência e *accountability*, surgiram diversas tipologias para abordar o tema. Por exemplo, Fox (2007) distingue as práticas de transparência como "difusas" quando as informações divulgadas não refletem verdadeiramente o comportamento das organizações, e como "claras" quando são fornecidas informações confiáveis que detalham desempenhos e responsabilidades.

Além disso, há uma distinção entre a transparência que promove a *soft accountability*, na qual as organizações prestam contas ao público por suas ações, e a *hard accountability*, onde a transparência serve como um mecanismo para sancionar e exigir compensação por danos (Fox, 2007). Acerca desta questão, Selbst (2021) também observa que uma maior carga de obrigações de transparência pode gerar resistência por parte dos

agentes do setor privado, incentivando-os a fornecer documentação mais vaga para proteger informações sensíveis dos concorrentes e do escrutínio público.

É importante ressaltar que diversas outras tipologias podem ser encontradas na doutrina, por exemplo, transparência retrospectiva e em tempo real, transparência interna e externa, transparência para cima e para baixo etc. (Fox, 2007). Contudo, o presente trabalho não busca analisar especificamente as diferentes classificações de práticas de transparência, mas estudar sua aplicabilidade à Avaliação de Impacto Algorítmico e suas possíveis limitações.

Nesse sentido, Ananny e Crawford (2016) entendem que se a transparência não tiver efeitos significativos, então a ideia pode perder seu propósito. Para verificar se a transparência é significativa em determinado contexto, é importante compreender suas limitações. Em primeiro lugar, verifica-se que a transparência pode ser prejudicial se implementada sem prévia análise e determinação do motivo pelo qual alguma parte de um sistema deve ser revelada, o que pode causar exposição de vulnerabilidade e, até mesmo, danos à privacidade (Ananny; Crawford, 2016). Além disso, a transparência pode gerar oclusão quando produz quantidades tão grandes de informações que partes importantes acabam se tornando ocultas em meio a tanto conteúdo divulgado (Ananny; Crawford, 2016).

Também é importante notar que a ênfase na transparência como uma forma de *accountability* pode ser excessivamente onerosa para indivíduos, que terão que buscar as informações, interpretá-las e compreendê-las, invocando um modelo neoliberal de agência que, por vezes, ignora a existência de assimetrias, as quais fazem com que as informações não sejam igualmente visíveis e compreensíveis para todos (Ananny; Crawford, 2016).

Por fim, Ananny e Crawford (2016) apontam que a transparência pode privilegiar a visão em detrimento da compreensão, ou seja, ter acesso ao interior de um sistema não significa necessariamente compreender seu comportamento ou suas origens. Na verdade, compreender um sistema complexo exige interagir dinamicamente com eles para entender como se comportam em relação a seus ambientes, o que inclui, por exemplo, análises sobre o design do sistema e seu contexto de aplicação.

Nesse contexto, é possível adotar uma lógica de transparência “por camadas”, na qual destinatários distintos recebem diferentes informações. Trata-se do reconhecimento de que as possíveis partes interessadas têm diferentes níveis de conhecimento, interesse e capacidade de compreensão sobre o funcionamento e os resultados dos sistemas de IA.

Desse modo, pessoas afetadas por sistemas de IA podem ter contato com informações mais objetivas e relevantes, enquanto autoridades regulatórias, em razão das funções de supervisão e fiscalização, podem receber informações mais detalhadas e abrangentes.

De todo modo, a divulgação de informações decorrentes da AIA deve ser acompanhada de garantias que auxiliem na mitigação dos riscos de que a publicização de vulnerabilidades e pontos fracos de determinado sistema de IA seja explorada por agentes mal-intencionados. Nesse sentido, uma das garantias oferecidas pelo ordenamento jurídico brasileiro é a proteção do segredo comercial e industrial, de modo que as informações sensíveis e estratégicas para o negócio, fornecidas às autoridades por meio da AIA, não devem ser publicizadas.

Portanto, a transparência significativa deve ir além da simples divulgação de informações, de modo que é mais importante que as informações sobre o funcionamento de um sistema sejam debatidas — ainda que fora da esfera pública — do que simplesmente estejam visíveis e publicizadas. Sendo assim, a transparência significativa não se trata apenas de divulgar informações e documentações técnicas sobre os sistemas, como é o caso das Avaliações de Impacto Algorítmico, mas também de discutir as implicações éticas e sociais do funcionamento de determinado sistema de IA, podendo incluir, ou não, a participação de outras partes interessadas, a depender do fórum que irá avaliar a AIA.

5.4 DESENHO DA AVALIAÇÃO DE IMPACTO ALGORÍTMICO A PARTIR DO TRIPÉ DE *ACCOUNTABILITY*, PRECAUÇÃO E TRANSPARÊNCIA

A partir dos entendimentos apresentados nos subcapítulos anteriores acerca dos princípios de *accountability*, precaução e transparência, é possível desenhar a Avaliação de Impacto Algorítmico como um instrumento amparado em um tripé formado por tais princípios. O presente trabalho opta por utilizar a representação de um tripé porque se trata de uma estrutura estável e equilibrada, devido à distribuição uniforme do peso em três elementos inter-relacionados. Nesse sentido, o termo “tripé” é usado metaforicamente para representar três pilares ou elementos fundamentais de um conceito, no caso deste trabalho, trata-se da Avaliação de Impacto Algorítmico.

Em relação a *accountability*, verifica-se que o elemento é importante para que a AIA seja submetida a um fórum de avaliação, possibilitando questionamentos sobre decisões tomadas e justificativas apresentadas, alterações e recomendações, bem como

imposição de consequências em caso de inadequações. A participação de um fórum possibilita que os agentes que desenvolvem e implementam sistemas de IA sejam — de fato — submetidos a um processo de prestação de contas. Em síntese, a compreensão da AIA sob a ótica de *accountability* possibilita que o instrumento seja compreendido em uma perspectiva relacional entre ator e fórum, o que enriquece o processo de avaliação.

Por sua vez, a precaução é um princípio relevante para desenvolvimento da AIA porque possibilita uma abordagem proativa na identificação e gestão de potenciais riscos associados ao desenvolvimento ou implementação de sistemas de IA, desde o momento inicial de concepção do sistema ou ideação de seu uso. A partia da AIA torna-se possível antecipar e gerenciar eventuais impactos negativos de maneira preventiva, o que auxilia na concretização do princípio da precaução.

A AIA possibilita que o gerenciamento de riscos seja feito de acordo com casos concretos de desenvolvimento ou aplicação de sistemas de IA, evitando vedações abstratas e genéricas que poderiam prejudicar a inovação e o desenvolvimento tecnológico. Trata-se, portanto, de uma forma de reconhecer a incerteza e a falta de dados completos sobre os possíveis impactos gerados por sistemas de IA e, por meio da AIA, trabalhar para identificar e mitigar riscos preventivamente.

Por fim, verifica-se que a transparência é um elemento importante para a AIA, uma vez que possibilita o acesso a informações e documentações sobre o funcionamento do sistema de IA em questão. Contudo, é importante checar se a transparência é verdadeiramente significativa e útil perante o fórum responsável pela avaliação da AIA, funcionando como um instrumento para a *accountability*. A transparência não deve se resumir a mera publicização e divulgação de informações e documentações sobre os sistemas de IA, contrariamente, deve possibilitar que o fórum responsável pela análise da AIA tenha capacidade de questionar e avaliar, de forma crítica, as decisões adotadas e justificativas apresentadas.

A AIA, enquanto instrumento de governança, é uma ferramenta na qual os agentes responsáveis pelo uso e desenvolvimento de sistemas de IA devem explorar detalhes do sistema, enfrentando seus pontos fortes e pontos fracos e identificando os riscos e impactos.

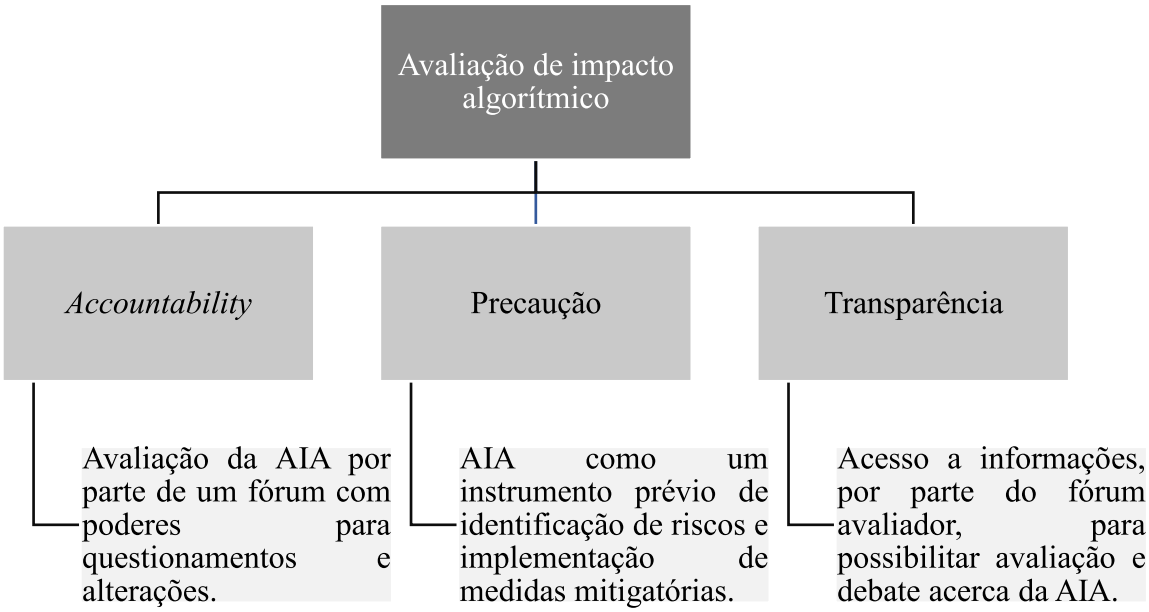
Caso haja exigência no sentido de que a AIA seja integralmente publicizada, é possível enxergar dois potenciais impactos. Em primeiro lugar, os agentes, ao elaborarem a AIA, podem não explorar com profundidade todas as circunstâncias de riscos que deveriam constar da AIA com receio de impactos no segredo de negócio de suas

atividades e em sua imagem e reputação perante o mercado; e, além disso, há o risco de que, ao publicizar os detalhes de funcionamento e os elementos de fragilidade de um sistema de IA, eventuais ataques sejam facilitados, interferindo na segurança do sistema.

