



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
FACULDADE NACIONAL DE DIREITO  
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO  
MESTRADO EM DIREITO

ELOISA SAMY SANTIAGO

**AVALIAÇÃO DE IMPACTO ALGORÍTMICO:  
Desafios Éticos e Jurídicos na Identificação e Mitigação do Viés de Gênero**

RIO DE JANEIRO  
2025

Eloisa Samy Santiago

AVALIAÇÃO DE IMPACTO ALGORÍTMICO:  
Desafios Éticos e Jurídicos na Identificação e Mitigação do Viés de Gênero

Número de volumes: 1

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Direito da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Direito (Teorias Jurídicas Contemporâneas).

Orientadora: Daniela Silva Fontoura de Barcellos

Rio de Janeiro  
2025

## CIP - Catalogação na Publicação

S48a Santiago, Eloisa Samy  
Avaliação de Impacto Algorítmico: Desafios Éticos e Jurídicos na Identificação e Mitigação do Viés de Gênero / Eloisa Samy Santiago. -- Rio de Janeiro, 2025.  
171 f.

Orientadora: Daniela Silva Fontoura de Barcellos.  
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Faculdade Nacional de Direito, Programa de Pós-Graduação em Direito, 2025.

1. Direitos Humanos e Tecnologia. 2. Vulnerabilidade Existencial. 3. Viés de Gênero Algorítmico. 4. Avaliação de Impacto Algorítmico. 5. Ética Algorítmica. I. Barcellos, Daniela Silva Fontoura de, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

ELOISA SAMY SANTIAGO

AVALIAÇÃO DE IMPACTO ALGORÍTMICO:  
Desafios Éticos e Jurídicos na Identificação e Mitigação do Viés de Gênero

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Direito da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Direito (Teorias Jurídicas Contemporâneas).

Aprovada em 27 de março de 2025

Orientadores:

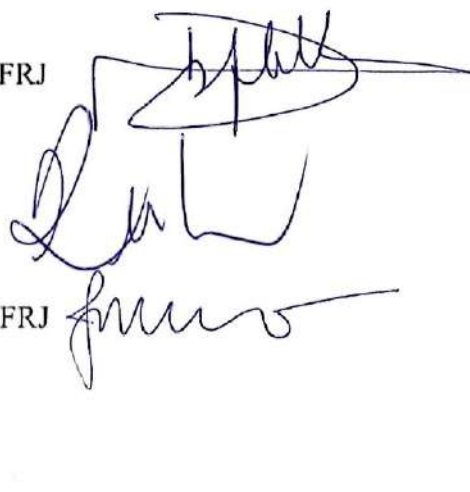
1) Daniela Silva Fontoura de Barcellos - PPGD/UFRJ

Banca examinadora (Nome Completo)

1) Clóvis Ricardo Montenegro de Lima - IBICT

2) Juliana Neuenschwander Magalhães - PPGD/UFRJ

3) Rafael Viola - UERJ



Rio de Janeiro  
2025

## **DEDICATÓRIA**

À memória de minha mãe, que sempre foi minha inspiração, e a todos aqueles que acreditam e confiam que somos capazes de construir uma sociedade mais justa e inclusiva.

## AGRADECIMENTOS

Aos meus demônios, que eu soube adestrar;

Aos meus fantasmas, que eu soube expurgar;

Aos meus antepassados, cujas histórias se reúnem em mim;

À minha psicanalista, que me ajudou a conviver com as minhas sombras;

Aos mestres do Programa de Pós-Graduação em Direito da UFRJ por todo o conhecimento jurídico e interdisciplinar que me passaram, meus mais profundos agradecimentos.

*Não acredito que existam qualidades, valores, modos de vida especificamente femininos: seria admitir a existência de uma natureza feminina, quer dizer, aderir a um mito inventado pelos homens para prender as mulheres na sua condição de oprimidas. Não se trata para a mulher de se afirmar como mulher, mas de tornarem-se seres humanos na sua integridade.*

Simone de Beauvoir

## LISTA DE ABREVIATURAS E SIGLAS

AIA	Avaliação de Impacto Algorítmico
AIR	Avaliação de Impacto Regulatório
ADM	<i>Automated Decision-Making</i> (Modelos de Decisão Automatizados)
ANN	<i>Artificial Neural Networks</i> (Redes Neurais Artificiais)
ANPD	Autoridade Nacional de Proteção de Dados
CEDAW	Convenção sobre a Eliminação de Todas as Formas de Discriminação Contra a Mulher
CNN	<i>Convolutional Neural Networks</i> (Redes Neurais Convolucionais)
CRFB/88	Constituição da República Federativa do Brasil de 1988
EBIA	Estratégia Brasileira de Inteligência Artificial
EC	Emenda Constitucional
GAN	<i>Generative Adversarial Networks</i> (Redes Adversárias Generativas)
GDPR	<i>General Data Protection Regulation</i> (Regulamento Geral sobre a Proteção de Dados da União Europeia)
GRC	Governança, Riscos e Compliance
HIC	<i>Human-in-Command</i> (Humano no Comando)
HITL	<i>Human-in-the-Loop</i> (Humano no Ciclo)
HOTL	<i>Human-on-the-Loop</i> (Humano no Ciclo, porém não na Malha)
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados
MCTIC	Ministério da Ciência, Tecnologia, Inovações e Comunicações
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ONU	Organização das Nações Unidas
PL	Projeto de Lei
PRO-REG	Programa de Fortalecimento da Capacidade Institucional para Gestão em Regulação
RIPD	Relatório de Impacto sobre a Proteção de Dados Pessoais
SHAP	<i>SHapley Additive exPlanations</i> (Explicações Aditivas de Shapley)
SNN	<i>Simulated Neural Networks</i> (Redes Neurais Simuladas)
STF	Supremo Tribunal Federal

UE	União Europeia
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
WebTAP	<i>Web-Based Thematic Analysis Platform</i> (Plataforma de Análise Temática Baseada na Web)