Ao integrar *accountability*, precaução e transparência, a Avaliação de Impacto Algorítmico se apresenta como um processo holístico, que busca equilibrar a promoção de inovação tecnológica e a responsabilidade e prestação de contas. Esse tripé proporciona uma estrutura robusta para desenvolvimento da AIA, o que, em última medida, busca garantir que o instrumento não seja utilizado de forma meramente simbólica e retórica, mas como uma medida efetiva para prevenção e mitigação de riscos decorrentes do desenvolvimento e uso de sistemas de IA.

Por fim, apresenta-se abaixo a representação gráfica do tripé utilizado para delinear os contornos jurídicos da Avaliação de Impacto Algorítmico.

Figura 1 – Representação visual da compreensão da Avaliação de Impacto Algorítmico a partir do tripé formado pelos parâmetros de *accountability*, precaução e transparência



Fonte: Elaborado pela autora (2024).

6 AVALIAÇÕES DE IMPACTO ALGORÍTMICO: ABORDAGEM CONTEXTUAL À LUZ DA PROTEÇÃO DA PESSOA HUMANA

Conforme examinado anteriormente, as Avaliações de impacto desempenham um papel relevante no desenvolvimento e na implementação de sistemas de IA, possibilitando que tais sistemas não apenas cumpram com os objetivos pretendidos, mas também observem a promoção de direitos humanos por meio da prévia identificação e mitigação de riscos. Nesse sentido, as AIAs são instrumentos que carregam o potencial de garantir o desenvolvimento e a implementação responsável de sistemas de IA, possibilitando que desenvolvedores e implantadores identifiquem e abordem questões éticas, técnicas e jurídicas de maneira abrangente e proativa.

As AIAs, portanto, não se limitam a analisar a conformidade técnica e jurídica dos sistemas de IA, mas estendem seu escopo para avaliar os impactos sobre os direitos de pessoas e grupos afetados. O presente capítulo abordará duas situações de aplicação concreta da AIA para demonstrar sua relevância na proteção dos direitos humanos. A escolha pela análise de tais situações se justifica pela capacidade de fornecer uma análise detalhada e contextualizada de como as AIAs podem ser implementadas em diferentes situações, permitindo empreender as nuances e a complexidade envolvidas.

6.1 ANÁLISE DE SITUAÇÕES DE APLICAÇÃO PRÁTICA DA AVALIAÇÃO DE IMPACTO ALGORÍTMICO

O Canadá conta com a Diretiva sobre Tomada de Decisão Automatizada, que busca garantir que os sistemas de decisão automatizados sejam implementados de forma a reduzir os riscos para a sociedade canadense, além de resultar em decisões mais eficientes, precisas, consistentes e interpretáveis, tomadas de acordo com a legislação canadense (Canadá, 2019). Trata-se de uma norma aplicável a qualquer sistema, ferramenta ou modelo estatístico utilizado para tomar uma decisão ou realizar uma avaliação no âmbito do poder público.

Um dos objetivos da Diretiva é que as decisões sejam avaliadas e os eventuais resultados negativos sejam reduzidos, quando encontrados. Por tal razão, um dos requisitos impostos pela norma é a realização de uma Avaliação de Impacto Algorítmico, realizada a partir de ferramenta pré-definida. A ferramenta é um questionário que

determina o nível de impacto de um sistema de decisão automatizado. O questionário é composto de 51 (cinquenta e uma) perguntas sobre riscos e 34 (trinta e quatro) sobre mitigação. As pontuações de avaliação baseiam-se em diversos fatores, incluindo o projeto do sistema, o algoritmo, o tipo de decisão, o impacto e os dados.

Os impactos variam de acordo com critérios de reversibilidade e duração prevista, classificando decisões automatizadas em categorias distintas. Aquelas com impacto mínimo ou inexistente são consideradas reversíveis e de curta duração, ao passo que aquelas com impacto significativo são tidas como irreversíveis e de longa duração. Esses níveis de impacto são determinantes para as medidas de mitigação prescritas pela Diretiva de Tomada de Decisão Automatizada. Tais medidas incluem requisitos como revisão por pares, transparência, envolvimento direto de seres humanos (*human-in-the-loop*), explicabilidade e treinamento.

A ferramenta de Avaliação de Impacto Algorítmico foi elaborada com base em boas práticas de governança de sistemas de IA e em consulta com as partes interessadas internas e externas. Sendo assim, o modelo de AIA foi desenvolvido de forma aberta e está disponível ao público para compartilhamento e reutilização sob uma licença aberta. Para além da divulgação do modelo utilizado, o Governo do Canadá disponibiliza publicamente os resultados das avaliações já conduzidas.

Em razão da possibilidade de acesso à documentação de uma AIA, a presente pesquisa optou por utilizar o repositório canadense para fins de realização de uma análise da aplicação prática do instrumento em situações concretas. Em geral, verifica-se que avaliações conduzidas no âmbito do setor privado não são integralmente publicizadas, por exemplo, em razão da preservação de segredos de negócio, o que dificulta a utilização de tais insumos. Desse modo, para fins de realização desta análise, o presente trabalho utilizará duas AIAs conduzidas no âmbito da ferramenta disponibilizada pelo Governo do Canadá.

A primeira delas refere-se ao uso de decisões automatizadas em projeto de concessão de benefício de programa de saúde mental para veteranos de guerra do Canadá. A segunda, por sua vez, está relacionada a projeto de implementação de sistema de IA que busca simplificar a avaliação de elegibilidade em solicitações de permissão de trabalho no Canadá (*International Experience Canada Work Permit*) no âmbito do Departamento de Imigração, Refugiados e Cidadania.

Passando-se à análise do primeiro caso, de acordo com o Departamento de Assuntos de Veteranos do Canadá, responsável pelo preenchimento da AIA, atualmente

os veteranos precisam aguardar a decisão acerca da solicitação de benefícios por invalidez antes de se qualificarem para a cobertura de benefícios de tratamento médico, incluindo o programa de saúde mental. No entanto, o processo decisório pode ser demorado, atrasando o acesso aos benefícios. Por tal razão, o Departamento, por meio do programa em análise, busca permitir que os veteranos tenham acesso ao tratamento e suporte de saúde mental imediatamente após a solicitação por incapacidade, respeitado o limite de 2 (dois) anos, ou até que a decisão sobre o direito seja concedida.

Para tanto, o Departamento deseja introduzir automação no processo de tomada de decisão acerca da concessão do benefício, sinalizando que a principal motivação para a automatização é o acúmulo de trabalho e casos pendentes. A AIA sinaliza que a aplicação de um sistema de IA poderia gerar benefícios como a otimização de processos e automação do fluxo de trabalho, bem como a análise de grandes conjuntos de dados para identificação de anomalias, padrões de agrupamento e predições de resultados.

No momento de avaliação de riscos, foram analisados aspectos relevantes, como o fato de que o projeto seria realizado em uma área de intenso escrutínio público — por exemplo, devido a preocupações com a privacidade — e com litígios frequentes. Além disso, foi sinalizado que o público afetado é particularmente vulnerável e que as decisões são muito relevantes, com alto impacto para a saúde e bem-estar.

A partir da pontuação gerada pela AIA adotada pelo Governo do Canadá, o impacto é classificado em quatro níveis: nível 1, correspondente a impacto mínimo ou ausência de impacto; nível 2, correspondente a impacto moderado; nível 3, correspondente a alto impacto; e nível 4, correspondente a impacto muito alto. A AIA envolvendo o programa de acesso aos benefícios de saúde mental para veteranos resultou em um impacto de nível 2 (moderado), com permissão para que as decisões sejam tomadas sem envolvimento humano direto.

Diante do resultado, foram recomendadas medidas mitigatórias para implementação do projeto, incluindo a atuação de especialistas qualificados, a elaboração de aviso sobre o uso do sistema em linguagem simples e disponibilização em todos os canais em uso (internet, comunicação pessoal, correspondências ou telefone), e a garantia de que explicações significativas e compreensíveis poderiam ser fornecidas em caso de decisões negativas do benefício.

A partir desse caso, é possível verificar que a AIA conduzida abrange a documentação e avaliação de temas relevantes, como as características do sistema e do algoritmo a ser aplicado, as particularidades do contexto de tomada de decisão e do

público afetado e os controles adotados para uso de dados. Contudo, no caso de tecnologias que impactam diretamente os direitos humanos, é necessário que os processos de avaliação tenham como foco o público afetado, e não apenas nos aspectos gerenciais de desenvolvimento e aplicação do sistema.