## RESUMO

Esta dissertação investiga os impactos sociais, éticos e jurídicos da IA, com ênfase no viés de gênero, propondo diretrizes para a avaliação de impacto algorítmico e mitigação de riscos ao desenvolvimento tecnológico responsável. Preconceitos e discriminações presentes nas decisões humanas, muitas vezes de forma inconsciente, também se refletem em softwares e sistemas de Inteligência Artificial, contrariando a expectativa inicial de um ambiente digital imparcial e objetivo. A constatação de que vieses algorítmicos podem reforçar desigualdades e ameaçar direitos humanos torna essencial a implementação de regulamentações legais, tecnológicas e éticas para mitigar esses problemas. Considerando que a IA opera por meio de dados e aprendizado automático, a supervisão criteriosa das etapas de coleta, análise e interpretação é indispensável para evitar discriminações baseadas em gênero, etnia ou classe social. A pesquisa adota metodologia qualitativa, sendo descritiva e explicativa quanto aos objetivos, com abordagem crítica e interdisciplinar que combina análise de teorias éticas e jurídicas com revisão da literatura especializada. Para fundamentar as reflexões, utilizam-se livros, artigos científicos, legislação aplicável e documentos oficiais. No primeiro capítulo, discute-se o conceito de gênero sob a ótica das teorias feministas, analisando como normas sociais moldam representações simbólicas e perpetuam estereótipos discriminatórios, além de explorar a relação entre gênero, poder e desigualdade na sociedade contemporânea. O segundo capítulo examina os processos de elaboração e treinamento de sistemas de IA, destacando heurísticas intuitivas e seus impactos na tomada de decisões automatizadas, com apresentação de exemplos de vieses recorrentes e suas consequências. O terceiro capítulo analisa os desafios éticos no desenvolvimento de sistemas de IA, visando prevenir violações à dignidade humana e reforçar o Estado Democrático de Direito num contexto de regulação internacional, em que se discute estratégias para mitigar riscos éticos e jurídicos associados a algoritmos de IA, com ênfase na governança algorítmica e avaliações de impacto que integrem uma análise crítica de gênero. Propõe-se práticas como auditorias independentes, testes de robustez e diversificação de bases de dados para minimizar vieses discriminatórios, garantindo que a tecnologia promova equidade.

**PALAVRAS-CHAVE:** Direitos Humanos e Tecnologia; Vulnerabilidade Existencial; Viés de Gênero Algorítmico; Avaliação de Impacto Algorítmico; Ética Algorítmica.

## ABSTRACT

This dissertation investigates the social, ethical, and legal impacts of AI, with an emphasis on gender bias, proposing guidelines for algorithmic impact assessment and risk mitigation to ensure responsible technological development. Biases and discriminations present in human decision-making, often unconsciously, are also reflected in software and Artificial Intelligence systems, contradicting the initial expectation of an impartial and objective digital environment. The recognition that algorithmic biases can reinforce inequalities and threaten human rights makes it essential to implement legal, technological, and ethical regulations to mitigate these issues. Considering that AI operates through data and machine learning, rigorous oversight of the stages of data collection, analysis, and interpretation is indispensable to prevent discrimination based on gender, ethnicity, or social class. The research adopts a qualitative methodology, being descriptive and explanatory in its objectives, with a critical and interdisciplinary approach that combines the analysis of ethical and legal theories with a review of specialized literature. To support the reflections, books, scientific articles, applicable legislation, and official documents are used. In the first chapter, the concept of gender is discussed from the perspective of feminist theories, analyzing how social norms shape symbolic representations and perpetuate discriminatory stereotypes, in addition to exploring the relationship between gender, power, and inequality in contemporary society. The second chapter examines the processes of developing and training AI systems, highlighting intuitive heuristics and their impacts on automated decision-making, with examples of recurring biases and their consequences. The third chapter analyzes the ethical challenges in AI system development, aiming to prevent violations of human dignity and strengthen the Democratic Rule of Law within the context of international regulation. Strategies to mitigate ethical and legal risks associated with AI algorithms are discussed, with an emphasis on algorithmic governance and impact assessments that incorporate a critical gender analysis. Proposed practices include independent audits, robustness testing, and database diversification to minimize discriminatory biases, ensuring that technology promotes equity.

**KEYWORDS:** Human Rights and Technology; Existential Vulnerability; Algorithmic Gender Bias; Algorithmic Impact Assessment; Algorithmic Ethics.

## SUMÁRIO

<b>INTRODUÇÃO.....</b>	<b>13</b>
<b>CAPÍTULO 1 – NORMAS DE GÊNERO E SUA IMPORTÂNCIA NA AVALIAÇÃO DE IMPACTO ALGORÍTMICO.....</b>	<b>18</b>
1.1 CONCEITO DE GÊNERO.....	21
1.2 IDENTIDADE E PERSONALIDADE.....	34
1.3 VULNERABILIDADE EXISTENCIAL.....	36
1.4 MANIPULAÇÃO DIGITAL E VULNERABILIDADE HUMANA.....	39
<b>CAPÍTULO 2 – PROCESSOS DE ELABORAÇÃO, DESENVOLVIMENTO E TREINAMENTO DE IA.....</b>	<b>47</b>
2.1 VIESES ALGORÍTMICOS E HEURÍSTICAS INTUITIVAS.....	60
2.1.1 Heurística da Disponibilidade.....	64
2.1.2 Heurística da Referência ou da Ancoragem.....	66
2.1.3 Heurística do Afeto.....	67
2.1.4 Heurística da Correlação Ilusória.....	68
2.1.5 Viés Egocêntrico e Excesso de Confiança.....	68
2.1.6 Viés de Confirmação.....	69
2.1.7 Viés Retrospectivo.....	70
2.1.8 Erro Fundamental de Atribuição ou Viés de Correspondência.....	71
2.1.9 Viés do Ator-Observador.....	74
2.1.10 Viés da Perspectiva de Câmera.....	75
<b>CAPÍTULO 3 – GOVERNANÇA ALGORÍTMICA.....</b>	<b>81</b>
3.1 NORMATIVIDADE ALGORÍTMICA CONCORRENTE.....	86
3.2 DIREITOS FUNDAMENTAIS E O RISCO TECNOLÓGICO.....	99
3.3 INTELIGÊNCIA ARTIFICIAL NO BRASIL: ESTRATÉGIAS DE GOVERNANÇA, ÉTICA E DESENVOLVIMENTO SUSTENTÁVEL.....	107
3.4 PRÁTICAS DE GOVERNANÇA E GESTÃO DE RISCOS NA IMPLEMENTAÇÃO DE ALGORITMOS.....	115
3.4.1 Princípios Fundamentais da Proteção no Cenário Jurídico Global.....	117
3.4.2 Princípios Fundamentais da Proteção no Cenário Jurídico Nacional.....	123
3.4.3 Princípios e Limitações na Regulação da IA no Brasil.....	128
3.4.4 Governança Pública, Ética e Inteligência Artificial.....	134
3.5. DA AVALIAÇÃO DE IMPACTO À MITIGAÇÃO DE RISCOS.....	144
<b>CONSIDERAÇÕES FINAIS.....</b>	<b>151</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>158</b>