Em tais casos, entende-se que a AIA deve ter ênfase no público afetado, assegurando que suas necessidades e preocupações sejam devidamente consideradas. Isso implica uma análise mais profunda das possíveis repercussões que a tecnologia pode ter sobre os direitos das pessoas afetadas, indo além dos aspectos técnicos e operacionais. Desse modo, compreende-se que a AIA deve levar em conta como as tecnologias em análise afetam os direitos humanos em termos práticos, considerando os impactos reais e potenciais sobre as vidas das pessoas e a sociedade em geral.

Nessa direção, o segundo caso relevante para a análise em questão é o de projeto de implementação de sistema de IA que busca simplificar a avaliação de elegibilidade em solicitações de permissão de trabalho no Canadá (*International Experience Canada Work Permit*), para auxiliar os tomadores de decisão do Departamento de Imigração, Refugiados e Cidadania do Canadá a processar as requisições com mais eficiência. O referido modelo faz a triagem das solicitações e agrupa arquivos com características semelhantes, com base nos requisitos legislativos, regulamentares e contratuais de cada subprograma e país participante.

O modelo foi projetado para aplicar critérios de triagem predefinidos (os mesmos que os tomadores de decisão já examinariam) e identificar casos “simples" em que a parte de requisição pode receber uma avaliação de elegibilidade positiva automatizada. Para solicitações em que a avaliação de elegibilidade positiva é automatizada pelo sistema, o sistema determina apenas que um solicitante é elegível, antes que a solicitação seja enviada a um tomador de decisões para fazer a triagem de admissibilidade e tomar a decisão final. Desse modo, o modelo nunca recusa ou recomenda a recusa de solicitações, e os funcionários do Departamento de Imigração, Refugiados e Cidadania do Canadá tomam a decisão final de aprovar ou recusar todas as solicitações.

No âmbito da avaliação de impacto, o Departamento argumentou que, para garantir a justiça e a transparência, as regras do sistema baseiam-se apenas em elementos de dados com um vínculo claro com requisitos legislativos, regulamentares e contratuais. Além disso, como salvaguarda adicional, o Departamento alega que as regras do sistema passam por um processo de revisão na medida em que, antes de serem introduzidas e, posteriormente, em intervalos regulares, são revisadas por funcionários experientes,

especialistas em direito, políticas e ciência de dados para garantir que sejam lógicas, compreensíveis, não discriminatórias e alinhadas com os critérios de elegibilidade estabelecidos.

Nesse sentido, o Departamento argumenta que o monitoramento regular e as medidas de garantia de qualidade ajudam a garantir que o sistema funcione como pretendido e que qualquer impacto negativo imprevisto, como existência de viés ou discriminação, possa ser identificado antecipadamente e atenuado. Diante das informações fornecidas, a avaliação resultou em um impacto de nível 2 (considerado moderado), que abrange as medidas mencionadas anteriormente.

No caso em questão, é interessante notar que houve, ainda, a determinação de realização de uma análise adicional baseada em gênero (*Gender-based Analysis Plus* – GBA+), que deve abordar os impactos do projeto (incluindo o sistema, os dados e a decisão) sobre o gênero e/ou outros fatores de identidade (como idade e localização), bem como medidas mitigatórias planejadas ou existentes para lidar com os riscos identificados, incluindo vieses e potenciais resultados discriminatórios. A partir da GBA+ é possível avaliar impactos de forma contextualizada a um grupo específico, o que proporciona melhor identificação e compreensão dos riscos associados à implementação do sistema de IA.

Desse modo, verifica-se que desenvolvimento de um modelo de Avaliação de Impacto Algorítmico, que se baseia em parâmetros específicos para gerar um cálculo de risco com base nas informações fornecidas, apresenta vantagens e desvantagens distintas. Uma vantagem desse modelo é a objetividade, uma vez que, ao seguir parâmetros claros e definidos, o modelo evita que o agente responsável pela avaliação apresente conclusões amplas e vagas, possibilitando uma análise mais precisa e fundamentada, o que é especialmente relevante para fins de controle social.

Além disso, modelos de AIA baseados em parâmetros pré-definidos e específicos podem ser mais fáceis de implementar e padronizar. A uniformidade nos critérios de avaliação pode simplificar o processo de auditoria e revisão, proporcionando uma abordagem sistemática e a possibilidade de comparação entre diferentes avaliações e sistemas.

Por outro lado, a objetividade também pode ser considerada uma desvantagem. A rigidez dos parâmetros pode reduzir a flexibilidade do modelo para se adaptar a novas informações ou mudanças no contexto. Além disso, modelos de AIA altamente objetivos

podem negligenciar a profundidade necessária para avaliar aspectos qualitativos ou mais sutis, especialmente aqueles relacionados a direitos humanos e questões éticas.

Por fim, uma desvantagem significativa é que a objetividade pode limitar a capacidade do modelo de considerar aspectos contextuais relevantes. Ao se concentrar estritamente nos parâmetros estabelecidos pelo modelo, a AIA pode deixar de capturar nuances e complexidades do mundo real, reduzindo sua eficácia em certas situações. Sendo assim, embora a objetividade e a clareza sejam benefícios importantes para a avaliação de riscos, entende-se que a AIA deve incorporar uma análise aprofundada e contextualizada, tendo em vista impactos para direitos humanos.

6.2 ABORDAGEM CONTEXTUAL E PROTEÇÃO DE DIREITOS HUMANOS

Para que a AIA possa considerar a proteção de direitos de pessoas pertencentes a diferentes grupos sociais, entende-se que é necessário adotar uma abordagem contextual de gerenciamento de riscos. Nessa direção, Mantelero (2022) esclarece que a avaliação específica realizada caso a caso é mais eficaz em termos de prevenção e mitigação de riscos do que estratégias pautadas em presunções de risco, baseadas em uma classificação abstrata de setores de alto risco ou usos/finalidades de alto risco.

Mantelero (2022) aponta que setores, usos e finalidades são categorias muito amplas, que, consequentemente, englobam diferentes tipos de aplicações de sistemas de IA — por vezes, sistemas em constante evolução — que acarretam uma variedade de potenciais impactos para direitos e liberdades e não podem ser agrupados com base em limitações de risco realizadas de forma prévia e abstrata. Acerca da FRIA, presente no contexto regulatório europeu, Mantelero (2024) esclarece, inclusive, que uma abordagem como foco apenas no sistema de IA ignora o fato de que esses produtos ou serviços operam em contextos específicos e que essa dimensão contextual, especialmente no que diz respeito ao impacto sobre os indivíduos, é relevante para a prevenção de riscos.

Nesse sentido, Mantelero (2024) aponta, por exemplo, que ao limitar a avaliação de riscos àqueles que “podem ser razoavelmente mitigados ou eliminados por meio do desenvolvimento ou do projeto do sistema de IA de alto risco, ou do fornecimento de informações técnicas adequadas”, o *AI Act* não reconhece que os sistemas de IA são frequentemente sistemas sociotécnicos e, por isso, o *design* a ser considerado não é apenas o *design* do sistema de IA, mas também o *design* resultante da interação e da modificação mútua que esses sistemas geram na sociedade.

Para ilustrar este cenário, Mantelero (2024) sugere como exemplo o fato de que há uma diferença entre o uso de um sistema de IA para suporte à decisão de autoridades públicas competentes no contexto de emergências humanitárias e o uso do mesmo sistema em condições normais, pois o estado de estresse de todas as pessoas envolvidas no primeiro cenário pode exacerbar a qualidade dos dados, a interação deficiente entre humanos e IA e os vieses. No mesmo sentido, Metcalf, Moss, Watkins, Singh e Elish (2023) destacam que os danos são inerentemente dependentes do contexto, pois afetam indivíduos e comunidades devido às particularidades de suas próprias circunstâncias.

Por tal razão, é importante que a avaliação de danos do "mundo real" faça parte da estrutura da AIA. A análise das implicações dos sistemas de IA deve levar em consideração os direitos e interesses das partes envolvidas, mas é essencial reconhecer que diferentes grupos de pessoas são afetados de maneiras distintas, mesmo quando interagem com sistemas de IA semelhantes (Negri *et al*., 2024). A complexidade dos impactos se deve ao fato de que as experiências e consequências para esses grupos variam amplamente com base no contexto específico de implementação e nas características individuais e sociais de cada grupo. As características específicas de cada grupo, como status socioeconômico, histórico de crédito, e contextos sociais e políticos, influenciam como eles são impactados.

Nesse sentido, Brown, Davidovic e Hasan (2021) destacam que a análise contextual é essencial para compreender que os grupos afetados não são homogêneos. Em vez disso, eles são compostos por indivíduos que vivenciam diferentes ameaças e impactos baseados em suas circunstâncias particulares. Esses impactos não são apenas uma função do tipo de interação com o sistema de IA, mas também são moldados por fatores sociopolíticos e dinâmicas de poder que variam entre os grupos. Nesse processo, é importante reconhecer que o desenvolvimento da AIA pressupõe a percepção de que os riscos associados aos usos de novas tecnologias podem não se distribuir de forma linear entre pessoas e grupos, especialmente no caso de grupos historicamente marginalizados e vulnerabilizados, que são recorrentemente expostos e submetidos a opressões e violências (Machado; Negri; Giovanini, 2023).

Para além de questões técnicas, os aspectos sociais, históricos, culturais e geográficos também influenciam o desenvolvimento e a implementação de sistemas de IA e, por isso, devem ser considerados durante a elaboração de uma Avaliação de Impacto Algorítmico. Nesse ponto, cabe, inclusive, a incorporação de uma perspectiva decolonial

do Sul Global, que considere as especificidades e diferenças culturais e históricas entre os países.

Uma perspectiva decolonial implica a necessidade de se considerar as desigualdades sociais e econômicas, uma vez que novas tecnologias podem perpetuar as injustiças já existentes, caso não sejam desenvolvidas e implementadas de forma responsável. Essas particularidades contextuais devem ser abordadas pela AIA, exigindo uma análise crítica dos impactos sociais, considerando as experiências históricas marcadas por colonização, exploração e marginalização. Desse modo, uma abordagem decolonial requer a incorporação de perspectivas e dinâmicas locais e o reconhecimento de diversidades culturais, geográficas e históricas.

Conforme aponta Nas (2023), a decolonização envolve o reconhecimento de que o conhecimento é produzido a partir de nossas experiências individuais e culturais. A autora destaca que, para efetivar a decolonização em processos de escolha e decisão, é importante compreender que não há categorias universais absolutas. Em vez disso, é necessário explorar e debater os pontos de interseção e consenso entre as diversas culturas que formam a contemporânea cultura globalizada.

Diante desse cenário, entende-se que é necessário reconhecer que o desenvolvimento e a implementação de sistemas de IA são processos profundamente influenciados por experiências individuais e coletivas. Ao invés de adotar uma abordagem universalista que pressupõe uma verdade objetiva e única, a decolonização convida a considerar as múltiplas perspectivas e saberes que surgem de diferentes contextos envolvendo o uso de sistemas de IA. Segundo Nas (2023), esse processo implica em questionar os conhecimentos tidos como universais, incluindo suas teorias, práticas e métodos. Adicionalmente, a autora destaca que a decolonização envolve a promoção de perspectivas diversas daquelas que são predominantemente baseadas na autoridade dos conhecimentos acumulados.

Na mesma direção, Arun (2019) aponta que, além de mudar como as decisões sobre *design*, dados e implantação de tecnologias na sociedade são tomadas, é necessário fornecer às populações do Sul as ferramentas para se envolverem de maneira eficaz com as questões que as afetam. Para Arun (2019), os marcos institucionais dos países do Sul devem ser levados em conta em análises sobre qual impacto a IA pode ter no Sul, incluindo não apenas aspectos relacionados aos direitos políticos e civis, mas também de outras questões sociais e econômicas, como educação e saúde.

O presente trabalho não aprofundará em discussões acerca de perspectivas decoloniais e debates sobre colonialismo digital porque, dado o tema específico de Avaliações de Impacto Algorítmico, optou-se por direcionar o foco para outras abordagens teóricas, as quais melhor se alinham com os objetivos e a estrutura da investigação, com ênfase na análise de aspectos normativos. No entanto, compreende-se que é necessário reconhecer a importância da abordagem decolonial para o desenvolvimento de Avaliações de Impacto Algorítmico, pois desafia as estruturas de poder e os padrões de pensamento dominantes que podem estar embutidos em sistemas de IA.

Por tal razão, é importante que as AIAs sejam desenvolvidas com aplicação de abordagem contextual, o que implica considerar as especificidades de cada comunidade, ou contexto, e como determinado sistema de IA pode afetá-los de maneiras diferentes. Assim, é fundamental considerar os interesses coletivos durante a avaliação de impacto de sistemas de inteligência artificial. Isso se deve ao fato de que muitos modelos de negócios lidam com dados pessoais para fins de perfilização, classificação e monitoramento do comportamento de grupos, revelando a tensão entre pessoa e mercado (Negri *et al*., 2024). Por exemplo, um sistema de IA que automatiza processos de contratação pode, inadvertidamente, perpetuar preconceitos existentes, caso não sejam consideradas as disparidades históricas e estruturais no acesso a oportunidades de emprego.

Um caso que poderia ter sido explorado em uma perspectiva contextual é foi a suspensão pela Google da criação de imagens de pessoas por meio da IA Gemini. A controvérsia surgiu quando a ferramenta começou a gerar imagens de figuras históricas com diversidade étnica e de gênero que não correspondiam aos dados históricos, como oficiais nazistas negros e um Papa mulher indiana. Conforme aponta Souza (2024), a situação destaca os desafios de calibragem de algoritmos para atender à demanda por inclusão sem perder precisão histórica. A empresa justificou que o modelo Gemini falhou ao tentar aplicar diversidade em contextos inadequados e, ao mesmo tempo, tornou-se excessivamente conservador em outras situações.

No caso da IA Gemini do Google, é razoável vislumbrar que a aplicação de uma AIA com perspectiva contextual poderia ter identificado previamente os riscos associados à representação histórica de figuras públicas com diversidade étnica e de gênero, antecipando potenciais conflitos culturais e evitando a geração de conteúdo considerado inadequado. Ao examinar o impacto do modelo em comunidades específicas,

especialmente em questões sensíveis como representações históricas e raciais, a empresa poderia ter identificado que a introdução de diversidade em figuras historicamente reconhecidas por suas características étnicas ou culturais poderia ser percebida como uma distorção dos fatos. Em síntese, uma abordagem contextual permitiria não apenas identificar riscos jurídicos, mas também aqueles relacionados à percepção pública e às dinâmicas culturais.

Nesse sentido, entende-se que uma abordagem contextual na elaboração de Avaliações de Impacto Algorítmico é fundamental para a promoção dos direitos humanos, na medida em que possibilita identificar e mitigar riscos associados a um determinado contexto, permitindo compreensão específica de dinâmicas sociais, culturais e históricas que moldam a realidade de cada comunidade. Referida abordagem contextual pode auxiliar na identificação de desigualdades e violações já existentes, de modo que a AIA assume relevância como um instrumento para evitar a perpetuação ou ampliação de tais injustiças.

A título de exemplificação, no âmbito da União Europeia, o AI Act adota, em determinadas disposições, abordagens contextuais que podem auxiliar na realização de avaliações de impacto para direitos humanos. Nesse sentido, é possível citar que, ao tratar das obrigações de governança de dados aplicáveis aos fornecedores de sistemas de IA de risco elevado, a regulação prevê que os conjuntos de dados devem levar em consideração as características ou os elementos que são particulares do enquadramento geográfico, contextual, comportamental ou funcional específico no qual o sistema de IA se destina a ser utilizado.

Outro exemplo relevante é que o fornecedor de um sistema de IA de alto risco deve fornecer instruções detalhadas aos responsáveis por sua implantação. Essas informações devem incluir as características, capacidades e limitações de desempenho do sistema, abrangendo, entre outros aspectos, seu desempenho em relação a determinadas pessoas ou grupos específicos para os quais o sistema foi projetado. Desse modo, ainda que o AI Act não determine explicitamente a adoção de uma abordagem contextual para a avaliação de impactos, essa interpretação pode ser inferida a partir de dispositivos que exigem que fornecedores e responsáveis pela implantação considerem os fatores que envolvem ou influenciam determinado sistema de IA, como o ambiente físico e social em que o sistema está inserido.

De toda forma, compreende-se que abordagem de desenvolvimento da Avaliação de Impacto Algorítmico (AIA) proposta neste trabalho pode enfrentar objeções.

Primeiramente, uma abordagem contextualizada pode aumentar a complexidade do processo, exigindo a consideração de uma vasta gama de variáveis e nuances específicas de cada situação. Essa abordagem pode demandar recursos adicionais, tanto financeiros quanto humanos, como a realização de pesquisas de campo, entrevistas com pessoas afetadas e a participação direta do público. Além disso, a aplicação universal dos resultados da AIA pode ser dificultada, uma vez que suas conclusões são fortemente influenciadas pelo contexto específico em que foram obtidas, o que poderia limitar a capacidade de extrapolar esses resultados para outros contextos ou cenários.

Nesse sentido, embora este trabalho sustente que uma abordagem contextualizada na elaboração de AIA pode contribuir para a promoção de direitos humanos, reconhece-se que há complexidades e desafios significativos em sua implementação. Diante desse contexto, entende-se que uma solução parcial seria o desenvolvimento de diretrizes gerais e padrões claros para a elaboração da AIA, o que poderia ajudar a garantir a consistência e a qualidade dos resultados, estabelecendo critérios para a seleção de contextos relevantes. Assim, buscando auxiliar na construção de uma metodologia para desenvolvimento da AIA a partir de uma abordagem contextual, retoma-se a última questão elencada para desenvolvimento do problema de pesquisa apresentado neste trabalho: como a AIA pode ser realizada considerando aspectos contextuais relacionados ao desenvolvimento e/ou ao uso de sistemas de IA?

Primeiramente, é essencial mapear o contexto em que o sistema de IA será implantado. Isso inclui a análise de fatores como a delimitação do público-alvo, a identificação de pessoas ou grupos que possam ser impactados além desse público, o ambiente socioeconômico, aspectos culturais e políticos relevantes, além de fatores temporais ou momentâneos que possam influenciar os impactos gerados pelo sistema de IA, como eleições, calamidades públicas, entre outros.

Além disso, conforme identificado anteriormente, a AIA deve identificar desigualdades e violações de direitos já existentes, que podem ser exacerbadas ou perpetuadas pelos sistemas de IA. Para tanto, deve englobar análise dos possíveis impactos do sistema de IA sobre grupos marginalizados e vulneráveis, o que pode ser realizado por meio de coleta e levantamento de informações que reflitam a diversidade das comunidades afetadas. A abordagem contextual, nesse sentido, enfatiza a importância de incluir as vozes e perspectivas dessas comunidades no processo de avaliação, reconhecendo sua expertise em relação às suas próprias experiências e necessidades.