## INTRODUÇÃO

Esta dissertação investiga os impactos sociais, éticos e jurídicos da Inteligência Artificial (IA), com especial atenção ao viés de gênero, propondo diretrizes para a avaliação de impacto algorítmico e a mitigação de riscos associados ao desenvolvimento tecnológico responsável. Embora a IA seja frequentemente concebida como um sistema neutro e objetivo, desvinculado das limitações e preconceitos humanos, sua estruturação, treinamento e aplicação demonstram que algoritmos não são entidades autônomas imunes a discriminações. Pelo contrário, preconceitos e assimetrias presentes nas decisões humanas, muitas vezes de forma inconsciente, acabam reproduzidos em softwares e sistemas inteligentes, comprometendo a equidade e a justiça social.

Esse fenômeno contraria a ideia de que o ambiente digital poderia constituir um espaço isento de vieses, revelando que a tecnologia não apenas reflete desigualdades preexistentes, mas também pode amplificá-las e consolidá-las de maneira sistêmica.

A constatação de que vieses algorítmicos podem reforçar dinâmicas discriminatórias e ameaçar direitos humanos exige uma resposta estruturada e multidimensional, que envolva tanto o aprimoramento das arquiteturas tecnológicas quanto a formulação de marcos regulatórios e protocolos éticos rigorosos. Considerando que os sistemas de IA dependem da qualidade e da representatividade dos dados utilizados para seu treinamento, a supervisão criteriosa das etapas de coleta, processamento, análise e interpretação torna-se fundamental para evitar discriminações relacionadas a gênero, etnia ou classe social. Sem mecanismos robustos de governança e transparência, a automação de processos decisórios pode perpetuar desigualdades históricas, legitimadas sob a falsa premissa de imparcialidade algorítmica.

Dessa forma, compreender e enfrentar os riscos decorrentes da incorporação de vieses nos sistemas inteligentes não é apenas uma necessidade técnica, mas um imperativo ético e jurídico para garantir que o avanço da IA se alinhe aos princípios de justiça, inclusão e respeito à dignidade humana.

Do ponto de vista jurídico, a formulação de normas específicas deve assegurar transparência, responsabilidade e equidade no desenvolvimento e na aplicação de modelos preditivos e sistemas automatizados de tomada de decisão.

No âmbito tecnológico, a implementação de metodologias de auditoria, testes de viés e estratégias de mitigação devem ser incorporadas desde as fases iniciais de projeto e treinamento

dos algoritmos, garantindo que os dados utilizados sejam representativos e livres de distorções que perpetuem desigualdades. Ademais, no plano ético, a discussão sobre os impactos sociais da IA deve envolver diversos atores, incluindo pesquisadores, formuladores de políticas públicas, empresas de tecnologia e organizações da sociedade civil, a fim de estabelecer diretrizes que assegurem a proteção dos direitos fundamentais. Assim, enfrentar a discriminação no ambiente digital não se limita a um desafio técnico, mas requer um esforço coordenado e interdisciplinar para que a tecnologia atue como um instrumento de inclusão e justiça, e não como um vetor de exclusão e reforço de assimetrias sociais.

Dado que a Inteligência Artificial opera por meio da padronização das ações e do aprendizado automático a partir de dados, é fundamental supervisionar as fases em que tais vieses surgem, pois podem ser posteriormente reproduzidos, resultando em decisões discriminatórias para indivíduos ou grupos com base em características como raça, religião, etnia, sexo ou classe social. O desafio, tanto no âmbito tecnológico quanto jurídico, é identificar esses vieses e estabelecer os procedimentos necessários para minimizá-los ou neutralizá-los.

Na esfera coletiva da sociedade, a imagem da mulher é influenciada por diversos fatores que a colocam em posição de desvantagem social, levando-a a adotar determinados papéis e comportamentos que refletem sua subordinação à autoridade masculina. Esses fatores, conhecidos como estereótipos de gênero, destacam-se como um ponto crítico na análise, pois evidenciam a importância das influências socioculturais na formação da identidade individual e nas interações sociais.

No presente estudo, o conceito de gênero será explorado à luz das teorias feministas, que servirão como base analítica. Procura-se demonstrar como as representações simbólicas da mulher nos meios de comunicação se conectam às normas sociais que regem gênero, poder e desigualdade, investigando a construção de percepções pré-reflexivas sobre o gênero e a forma como essas percepções, profundamente enraizadas, fomentam preconceitos e práticas discriminatórias contra as mulheres. Embora o progresso e a aplicação das tecnologias de Inteligência Artificial sejam indubitavelmente benéficos, é imprescindível examinar os impactos negativos que podem surgir, especialmente no que diz respeito à proteção e ao exercício dos direitos fundamentais.

A Constituição Federal de 1988, em seu artigo 5º, consagra o princípio da igualdade ao estabelecer que “todos são iguais perante a lei”, com o inciso I reforçando a igualdade entre

homens e mulheres como um dos pilares fundamentais da República Federativa do Brasil, orientada para o bem-estar de todos, sem distinções de sexo ou outras formas de discriminação. E, por tal razão, justifica-se o interesse nessa pesquisa. Primeiramente, por havermos chegado em um momento histórico de obscurantismo que ameaça a cultura dos direitos humanos e contribui para a manutenção da vulnerabilidade social de muitos segmentos sociais, em especial das mulheres. A isso, soma-se a carência de um referencial teórico adequado para a condução dos debates sobre os problemas enfrentados pelas mulheres advindos, principalmente, da negação aos seus direitos reprodutivos e sexuais, impedindo o alcance de soluções adequadas. Por tais motivos, busca-se promover a equidade entre homens e mulheres por meio da prevenção do viés de gênero, incentivando uma reflexão crítica sobre os impactos sociais, culturais e éticos da Inteligência Artificial na vida das mulheres.

Nos últimos anos, esforços provenientes de esferas governamentais, legislativas e acadêmicas têm se concentrado na formulação de diretrizes que, além de fundamentais, sejam efetivamente aplicáveis na abordagem desse desafio. Dessa forma, consolida-se um consenso acerca da importância de uma governança voltada ao desenvolvimento e ao uso responsável da Inteligência Artificial, com especial atenção à avaliação de riscos e aos impactos que cada aplicação pode gerar sobre os direitos e as liberdades individuais, analisando-se as ações e os elementos subjacentes que contribuem para os preconceitos algorítmicos, compreendidos como uma forma de desalinhamento. Examina-se, ainda, a tipologia desses preconceitos e sua manifestação em contextos específicos. Destaca-se a importância de investigar as razões pelas quais esses vieses, enraizados em valores, crenças, normas e culturas, também estão presentes nos algoritmos utilizados na Inteligência Artificial.

No âmbito da estratégia digital da UE, o conceito de risco abrange uma ampla gama de impactos negativos, incluindo, mas não se restringindo a: (a) violações de direitos fundamentais, como igualdade, não discriminação, privacidade e liberdade de expressão; (b) prejuízos à saúde ou segurança dos indivíduos, que podem envolver desde lesões físicas até perda de vida; (c) danos psicológicos, como perda de autoestima ou de autonomia pessoal; (d) impactos sociais ou econômicos, como perdas financeiras, restrições no acesso a serviços públicos ou privados e perda de propriedades; (e) comprometimentos à reputação ou estigmatização; (f) práticas de discriminação injusta, como diferenças no preço, no acesso a empregos, renda e serviços; (g) perda de controle individual, incluindo manipulações econômicas ou psicológicas.

Para mitigar esses riscos propôs-se a criação do Relatório de Impacto de Inteligência Artificial, documento destinado a avaliar os efeitos sobre a proteção de dados pessoais, com especial atenção a possíveis discriminações resultantes do processamento algorítmico.

O ordenamento jurídico brasileiro, todavia, ainda carece de uma obrigatoriedade legal para a realização da Avaliação de Impacto Algorítmico, embora sua condução encontre justificativa em diversos aspectos, com subsídios em considerações econômicas, regulatórias (incluindo autorregulação) e a proteção dos direitos fundamentais. No entanto, embora não haja exigência legal específica, a função social da empresa, consagrada no artigo 170, inciso II, da Constituição Federal de 1988, reforça o princípio do bem-estar coletivo. Nesse sentido, negligenciar aspectos como supervisão humana, prevenção de danos, justiça, não discriminação, responsabilidade social e proteção de grupos vulneráveis pode configurar violação do ordenamento jurídico, resultando na responsabilização do agente infrator. Assim, algoritmos que perpetuem discriminações são considerados não apenas antiéticos, mas também ilegais, acarretando riscos financeiros e de reputação para as organizações.

Considerando que o estudo do direito deve, precipuamente, voltar-se para a resolução de problemas concretos, nosso foco se deterá especialmente na análise de conteúdo relacionada à representação simbólica da mulher e ao viés de gênero na comunicação, em sua confrontação com a normativa doméstica e internacional de proteção à mulher, bem como as regras e princípios que norteiam a efetivação dos direitos fundamentais, relativamente ao direito à identidade.

Para compreender melhor este fenômeno, examinaremos as maneiras pelas quais a mulher é simbolicamente representada nas relações sociais, incluindo estereótipos de gênero, papéis tradicionais e traços de personalidade atribuídos a elas, em que buscamos elucidar como nossas percepções automáticas e não conscientes podem moldar preconceitos e comportamentos discriminatórios em relação às mulheres, contribuindo para a persistência da desigualdade de gênero. Desse modo, a pesquisa tem como objetivo formular diretrizes para Avaliação de Impacto Algorítmico, visando garantir que sistemas inteligentes sejam empregados de maneira ética e alinhada aos princípios de igualdade e não discriminação.

Procura-se, assim, contribuir para a promoção da ordem e do bem-estar social, prevenindo o agravamento das desigualdades por meio do uso responsável da tecnologia.

Como resultados, espera-se consolidar um referencial teórico e prático para o desenvolvimento e a governança da IA, alinhado a iniciativas internacionais, como as diretrizes

da União Europeia, e fundamentado nos princípios constitucionais brasileiros, especialmente no que tange à função social da empresa e à proteção dos direitos fundamentais.

O estudo não pretende aprofundar o entendimento das diversas técnicas de aprendizado de máquina, mas sim investigar os processos de criação e modelagem de dados nos quais vieses prejudiciais podem ser replicados.