Em última instância, a análise deve considerar que grupos marginalizados tendem a ser mais afetados por novas tecnologias, mas frequentemente são menos representados nos processos estratégicos de tomada de decisão. A inclusão dessas vozes no desenvolvimento da AIA é importante para garantir uma contextualização adequada, permitindo que a identificação e mitigação de impactos sejam mais eficazes. Em síntese, o envolvimento direto das pessoas afetadas no processo de análise enriquece a AIA, proporcionando uma compreensão mais detalhada e multifacetada de suas necessidades e preocupações reais.

Desse modo, a contextualização dos processos de identificação e mitigação de impactos, levando em conta as nuances culturais, sociais e econômicas das pessoas afetadas, torna a AIA mais sensível e relevante para situações específicas. Além disso, a inclusão de diferentes perspectivas promove a detecção de problemas e desafios que poderiam ser negligenciados em abordagens mais convencionais.

Evidentemente, operacionalizar a inclusão das vozes e perspectivas das comunidades afetadas no desenvolvimento de Avaliações de Impacto Algorítmico requer um esforço deliberado e estruturado para garantir que suas contribuições sejam efetivamente consideradas. Para incorporar uma abordagem contextual à AIA, os responsáveis pelo seu desenvolvimento podem adotar ações como: realizar entrevistas e consultas com membros das comunidades afetadas para entender suas preocupações, experiências e necessidades em relação ao sistema de IA; colaborar com organizações comunitárias e grupos locais para facilitar o envolvimento das pessoas afetadas; e implementar mecanismos para a coleta de *feedback* regular, permitindo ajustes contínuos no sistema de IA.

A partir do modelo estruturado entre elementos “ator” e “fórum”, apresentado anteriormente pelo presente trabalho, é possível compreender a participação pública como uma modalidade de fórum externo. Nesse contexto, o ator deve justificar suas escolhas em um ambiente onde será interrogado, avaliado e sujeito a sugestões de alterações no processo apresentado. No entanto, conforme discutido anteriormente, nem sempre o fórum possui a capacidade de impor consequências ao ator. No caso da participação pública, essa característica é particularmente relevante, pois a ausência de competências institucionais, como a capacidade de impor sanções, limita o impacto direto que o fórum pode ter sobre o ator.

A falta de consequências pode dificultar a implementação das recomendações ou sugestões resultantes do fórum externo de participação pública, uma vez que os atores

podem não se sentir obrigados a agir de acordo com as decisões do público. Esse cenário pode criar uma lacuna entre as discussões realizadas no fórum e as ações concretas efetivamente executadas pelo agente. Portanto, assim como o desenvolvimento da AIA deve adotar uma abordagem contextual, a definição de um fórum de participação pública também deve ser contextualizada, pois a natureza dos fóruns pode variar dependendo do estágio do ciclo de vida do sistema de IA e dos agentes envolvidos.

Desse modo, entende-se que, ao identificar contextualmente a necessidade de estruturar um fórum externo de participação pública, é essencial que ele seja complementado por fóruns (internos ou externos) com poder para impor consequências. Por exemplo, um fórum externo composto por autoridades competentes, capazes de tomar decisões vinculativas, ou um fórum interno com atribuições organizacionais significativas, seria necessário para garantir que as recomendações e sugestões sejam efetivamente implementadas e que os atores sejam responsabilizados por suas ações.

A abordagem contextual na Avaliação de Impacto Algorítmico (AIA) não se restringe apenas à identificação de riscos e desafios específicos enfrentados por pessoas afetadas, mas busca também promover ativamente seus direitos e interesses. Ao integrar essa perspectiva, a AIA deve considerar não apenas as métricas tradicionais de desempenho, como precisão e eficiência, mas também os fatores históricos, éticos e sociais que podem ter contribuído para a vulnerabilização de certos grupos. Em síntese, a abordagem contextual reconhece que os impactos gerados por sistemas de IA não são lineares, podendo variar significativamente dependendo do contexto social, cultural e econômico em que são aplicados e, até mesmo, perpetuar ou agravar desigualdades já existentes.

A participação pública, por si só, também exige uma avaliação contextual que verifique, em primeiro lugar, a necessidade de realização deste processo e, posteriormente, o poder de influência deste fórum no processo de desenvolvimento ou aplicação de determinado sistema de IA. Assim, entende-se que a participação pública não deve ser vista como um processo automático, mas como uma ferramenta que poderá ser aplicada em determinados contextos, quando necessário e efetivamente possível.

Desse modo, a compreensão da AIA a partir de uma abordagem contextual permite traçar as seguintes premissas acerca deste instrumento:

1) A AIA pode ser elaborada pelo agente responsável pelo desenvolvimento de um sistema de IA ou pelo agente responsável pela implementação prática do sistema.

Tais avaliações não são excludentes, mas complementares, pois possuem focos distintos.

2) A AIA conduzida no processo de desenvolvimento possui ênfase na antecipação e identificação de potenciais impactos decorrentes da forma como o sistema foi concebido e programado, ou seja, são impactos previstos antes mesmo que o sistema seja implantado em um ambiente.
3) Por outro lado, a AIA conduzida durante a implementação prática do sistema de IA considera um contexto real, no qual já é possível identificar parâmetros como (i) público afetado e seu respectivo contexto socioeconômico e cultural; (ii) características das pessoas que possuem interação com o sistema, incluindo aspectos relacionados a vulnerabilidades; (iii) desigualdades e disparidades já existentes no ambiente de implementação do sistema de IA etc. Nesse sentido, o foco principal está na monitorização contínua dos efeitos do algoritmo e na adaptação às mudanças no ambiente operacional.
4) A AIA deve avaliar o impacto de um sistema de IA sobre os direitos humanos das pessoas afetadas, sendo essencial reconhecer que os impactos dessas iniciativas podem variar significativamente com base nas características individuais e contextuais das pessoas afetadas. Isso inclui avaliar os impactos sob uma ótica que considere diferentes perspectivas, incluindo gênero, raça, orientação sexual, deficiência e características identitárias relevantes.
5) A avaliação de impacto aos direitos humanos, a ser realizada no âmbito da AIA, refere-se exclusivamente aos direitos de pessoas naturais que são afetadas pelo desenvolvimento ou implantação de um determinado sistema de IA. Em síntese, significa dizer que não se trata de reconhecimento de titularidade de direitos a entes abstratos e, tampouco, de avaliação de interesses associados ao próprio sistema de IA.
6) A AIA não deve ser confundida com um instrumento de gerenciamento de riscos corporativos. Embora práticas de *accountability* e gerenciamento de riscos sejam comuns nas rotinas de compliance de diversas organizações, é importante distinguir entre esses instrumentos tradicionais e aqueles destinados a lidar com os riscos associados às novas tecnologias que afetam direitos humanos. Os instrumentos tradicionais de *compliance* concentram-se na gestão de riscos que impactam diretamente a organização e suas operações, buscando proteger a integridade do negócio e garantir a conformidade com regulamentações. Esses

> mecanismos avaliam como as violações podem afetar a própria atividade empresarial. Por outro lado, a AIA se dedica a avaliar e mitigar os impactos específicos que sistemas de IA podem ter sobre os direitos das pessoas afetadas, indo além das preocupações tradicionais com o risco corporativo. Em vez de se concentrar exclusivamente nos riscos para a organização, entende-se que a AIA deve priorizar a proteção dos direitos humanos e a minimização dos impactos adversos que a tecnologia pode ter sobre indivíduos e comunidades.

Evidentemente, a adoção de uma abordagem contextual na Avaliação de Impacto Algorítmico não é uma solução definitiva ou única para todos os desafios associados ao desenvolvimento e aplicação de sistemas de IA. A AIA contextualiza a análise dos impactos considerando aspectos específicos dos grupos afetados, suas condições socioeconômicas, culturais e políticas, o que ajuda a identificar e mitigar riscos de forma mais precisa e eficaz. No entanto, essa abordagem deve ser vista como parte de um quadro mais amplo de proteção dos direitos humanos.

O problema de pesquisa abordado pelo presente também deve ser analisado em perspectiva ampla a partir das dinâmicas e estruturas de poder que permeiam a sociedade. Nesse sentido, compreende-se que a AIA deve ser complementada por políticas públicas e regulamentações que abordem dinâmicas sociopolíticas já existentes. Além disso, fortalecer mecanismos sólidos de supervisão e fiscalização pode ser relevante para garantir que as práticas de AIA sejam implementadas de forma eficaz e que os direitos das pessoas afetadas sejam protegidos de maneira consistente.

Em conclusão, a implementação de uma abordagem contextual no processo de desenvolvimento da Avaliação de Impacto Algorítmico pode representar um avanço significativo na proteção dos direitos humanos para grupos marginalizados e em situação de vulnerabilidade, mas é importante reconhecer suas limitações e a necessidade de complementação deste arranjo com outros aspectos, como regulamentações robustas e fortalecimento de mecanismos sólidos de supervisão e fiscalização.

7 CONCLUSÃO

O avanço da tecnologia trouxe consigo uma proliferação significativa de sistemas de inteligência artificial, os quais têm se integrado de forma cada vez mais intrínseca ao cotidiano das pessoas. Desde assistentes virtuais em *smartphones* até algoritmos de recomendação em plataformas de *streaming*, a presença da IA pode ser considerada como ubíqua. No entanto, à medida que esses sistemas se tornam onipresentes, é imperativo reconhecer os potenciais riscos e impactos que acompanham seu desenvolvimento e implementação, bem como os desafios éticos e sociais que os acompanham.