Utiliza-se o termo Inteligência Artificial de forma abrangente, incluindo qualquer produto ou serviço que incorpore componentes algorítmicos capazes de imitar habilidades originalmente humanas.

## **CAPÍTULO 1 – NORMAS DE GÊNERO E SUA IMPORTÂNCIA NA AVALIAÇÃO DE IMPACTO ALGORÍTMICO**

A luta contra a discriminação tem sido um dos pilares dos direitos humanos, refletindo o compromisso com a construção de sociedades mais justas e igualitárias. Contudo, com o avanço das tecnologias digitais e a crescente adoção de sistemas de Inteligência Artificial em diversas esferas da vida social, econômica e política, torna-se evidente que a discriminação não se limita às interações exclusivamente humanas, mas também se manifesta por meio de algoritmos e processos automatizados. A incorporação de vieses nesses sistemas pode agravar desigualdades estruturais, afetando desproporcionalmente grupos historicamente marginalizados, como mulheres, populações racializadas e indivíduos em situação de vulnerabilidade socioeconômica. Nesse contexto, a implementação de regulamentações apropriadas torna-se urgente, abrangendo os aspectos legais, tecnológicos e éticos da governança algorítmica. A análise das normas de gênero, enquanto estruturas reguladoras das relações sociais, é fundamental para compreender como essas normas impactam a construção e a reprodução de desigualdades no ambiente digital. Assim, este capítulo se estrutura a partir de quatro eixos principais: a definição do conceito de gênero e sua relevância na análise da Inteligência Artificial; a relação entre identidade, personalidade e discriminação algorítmica; os impactos da vulnerabilidade existencial na perpetuação de desigualdades; e, por fim, o papel da manipulação digital na amplificação da vulnerabilidade humana.

Os desafios éticos e jurídicos na identificação e mitigação do viés de gênero envolvem a necessidade de garantir que os sistemas de Inteligência Artificial sejam desenvolvidos e utilizados de maneira justa, sem reforçar desigualdades de gênero, dentro de um marco legal que, no Brasil, ainda carece de regulamentações específicas.

Apesar da expectativa de igualdade, as normas constitucionais não fornecem uma definição clara desse princípio, e o tema da discriminação contra a mulher é raramente abordado de forma direta pelos teóricos do direito. Isso torna difícil, senão impossível, transformar nossa sociedade, dado o entendimento limitado sobre o funcionamento dos sistemas de marginalização social.

Ao analisar a representação feminina na esfera pública, é fundamental abordar a função ideacional, que revela como padrões de experiência, realidade e processos internos são representados, refletindo concepções sobre o mundo.

Nesse contexto, os homens, enquanto integrantes do grupo social dominante, possuem o poder de definir os significados culturais atribuídos às mulheres, perpetuando relações de poder que reforçam sua sujeição e subordinação dentro da ordem patriarcal.

PEREZ<sup>1</sup>, em seu extenso levantamento histórico sobre a invisibilidade feminina, constata que a prevalência da técnica de Inteligência Artificial (IA) baseada em dados nas aplicações atuais resulta em decisões enviesadas por gênero, muitas vezes de forma inadvertida. Um exemplo ilustrativo é o diagnóstico equivocado após um ataque cardíaco: na Inglaterra, as mulheres têm 50% mais probabilidade de receber um diagnóstico incorreto, consequência da predominância masculina nos estudos científicos relacionados à insuficiência cardíaca. A prática de não coletar dados desagregados por sexo, considerando os homens como o “padrão humano”, distorce a suposta objetividade e precisão dos resultados dos modelos estatísticos baseados em IA.

A conexão entre a manutenção das normas patriarcais na sociedade, seus critérios regulatórios e os sistemas de Inteligência Artificial pode ser explorada em um contexto de análise crítica que inclui considerações acerca da perpetuação do viés de gênero. Isso porque os sistemas de IA são frequentemente treinados com conjuntos de dados que refletem e perpetuam vieses e desigualdades existentes na sociedade, incluindo normas patriarcais, o que resulta em algoritmos que reproduzem e amplificam preconceitos de gênero, como a sub-representação das mulheres em certos campos profissionais, a perpetuação de estereótipos de gênero em recomendações de produtos ou através de algoritmos de recrutamento de emprego que aprendem a preferir candidatos com características historicamente associadas ao masculino, o que pode acarretar a perda de oportunidades no mercado de trabalho.

Em 2018, a Amazon identificou que a Inteligência Artificial utilizada para realizar a triagem inicial de candidatos a vagas de emprego estava enviesada contra mulheres. O algoritmo foi treinado com dados que resultaram em uma preferência por candidatos do sexo masculino e em uma avaliação negativa de currículos contendo expressões associadas a mulheres.

Diante desse viés presente na tecnologia, a Amazon decidiu interromper seu uso nos processos seletivos<sup>2</sup>.

---

<sup>1</sup>PEREZ, Caroline Criado. **Invisible Women**: Data Bias in a World Designed for Men. New York: Adam Press, 2019. [recurso eletrônico AZW3].

<sup>2</sup>REUTERS. “Amazon ditched AI recruiting tool that favored men for technical jobs”. **The Guardian**. 11 out 2018. Disponível em: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>. Acesso em: 8 mar. 2025.

Como “caixas-pretas”, os sistemas de IA podem tornar-se ferramentas poderosas para manter normas patriarcais, uma vez que os processos internos de tomada de decisão podem ser obscuros e difíceis de serem responsabilizados, dificultando a identificação e a correção de viés de gênero nesses sistemas.

Um problema mais desafiador é que muitos sistemas de IA são projetados para o conceito de universalismo, ou seja, diante de uma sociedade cada vez mais plural, a atribuição de sentidos ao humano, sem levar em consideração a enorme diversidade étnica e cultural dos povos, torna-se um instrumento hábil a perpetuar, concretizar e estabilizar sistemas de marginalização social, impedindo ou dificultando o tratamento sem discriminação e o acesso igualitário a bens e serviços.

### 1.1 CONCEITO DE GÊNERO

A construção histórica do conceito de gênero e sua percepção como um marcador das diferenças entre homens e mulheres são exploradas por Simone de Beauvoir em *O Segundo Sexo* (2009). De acordo com a autora, a disparidade entre as categorias de masculino e feminino revela-se claramente na configuração unilateral dos mitos sexuais.

A construção e a representação do mundo, historicamente, têm sido empreendimentos conduzidos pelos homens, que impõem sua perspectiva particular como se fosse uma verdade universal<sup>3</sup>.

No curso desta obra, a filósofa discorre acerca de duas ideias centrais que ela divide em dois volumes, que se complementam. No primeiro, acerca dos mitos da sexualidade, sugere que os papéis de gênero e as concepções de masculinidade e feminilidade não são naturais, mas sim construções sociais reforçadas por meio das dinâmicas das interações sexuais e das percepções sociais associadas a essas interações, questionando a naturalização das diferenças sexuais e apontando para a influência da atividade sexual dos homens como estruturante dessas relações. No segundo volume, ela explora a noção de que o corpo humano possui uma natureza dual, funcionando simultaneamente como objeto físico no mundo e como uma perspectiva interpretativa. O corpo é, ao mesmo tempo, uma entidade material inserida na realidade objetiva e um meio subjetivo por meio do qual o indivíduo compreende e interpreta o mundo ao seu redor.

---

<sup>3</sup>BEAUVOIR, Simone. **O Segundo Sexo**. 1 v. Tradução de Sérgio Milliet. 2a ed. Rio de Janeiro: Nova Fronteira, 2009, p. 182).

Essas ideias denotam a complexidade da interação entre corpo, subjetividade e construções sociais, além de sublinharem a importância de desconstruir percepções naturalizadas sobre os gêneros e a corporeidade.

Joan Scott, por sua vez, explica a origem do conceito de gênero:

Na sua utilização mais recente, o termo “gênero” parece ter feito sua aparição inicial entre as feministas americanas, que queriam enfatizar o caráter fundamentalmente social das distinções baseadas no sexo. A palavra indicava uma rejeição do determinismo biológico implícito no uso de termos como “sexo” ou “diferença sexual”. O termo “gênero” enfatizava igualmente o aspecto relacional das definições normativas da feminilidade. Aquelas que estavam preocupadas pelo fato de que a produção de estudos recíprocos e não se poderia compreender qualquer um dos sexos por meio de um estudo inteiramente separado.<sup>4</sup>

Ela aponta que o termo “gênero”, além de um substituto para o termo “mulheres”, é também utilizado para sugerir que qualquer informação sobre as mulheres é necessariamente informação sobre os homens, que um implica o estudo do outro. Essa abordagem destaca que o universo feminino está intrinsecamente ligado ao universo masculino, sendo gerado dentro e por este contexto, e refuta a validade da interpretação que propõe esferas separadas, defendendo que estudar as mulheres de forma isolada perpetua o mito de que uma esfera, a experiência de um sexo, tenha muito pouco ou nada a ver com o outro sexo. Ademais, o termo “gênero” também é utilizado para designar as relações sociais entre homens e mulheres. Seu uso rejeita explicitamente explicações biológicas, como aquelas que encontram um denominador comum para diversas formas de subordinação feminina, como no fato de que as mulheres têm a capacidade para dar à luz e de que os homens têm força muscular superior.

Em vez disso, a autora argumenta que o termo “gênero” torna-se uma forma de indicar “construções culturais” – a criação inteiramente social de ideias sobre os papéis adequados aos homens e às mulheres. Trata-se de uma forma de se referir às origens exclusivamente sociais das identidades subjetivas de homens e de mulheres. Gênero é, segundo esta definição, uma categoria social imposta sobre um corpo sexuado.<sup>5</sup>

É nas relações entre homens e mulheres que o gênero se manifesta e se consolida. É também por essa razão que uma das passagens mais citadas do texto da autora é aquela em que ela afirma que o gênero é um elemento constitutivo das relações sociais fundamentadas nas diferenças percebidas entre os sexos, sendo o gênero o primeiro meio de dar significado

---

4SCOTT, Joan. “Gênero: Uma Categoria Útil de Análise Histórica”. **Educação & Realidade**, v.15, n. 2, jul./dez., 1990, p. 72.

5Idem, *ibidem*, p. 75.

às relações de poder<sup>6</sup>, ou por outra, que “[...] o gênero é um campo primário no interior do qual, ou por meio do qual, o poder é articulado.”

Em sua análise sobre a história das mulheres e o conceito de gênero, a pesquisadora percebe que, enquanto o primeiro termo carrega uma carga política ao afirmar explicitamente a validade das mulheres como sujeitos históricos, o segundo termo, “gênero”, inclui as mulheres sem necessariamente nomeá-las, o que pode ser interpretado como uma forma menos direta de abordar a questão, sugerindo que o uso do termo “gênero” pode não representar uma ameaça tão evidente às estruturas de poder estabelecidas quanto o termo “história das mulheres”<sup>7</sup>. Ela observa que o termo “história das mulheres” é explicitamente político, pois afirma a validade das mulheres como sujeitos históricos, contestando as práticas tradicionais que muitas vezes as excluíam ou marginalizam nas narrativas históricas. Ao nomear as mulheres diretamente, esse termo reivindica seu lugar na história e aponta para a necessidade de reconhecimento e inclusão de suas experiências e contribuições.

A disciplina histórica desempenhou papel preponderante na produção da “diferença sexual”. Ao salientar, muitas vezes, a participação exclusiva dos homens, essas narrativas acabam por construir e reforçar o conceito de gênero. Enquanto o sexo biológico é determinado por características físicas, o gênero constitui uma construção social, organizada por normas e papéis que estão ligados às percepções de masculinidade e feminilidade. Essa dinâmica evidencia as relações de poder que sustentam um sistema hierárquico, perpetuando formas de opressão e desigualdade. Assim, as normas de gênero não se limitam a meras classificações identitárias, mas funcionam como elementos estruturantes das relações sociais, promovendo a subordinação e marginalização das mulheres.