Ao mesmo tempo em que o avanço tecnológico promove benefícios e ganhos em produtividade e eficácia, é necessário assegurar a existência de instrumentos de identificação e mitigação de riscos associados às novas tecnologias. Trata-se, em última instância, do reconhecimento de que o progresso não está isento de riscos e desafios significativos. A busca por um equilíbrio entre inovação e responsabilidade torna-se, assim, um imperativo para garantir que sistemas de IA continuem a agregar valor à sociedade em observância aos direitos humanos e às diretrizes éticas.

Conforme discutido no presente trabalho, a AIA envolve a análise dos possíveis impactos que um determinado sistema de IA pode gerar para as pessoas afetadas por seu funcionamento. Dessa forma, ao ser desenvolvida sobre o tripé de *accountability*, precaução e transparência, a AIA pode ser compreendida como um instrumento de responsabilidade e prestação de contas, na medida em que as considerações e justificativas apresentadas são avaliadas em fórum com poderes para questionamentos e alterações. Além disso, a partir da precaução, a AIA é, também, um instrumento de identificação de riscos e implementação de medidas mitigatórias. Por fim, sob a perspectiva da transparência, a AIA é um instrumento que possibilita o acesso a informações por parte do fórum avaliador e, consequentemente, permite as avaliações e debates acerca do impacto de determinados sistemas de IA.

Após a compreensão dos contornos jurídicos da AIA, resta, então, a necessidade de retornar ao problema de pesquisa deste trabalho: em que medida a Avaliação de Impacto Algorítmico (AIA) pode ser considerada um instrumento eficaz na mitigação de riscos para direitos humanos, associados ao desenvolvimento e uso de sistemas de IA em diferentes contextos sociais, econômicos e territoriais?

Conforme é possível extrair da pesquisa realizada, a AIA pode ser considerada um instrumento para mitigação de riscos para os direitos humanos associados ao

desenvolvimento e uso de sistemas de IA em diversos contextos sociais, econômicos e territoriais, porém, sua efetividade deve ser avaliada, principalmente, diante de dois parâmetros: (i) contextualização social e econômica; e (ii) estruturação de fórum apto a questionar e avaliar as escolhas e decisões tomadas.

Em primeiro lugar, a eficácia da AIA enquanto instrumento para mitigação de riscos e violações a direitos humanos está diretamente relacionada ao reconhecimento de que impactos gerados por sistemas de IA podem variar significativamente dependendo do contexto em que são implementados. Os impactos de sistemas de IA não são homogêneos, motivo pelo qual as particularidades sociais, econômicas e culturais de pessoas e grupos afetados devem ser levadas em consideração pela AIA, especialmente diante de grupos marginalizados e marcados por vulnerabilidades.

O segundo parâmetro diz respeito à estruturação da AIA a partir do modelo ator-fórum, no qual é essencial que haja um fórum (interno ou externo) com poderes para questionar e avaliar as escolhas, decisões e justificativas apresentadas pelo agente responsável pelo desenvolvimento ou implementação de determinado sistema de IA, bem como competência para impor alterações e, em caso de descumprimento das solicitações apresentadas, consequências. Caso contrário, a AIA pode se tornar uma formalidade burocrática sem impacto substancial na mitigação de riscos para direitos humanos.

Sendo assim, o desenvolvimento da Avaliação de Impacto Algorítmico a partir dos parâmetros mencionados anteriormente pode representar um avanço importante na proteção dos direitos humanos, especialmente para grupos marginalizados e em situação de vulnerabilidade. No entanto, é essencial reconhecer que essa abordagem não é uma solução única e definitiva para todas as questões relacionadas à proteção dos direitos humanos diante do desenvolvimento e aplicação de sistemas de IA. É essencial considerar, também, as dinâmicas e estruturas de poder que permeiam a sociedade em uma análise mais ampla.

Portanto, a AIA — enquanto instrumento de governança — deve ser complementada por outras ferramentas, medidas e controles, como regulamentações robustas e fortalecimento de mecanismos de supervisão e governança. Regulamentações eficazes podem estabelecer padrões claros de responsabilidade e transparência para os agentes que desenvolvem ou implementam sistemas de IA, enquanto os mecanismos de supervisão e governança podem garantir o cumprimento dessas regulamentações e a prestação de contas adequada.

REFERÊNCIAS

ADA LOVELACE INSTITUTE. Meaningful public participation and AI: lessons and visions for the way forward. Blog, 2024. Disponível em: https://www.adalovelaceinstitute.org/blog/meaningful-public-participation-and-ai/. Acesso em 03 fev. 2024.

AMARAL, João. J. F. **Como fazer uma pesquisa bibliográfica**. Fortaleza, CE: Universidade Federal do Ceará, 2007. Disponível em: http://200.17.137.109:8081/xiscanoe/courses1/mentoring/tutoring/Como%20fazer%20pesquisa%20bibliografica.pdf. Acesso em: 24 dez. 2023.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New Media & Society**. v. 20. n. 3. dez. 2016, p.973-989. Disponível em: https://journals.sagepub.com/doi/10.1177/1461444816676645. Acesso em: 04 jan. 2024.

ARUN, Chinmayi. AI and the Global South: Designing for Other Worlds. In DUBBER, Markus D; PASQUALE, Frank; DAS, Sunit. **The Oxford Handbook of Ethics of AI**. Oxford, Inglaterra: Oxford University Press, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403010. Acesso em: 22 dez. 2023.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. Relatório de Impacto à Proteção de Dados Pessoais (RIPD), 2013. Disponível em: https://www.gov.br/anpd/pt-br/canais_atendimento/agente-de-tratamento/relatorio-de-impacto-a-protecao-de-dados-pessoais-ripd. Acesso em: 19 fev. 2024.

BABBIE, Earl. **The Practice of Social Research**. Canadá: Wadsworth, Cengage Learning, 2013. Disponível em: < http://old-eclass.uop.gr/modules/document/file.php/SEP187/BI%CE%92%CE%9B%CE%99%CE%91%20%CE%9C%CE%95%CE%98%CE%9F%CE%94%CE%9F%CE%9B%CE%9F%CE%93%CE%99%CE%91%CE%A3/Babbie_The_Practice_of_Social_Research.pdf>. Acesso em: 24 dez. 2023.

BALKIN, Jack. The path of robotics law. **California Law Review Circuit**, Berkeley, v. 06, p. 45-60, jun. 2015.

BIONI, Bruno; LUCIANO, Maria. O princípio da precaução da regulação da inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada? *In:* FRAZÃO, Ana; MULHOLLAND, Caitlin (org.). **Inteligência Artificial e Direito**. São Paulo: Thomson Reuters Brasil, 2019. p. 207-232.

BOSTROM, Nick. **Superinteligência**: perigos, caminhos e estratégias para um novo mundo. [S.L]: Darkside, 2018. 480 p.

BOVENS, Mark. Analysing and Assessing Accountability: a conceptual framework. **European Law Journal**, [S. L], v. 13, n. 4, p. 447-468, jul. 2007. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0386.2007.00378.x. Acesso em: 21 mai. 2023.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados (LGPD). Brasília, DF: Diário Oficial da União, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 03 mai. 2023.

BRASIL. **Lei nº 6.938, de 31 de agosto de 1981**. Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formulação e aplicação, e dá outras providências. Brasília, DF: Diário Oficial da União, 1981. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l6938.htm. Acesso em: 18 out. 2023.

BRASIL. **Projeto de Lei n. 5051, de 2019**. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/138790. Acesso em: 05 nov. 2023.

BRASIL. **Projeto de Lei n. 21, de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: https://www.camara.leg.br/propostas-legislativas/2236340. Acesso em: 05 nov. 2023.

BRASIL. **Projeto de Lei n. 872, de 2021**. Dispõe sobre os marcos éticos e as diretrizes que fundamentam o desenvolvimento e o uso da Inteligência Artificial no Brasil. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/147434. Acesso em: 05 nov. 2023.

BRASIL. **Projeto de Lei n. 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 05 nov. 2023.

BRASIL. **IA para o bem de todos**: proposta de plano brasileiro de inteligência artificial 2024-2028. 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: 10 ago. 2024.

BROUSSARD, Meredith. **More than a glitch**: confronting race, gender and ability bias in tech. Londres: The Mit Press, 2023. 211 p.

BROWN, Shea; DAVIDOVIC, Jovana; HASAN, Ali. The algorithm audit: scoring the algorithms that score us. **Big Data & Society**, *[S.l.]*, v. 8, n. 1, p. 1-8, jan. 2021. SAGE Publications. http://dx.doi.org/10.1177/2053951720983865. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951720983865. Acesso em: 21 nov. 2023.

BURWELL V. HOBBY LOBBY STORES. INC. Disponível em: www.supremecourt.gov/opinions/13pdf/13-354_olp1.pdf. Acesso em: 05 nov. 2023.

CANADÁ. Digital Charter Implementation Act, 2022. Disponível em: https://www.parl.ca/legisinfo/en/bill/44-1/c-27. Acesso em: 05 nov. 2023.

CANADÁ. Directive on Automated Decision-Making, 2023. Disponível em: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592. Acesso em: 21 nov. 2023.