A crescente importância dos debates feministas tem impulsionado iniciativas governamentais que buscam melhorar a eficácia e a aplicação das leis no combate à violência contra a mulher. Nesse cenário, a expansão das legislações voltadas para a proteção de grupos minoritários exige novas interpretações sobre as relações estruturais que conectam as diversas dimensões da igualdade às variadas formas de discriminação.

De acordo com PEREZ<sup>8</sup>, a maior parte da história registrada da humanidade é caracterizada por uma grande ausência de informações. Desde a formulação da teoria do Homem Caçador, os relatos históricos negligenciaram o papel das mulheres no desenvolvimento

---

<sup>6</sup>SCOTT, *Op. cit.*, p. 88.

<sup>7</sup>*Idem, ibidem*, p. 72.

<sup>8</sup>PEREZ, *Op. cit.*

cultural e biológico da humanidade.

Em contrapartida, as experiências masculinas passaram a ser compreendidas como manifestações de todas as experiências humanas. No que diz respeito às mulheres, o que frequentemente se observa é o silêncio. Esses vazios de representação estão presentes em diversas esferas da cultura, manifestando-se em filmes, notícias, literatura, ciência, planejamento urbano e economia.

As narrativas sobre o passado, o presente e o futuro, que são construídas pela sociedade, estão todas marcadas por uma “ausência presente” simbolizada pela figura feminina. Este é o fenômeno denominado “lacuna de dados de gênero”.

Contudo, a lacuna de dados de gênero não se resume apenas ao silêncio. Esses vazios têm implicações concretas e afetam o cotidiano das mulheres. O impacto pode ser percebido de forma sutil, como no desconforto causado por ambientes de trabalho ajustados a padrões masculinos, ou como na dificuldade de alcançar prateleiras altas projetadas de acordo com a altura média dos homens. Embora possam parecer triviais, são, de fato, injustos, e revelam um sistema normativo desigual.

A substituição do critério de sexo biológico pelo de identidade de gênero no acesso a espaços femininos, como banheiros e vestiários, por exemplo, ganha contornos ainda mais difíceis quando analisado sob a ótica das lacunas de dados de gênero em sistemas de Inteligência Artificial. Esses sistemas, cada vez mais presentes na formulação de políticas públicas e na gestão de espaços coletivos, frequentemente se baseiam em dados limitados ou enviesados, comprometendo a análise de segurança e privacidade em espaços segregados por sexo, o que pode resultar na incapacidade de diferenciar adequadamente o sexo biológico da identidade de gênero autopercebida e, por sua vez, agravar a vulnerabilidade das mulheres.

Tal deficiência compromete a avaliação de riscos associados à presença de indivíduos do sexo masculino nesses espaços e a análise dos impactos psicossociais sobre as mulheres das políticas implementadas.

A fluidez do conceito de gênero torna ainda mais difícil a parametrização de dados em sistemas que dependem de classificações objetivas para oferecer soluções seguras e eficazes. A invisibilidade das realidades femininas em bancos de dados tecnológicos perpetua desigualdades e fragiliza a garantia de seus direitos fundamentais. Os algoritmos, que operam a partir de correlações, construção de perfis e detecção de padrões, podem reproduzir estereótipos e preconceitos, refletindo uma das principais diferenças em relação à inteligência

humana.

É importante compreender que a distinção entre sexo e gênero muitas vezes negligencia o fato de que a nossa percepção do corpo humano não é neutra, mas influenciada e moldada por padrões culturais.

O corpo não existe de maneira isolada, livre de interpretações; ao contrário, é o discurso social que define como o entendemos, criando, a partir daí, hierarquias de poder.

LERNER<sup>9</sup> discute como o patriarcado se desenvolveu a partir da divisão do trabalho e do poder. Ela examina as estruturas sociais, econômicas e políticas que sustentam e perpetuam a subordinação das mulheres, incluindo a instituição do casamento, a transferência da propriedade por meio do sistema patrilinear e a criação de mitos e narrativas que justificam a supremacia masculina. Essa mudança levou a uma crescente desigualdade entre homens e mulheres, com as mulheres sendo subordinadas aos homens e restringidas, principalmente, ao espaço privado e doméstico.

Na estrutura da família patriarcal, as responsabilidades e obrigações são distribuídas de forma desigual entre aqueles que estão sob proteção. Para os filhos homens, a submissão à autoridade paterna é temporária, restringindo-se ao período em que ainda não assumiram suas próprias famílias. Em contrapartida, para filhas e esposas, essa subordinação é permanente. As filhas só conseguem romper com a autoridade paterna ao se casarem, transferindo, porém, essa relação de dependência para um novo homem, o marido.

O paternalismo, por sua vez, baseia-se em um acordo tácito: o homem provê sustento financeiro e proteção, enquanto a mulher se submete integralmente, oferecendo trabalho doméstico não remunerado, serviços sexuais e obediência em todas as dimensões da convivência.<sup>10</sup>

O fenômeno da assimetria sexual, com a atribuição de diferentes tarefas e papéis para homens e mulheres, é observado em praticamente todas as sociedades humanas conhecidas.

A explicação da assimetria sexual, segundo a historiadora, leva em consideração fatores biológicos relacionados aos homens para justificar a submissão feminina. A força física superior, a habilidade de correr mais rápido e levantar pesos mais pesados, bem como uma maior agressividade nos homens, resultam em sua função como caçadores. Por conseguinte, eles se tornam os provedores de alimentos nas tribos, sendo mais valorizados e

---

<sup>9</sup>LERNER, Gerda. **A Criação do Patriarcado**: história da opressão das mulheres pelos homens. Tradução de Luiza Sellera. São Paulo: Cultrix, 2019.