CASTRO, Daniel; MCLAGHLIN, Michael. Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence. **Information Technology And Innovation Foundation**, [S.I], p. 1-35, fev. 2019. Disponível em: https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/. Acesso em: 04 jan. 2024.

CERKA, Paulius. GRIGIENE, Jurgita. SIRBIKYTE, Gintare. Liability for damages caused by Artificial Intelligence, **Computer Law & Security Review**, Elsevier, v.31, n.3, p.376-389, jun. 2015, p.378.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2002.

CIPL. **Artificial Intelligence and Data Protection in Tension**, 2024. Disponível em: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_first_ai_report_-_ai_and_data_protection_in_tension__2_.pdf. Acesso em: 13 jul. 2024.

COECKELBERGH, Mark. **AI Ethics**. Londres: The Mit Press, 2020. 250 p.

CONSELHO EUROPEU. AI Glossary. Disponível em: https://www.coe.int/en/web/artificial-intelligence/glossary. Acesso em: 05 nov. 2023.

COOLEY, Thomas McIntyre. **A treatise on the law of torts**. Chicago: Callaghan, 1880. Disponível em: https://repository.law.umich.edu/books/11/. Acesso em: 13 jul. 2023.

CORTIZ, Diogo. Inteligência Artificial: equidade, justiça e consequências. **Panorama Setorial da Internet**: N.° 1, Ano 12, maio de 2020, pp. 1-5

COSTA, Luiz. Privacy and the precautionary principle. **Computer Law & Security Review**, [s. l], v. 28, n. 1, p. 14-24, fev. 2012. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0267364911001804?via%3Dihub. Acesso em: 31 jan. 2023.

CRAWFORD, Kate. **Atlas of AI**: power, politics, and the planetary costs of artificial intelligence. Londres: Yale University Press, 2021. 336 p.

CJSUBIA. Relatório final, 2022. Disponível em: https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf. Acesso em: 05 nov. 2023.

DIGICHINA. Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022. Disponível em: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/. Acesso em: 03 abr. 2023.

DONEDA, Danilo Cesar Maganhoto; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; AN, Norberto Nuno Martin Becerra Gomes de. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar - Revista de Ciências Jurídicas**, [S.L.], v. 23, n. 4, p. 1-17, 20 dez. 2018. Fundacao Edson Queiroz. Disponível em: https://ojs.unifor.br/rpen/article/view/8257. Acesso em: 23 dez. 2023.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**: elementos da lei geral de proteção de dados. 2. ed. São Paulo, SP: Thomson Reuters Brasil, 2019, p.23.

FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica; CHATILA, Raja; CHAZERAND, Patrice; DIGNUM, Virginia; LUETGE, Christoph; MADELIN, Robert; PAGALLO, Ugo; ROSSI, Francesca; SCHAFER, Burkhard; VALCKE, Peggy; VAYENA, Effy. AI4People—An Ethical Framework for a Good AI Society: opportunities, risks, principles, and recommendations. **Minds And Machines**, *[S.l.]*, v. 28, n. 4, p. 689-707, 26 nov. 2018. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11023-018-9482-5. Disponível em: https://link.springer.com/article/10.1007/s11023-018-9482-5. Acesso em: 05 nov. 2023.

FONSECA, João José Saraiva. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila. Disponível em: < http://www.ia.ufrrj.br/ppgea/conteudo/conteudo-2012-1/1SF/Sandra/apostilaMetodologia.pdf>. Acesso em: 24 dez. 2023.

FOX, Jonathan. The uncertain relationship between transparency and accountability. **Development In Practice**, [S.L.], v. 17, n. 4-5, p. 663-671, ago. 2007. Informa UK Limited. http://dx.doi.org/10.1080/09614520701469955. Disponível em: https://www.tandfonline.com/doi/full/10.1080/09614520701469955. Acesso em: 4 abr. 2024.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. São Paulo, SP: Atlas, 2002.

GODOY, Arilda Schmidt. Pesquisa Qualitativa: tipos fundamentais. **Revista de Administração de Empresas**, São Paulo, v. 26, n. 2, 1995. Disponível em: < https://www.scielo.br/j/rae/a/ZX4cTGrqYfVhr7LvVyDBgdb/?lang=pt>. Acesso em: 24 dez. 2023.

HASAN, Ali; BROWN, Shea; DAVIDOVIC, Jovana; LANGE, Benjamin; REGAN, Mitt. Algorithmic Bias and Risk Assessments: lessons from practice. **Digital Society**, *[S.l.]*, v. 1, n. 2, p. 1-20, 19 ago. 2022. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s44206-022-00017-z. Disponível em: https://link.springer.com/article/10.1007/s44206-022-00017-z. Acesso em: 21 nov. 2023.

HOLANDA. Fundamental Rights and Algorithms Impact Assessment (FRAIA). Disponível em: https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms. Acesso em: 24 dez. 2023.

HORNEBER, David; LAUMER, Sven. Algorithmic Accountability. **Business & Information Systems Engineering**, *[S.l.]*, p. 1-8, 24 maio 2023. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s12599-023-00817-8. Disponível em: https://link.springer.com/article/10.1007/s12599-023-00817-8#citeas. Acesso em: 20 nov. 2023.

INFORMATION COMMISSIONER'S OFFICE. **Guidance on the AI auditing framework**: draft guidance for consultation. Disponível em: https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf. Acesso em: 20 nov. 2023.

JAPÃO. Governance Guidelines for Implementation of AI Principles. Disponível em: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf. Acesso em: 05 nov. 2023.

KAMINSKI, Margot; MALGIERI, Gianclaudio. Algorithmic impact assessments under the GDPR: producing multi-layered explanations. **International Data Privacy Law**, *[S.l.]*, v. 11, n. 2, p. 125-144, 6 dez. 2020. Oxford University Press (OUP). http://dx.doi.org/10.1093/idpl/ipaa020. Disponível em: https://scholar.law.colorado.edu/faculty-articles/1510/. Acesso em: 29 dez. 2023.

KOSHIYAMA, Adriano; ENGIN, Zeynep. Algorithm Impact Assessment: Fairness, Robustness and Explainability in Automated Decision-Making. Data for Polícy, 2019. Disponível em: https://www.academia.edu/download/87169752/DFP_20Presentation_20-_20Zenodo.pdf. Acesso em: 05 nov. 2023

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. São Paulo, SP: Atlas, 2003. Disponível em: < https://docente.ifrn.edu.br/olivianeta/disciplinas/copy_of_historia-i/historia-ii/china-e-india/view>. Acesso em: 24 dez. 2023.

MACHADO, Joana de Souza; NEGRI, Sergio Marcos Carvalho de Ávila; GIOVANINI, Carolina Fiorini Ramos. Nem invisíveis, nem visados: inovação, direitos humanos e vulnerabilidade de grupos no contexto da Covid-19. **Liinc em Revista**, *[S. l.]*, v. 16, n. 2, p. e5367, 2020. Disponível em: https://revista.ibict.br/liinc/article/view/5367. Acesso em: 07 set. 2023.

MANTELERO, Alessandro. Beyond Data: human rights, ethical and social impact assessment in ai. **Information Technology And Law Series**, *[S.l.]*, v. 36, p. 1-215, 2022. T.M.C. Asser Press. http://dx.doi.org/10.1007/978-94-6265-531-7. Disponível em: https://link.springer.com/book/10.1007/978-94-6265-531-7. Acesso em: 05 nov. 2023.

MANTELERO, Alessandro. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: roots, legal obligations and key elements for a model template. **SSRN**, 2024. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4782126. Acesso em: 20. Abrl. 2024.

MCCARTHY, J; MINSKY, M. L; ROCHESTER, N; SHANNON, C. E. **A proposal for the Dartmouth summer research project on Artificial Intelligence**. [S.l.]: Dartmouth, 1956. Disponível em: http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. Acesso em: 25 jan. 2023.

METCALF, Jacob; SINGH, Ranjit; MOSS, Emanuel; TAFESSE, Emnet; WATKINS, Elizabeth Anne. Taking Algorithms to Courts: a relational approach to algorithmic

accountability. **2023 ACM Conference On Fairness, Accountability, And Transparency**, *[S.l.]*, v. 1, n. 1, p. 1450-1462, 12 jun. 2023. ACM. http://dx.doi.org/10.1145/3593013.3594092. Disponível em: https://dl.acm.org/doi/10.1145/3593013.3594092. Acesso em: 29 dez. 2023.

MILARÉ, Edis. **Direito do Ambiente**: doutrina, prática, jurisprudência, glossário. São Paulo: Revista dos Tribunais, 2000.

MÖKANDER, Jakob; FLORIDI, Luciano. Ethics-Based Auditing to Develop Trustworthy AI. **Minds And Machines**, *[S.l.]*, v. 31, n. 2, p. 323-327, 19 fev. 2021. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11023-021-09557-8. Disponível em: https://link.springer.com/article/10.1007/s11023-021-09557-8#citeas. Acesso em: 21 nov. 2023.

NAS, Elen. Como e por que decolonizar a inteligência artificial? **Jornal da USP**, 2023. Disponível em: https://jornal.usp.br/artigos/como-e-por-que-decolonizar-a-inteligencia-artificial/. Acesso em: 03 fev. 2024.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. AI Risk Management Framework, 2023. Disponível em: https://www.nist.gov/itl/ai-risk-management-framework. Acesso em: 05 nov. 2023.

NEGRI, Sergio Marcos Carvalho de Ávila. As razões da pessoa jurídica e a expropriação da subjetividade. **Civilística**, v. 5, n. 2, p. 1-18, 2016, p. 2. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/265. Acesso em: 26 dez. 2023.

NEGRI, Sergio Marcos Carvalho de Ávila. Robôs como pessoas: a personalidade eletrônica na Robótica e na inteligência artificial. **Pensar Revista de Ciências Jurídicas**. Fortaleza, 2020, p.1-14). Disponível em: https://periodicos.unifor.br/rpen/article/view/10178/pdf. Acesso em: 26 dez. 2023.

NEGRI, Sergio M. C. Ávila; MACHADO, Joana de Souza; GIOVANININ, Carolina Fiorini Ramos; BATISTA, Nathan Pascoalini Ribeiro. Sistemas de inteligência artificial e avaliações de impacto para direitos humanos. **Revista Culturas Jurídicas**, V. 10, n. 26, p. 153-181, 2024. Disponível em: https://periodicos.uff.br/culturasjuridicas/index. Acesso em: 05 mai. 2024.

O'NEIL, Cathy. **Algoritmos de Destruição em Massa**: como o big data aumenta a desigualdade e ameaça à democracia. Santo André: Editora Rua do Sabão, 2021. 342 p.

OCDE. Recommendation of the Council on Artificial Intelligence, 2019 (Amended on: 07/11/2023). Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 25 jun. 2023.

OTTERLO, Van. A machine learning view on profiling. *In:* HILDEBRANDT, M.; DE VRIES, K. (Eds.) **Privacy, due process and the computational turn**: philosophers of law meet philosophers of technology. Abingdon: Routledge, 2013 p. 41-64

OSWALD, Marion. Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism. **International Review of Law, Computers &**

Technology, *[S.l.]*, v. 34, n. 2, p. 214-231, 4 mar. 2020. Disponível em: https://www.tandfonline.com/doi/full/10.1080/13600869.2020.1733758. Acesso em: 26 dez. 2023.

PASQUALE, Frank. **The black box society**: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

POWERS, Thomas M.; GANASCIA, Jean-Gabriel. The Ethics of the Ethics of AI. *In*: DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (ed.). **The Oxford Handbook of Ethics of AI**. Nova Iorque: Oxford University Press, 2020.

REINO UNIDO. **AI regulation**: a pro-innovation approach. *[S.l.]*, 2023. Disponível em: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach. Acesso em: 06 jan. 2023.

RODOTÀ, Stefano. **A vida na sociedade da vigilância**: a privacidade hoje. Organização, seleção e apresentação de Maria Celina Bodin de Moraes. Tradução de Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

RODOTÀ, Stefano. Così l'umano può difendersi dal postumano. **MicroMega**: 28 de abril de 2015. Disponível em: http://temi.repubblica.it/micromega-online/cosi-l%E2%80%99umano-puo-difendersi-dal-postumano. Acesso em: 26 dez. 2023

RODOTÀ, Stefano. L'uso umano degli esseri umani. **MicroMega**: agosto de 2015. Disponível em: http://temi.repubblica.it/micromega-online/addio-a-stefano-rodota-una-vita-per-la-costituzione-la-laicita-e-i-diritti/?printpage=undefined. Acesso em: 26 dez. 2023

RUIZ, J. A. Metodologia Científica: guia para eficiência nos estudos. São Paulo, SP: Atlas, 2009. Disponível em <http://gestaouniversitaria.com.br/artigos/consideracoes-sobre-estado-da-arte-levantamento-bibliografico-e-pesquisa-bibliografica-relacoes-e-limites>. Acesso em: 24 dez. 2023.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.

RUSSEL, Stuart. Updates to the OECD's definition of an AI system explained. OEDC.AI Policy Observatory, 2023. Disponível em: https://oecd.ai/en/wonk/ai-system-definition-update. Acesso em: 30 nov. 2023.

SÁ-SILVA, Jackson Ronie; ALMEIDA, Cristovão Domingos.; GUINDANI, Joel Felipe Pesquisa documental: pistas teóricas e metodológicas. **Revista Brasileira de História e Ciências Sociais**, São Leopoldo, ano 1, n. 1, jul. 2009. Disponível em: <https://www.periodicos.furg.br/rbhcs/article/view/10351>. Acesso em: 24 dez. 2023.

SCHUETT, Jonas. Defining the scope of AI regulations. **Law, Innovation And Technology**, [S.L.], v. 15, n. 1, p. 60-82, 2 jan. 2023. Informa UK Limited. http://dx.doi.org/10.1080/17579961.2023.2184135. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3453632. Acesso em: 05 mar. 2024.

SEARLE, John. **Mente, Cérebro e Ciência**. Tradução de Artur Morão. Lisboa: Biblioteca de Filosofia Contemporânea, 2000.

SELBST, Andrew D. An Institutional View of Algorithmic Impact Assessments. **Harvard Journal of Law & Technology**, *[S.l.]*, v. 35, n. 1, p. 117-191, jun. 2021. Disponível em: https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf. Acesso em: 29 dez. 2023.

SEVERINO, Antônio Joaquim. **Metodologia do Trabalho Científico**. São Paulo, SP: Cortez, 2007. Disponível em: < https://edisciplinas.usp.br/pluginfile.php/3480016/mod_label/intro/SEVERINO_Metodologia_do_Trabalho_Cientifico_2007.pdf>. Acesso em: 24 dez. 2023.

SILVA, Tarcízio. **Racismo algorítmico**: inteligência artificial e discriminação nas redes digitais. *[S.l.]*: Democracia Digital, 2022. 223 p.

SINGAPURA. **National AI Strategy 2.0**: AI for the public good for singapore and the world. Singapura, 2023. Disponível em: https://file.go.gov.sg/nais2023.pdf. Acesso em: 06 jan. 2023.

SMITH, Andrew. Using Artificial Intelligence and Algorithms. 2020. Disponível em: https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms. Acesso em: 03 abr. 2023.

SOUZA, Carlos Affonso. Como a 'guerra cultural' fez o Google proibir IA de criar imagem de pessoas. Tilt UOL, 2024. Disponível em: https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2024/02/27/como-a-guerra-cultural-fez-o-google-proibir-ia-de-criar-imagem-de-pessoas.htm. Acesso em: 23 out. 2024.

SOUZA, Carlos Affonso. Inteligência artificial 'consciente' do Google: avanço real ou erro humano?. Tilt UOL, 2022. Disponível em: https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2022/06/13/inteligencia-artificial-consciente-do-google-avanco-real-ou-erro-humano.htm. Acesso em: 14 jan. 2024.

STEIBEL, F; VICENTE, V. F; JESUS, D. S. V. Possibilidades e potenciais da utilização da Inteligência Artificial *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. (Org.). **Inteligência Artificial e Direito**. São Paulo: Revista dos Tribunais, 2019, p. 55.

TEIXEIRA, João de Fernandes. **O cérebro e o robô**: inteligência artificial, biotecnologia e a nova ética. São Paulo: Paulus, 2015.

UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em: 05 nov. 2023.

UNIÃO EUROPEIA. Diretiva 2014/24/UE do Parlamento Europeu e do Conselho, de 26 de fevereiro de 2014, relativa aos contratos públicos e que revoga a Diretiva 2004/18/CE. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32014L0024. Acesso em: 15 jul. 2024.

UNIÃO EUROPEIA. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL)). Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html?redirect. Acesso em: 15 mai. 2023.

UNIÃO EUROPEIA. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Disponível em: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Acesso em: 15 mai. 2023.

UNIÃO EUROPEIA. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Act. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. Acesso em: 05 nov. 2023.

UNIÃO EUROPEIA. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 13 jul. 2024.

UNIÃO EUROPEIA. Directive (EU) 2024/1760 of the European Parliament and of the Council of 13 June 2024 on corporate sustainability due diligence and amending Directive (EU) 2019/1937 and Regulation (EU) 2023/2859. Disponível em: https://eur-lex.europa.eu/eli/dir/2024/1760/oj. Acesso em: 20 jul. 2024.

UNITED KINGDOM. Algorithmic Transparency Reports. Disponível em: https://www.gov.uk/government/collections/algorithmic-transparency-reports. Acesso em: 03 abr. 2023.

UNITED STATES CONGRESS. S.3771 – FUTURE of Artificial Intelligence Act of 2020. Disponível em: https://www.congress.gov/bill/116th-congress/senate-bill/3771/text. Acesso em: 05 nov. 2023.

U.S. DEPARTMENT OD DEFENSE. DOD Adopts Ethical Principles for Artificial Intelligence. Disponível em: https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/. Acesso em: 03 abr. 2023.

U.S NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH RESOURCE TASK FORCE. Strengthening andDemocratizing the U.S. Artificial Intelligence Innovation Ecosystem. Disponível em: https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf. Acesso em: 05 nov. 2023.

WATKINS, Elizabeth Anne; MOSS, Emanuel; METCALF, Jacob; SINGH, Ranjit; ELISH, Madeleine Clare. Governing Algorithmic Systems with Impact Assessments: six observations. **Proceedings Of The 2021 AAAI ACM Conference On AI, Ethics, And**

Society, *[S.l.]*, v. 1, n. 1, p. 1010-1021, 21 jul. 2021. ACM. http://dx.doi.org/10.1145/3461702.3462580. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3461702.3462580. Acesso em: 29 dez. 2023.

WHITE HOUSE. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. Disponível em: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. Acesso em: 05 nov. 2023.

WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY. Blueprint for an AI Bill of Rights. Disponível em: https://www.whitehouse.gov/ostp/ai-bill-of-rights/. Acesso em: 05 nov. 2023.