<sup>10</sup>LERNER, *Op. cit.*, p. 291.

respeitados do que as mulheres. Além disso, as habilidades desenvolvidas por meio da experiência na caça também lhes permite se tornar guerreiros.

Segundo a autora, a defesa da supremacia masculina, baseada no determinismo biológico, sofreu transformações ao longo do tempo, revelando-se incrivelmente adaptável e resiliente.

A historiadora prossegue identificando que, com a perda de sustentação do argumento religioso no século XIX, a explicação tradicionalista da inferioridade feminina adotou uma abordagem “científica”. As teorias darwinistas reforçaram a crença de que a sobrevivência da espécie era mais importante do que a busca pela autorrealização individual. Assim, essa visão sustentava a ideia de que a supremacia masculina era justificada em prol da preservação da espécie. Conforme exposto pela autora, a visão tradicionalista das diferenças sexuais, frequentemente fundamentada em uma explicação divina ou natural, desempenhou um papel fundamental na justificação da divisão sexual do trabalho, isentando os homens de responsabilidade pela desigualdade de gênero e pela dominação masculina. Nessa perspectiva, a mulher, considerada destinada a exercer uma função biológica distinta da do homem, deveria, conseqüentemente, ter responsabilidades sociais diferenciadas. De forma semelhante, a psicologia moderna, ao analisar as diferenças sexuais, tratou-as como características intrínsecas, sem contestação, o que levou à construção de uma concepção psicológica da mulher essencialmente determinada pela biologia, como nas teorias anteriores.

Ao estudar os papéis de gênero sem considerar o contexto histórico, os psicólogos, a partir de dados clínicos, chegaram a conclusões que acabaram por reforçar as normas de gênero estabelecidas.

A teoria freudiana, em suas versões simplificadas, especialmente em relação à criação dos filhos, e a literatura popular de autoajuda, consolidaram a ideia de que a principal função da mulher seria a maternidade e a educação dos filhos, visão amplamente disseminada pela mídia e por educadores, solidificando-se como uma referência social.<sup>11</sup>

No cerne da estrutura patriarcal, a dependência econômica das mulheres dentro da família tem sido um dos principais pilares de sua subordinação. A divisão sexual do trabalho, que atribui aos homens as funções públicas e às mulheres as responsabilidades domésticas, perpetua uma desigualdade estrutural que se reflete em todos os aspectos da vida social. Nesse contexto, a supremacia masculina é reforçada e consolidada não apenas por fatores econômicos,

---

<sup>11</sup>LERNER, *Op. cit.*, p. 49.

mas também por mitos, narrativas e representações que legitimam a hegemonia do masculino sobre o feminino. Dessa forma, mantém-se uma divisão artificial e arbitrária entre os sexos, sustentada por um sistema de valores androcêntrico, que perpetua a dominação e a marginalização das mulheres em diversos espaços.

Claude Lévi-Strauss<sup>12</sup> dedicou-se ao estudo do parentesco, abordando-o como uma construção cultural que vai além dos aspectos biológicos da procriação. Ele reconhece que a sexualidade humana exerce um papel essencial na formação das estruturas sociais, sendo um fator determinante na análise das organizações humanas, e rejeita a ideia de uma sociedade neutra, na qual os indivíduos seriam desprovidos de identidade sexual ou destino predeterminado. Ao contrário, ele vê a sociedade como uma construção na qual os indivíduos são definidos desde o nascimento como masculinos ou femininos, e essa distinção organiza comportamentos, responsabilidades e destinos. O teórico aprofundou sua análise ao identificar a troca de mulheres entre homens como o fundamento do sistema de parentesco. Essa prática, longe de se limitar a um símbolo, revela a opressão sexual nas sociedades humanas, transformando as mulheres em peças-chave para as trocas sociais, políticas e econômicas. A troca de mulheres, considerada uma “dádiva”, se revela, portanto, como um mecanismo de organização social.

O autor procurou identificar os princípios que regem o parentesco, investigando as regras e os tabus que estruturam essas trocas. Por meio dessa pesquisa, ele evidenciou uma lógica expressa nas normas matrimoniais e nos interditos de diversas culturas, com destaque para o tabu do incesto. Este tabu, juntamente com a ideia da “dádiva” das mulheres, formava um sistema que não apenas regulava a sexualidade, mas também delineava os papéis e destinos das mulheres dentro da sociedade. Com essa análise, Claude Lévi-Strauss propôs uma nova forma de compreender as relações sociais.

Embora tenha reconhecido as mulheres como subordinadas nesse sistema, ele não as via como figuras passivas. Em vez disso, elas são consideradas agentes dentro de um processo de troca, que, embora subordinado a uma lógica de poder desigual, é fundamental para o funcionamento social. Essa perspectiva sugere uma teoria subjacente da opressão sexual nas relações humanas. O parentesco, enquanto estrutura organizacional, está intimamente relacionado ao poder.

---

<sup>12</sup>LÉVI-STRAUSS, Claude. **As Estruturas Elementares do Parentesco**. Tradução de Mariano Ferreira. Petrópolis: Vozes, 1982.

As mulheres, ao se tornarem objetos de troca, são intermediárias nas relações estabelecidas entre os homens. Como responsáveis por realizar essas trocas, são eles os principais beneficiários dos resultados dessa organização social, consolidando o domínio masculino nas estruturas sociais.

Para BOURDIEU<sup>13</sup>, o “poder simbólico” faz parte de uma “estrutura estruturada” da sociedade ocidental, de modo que está incorporado no âmbito das relações socioculturais implicitamente e subjetivamente. É uma forma de poder que contribui sobremaneira para a manutenção do *status quo*, operando sutilmente a partir do controle ideológico de massas, possível a partir da comunicação e do discurso – que se transformam em performatividades.

Os símbolos são entendidos como elementos que, no sentido antropológico, produzem sentidos e significados no campo social, de modo que criam “integração social”, possibilitando o “consenso acerca do sentido do mundo social que contribui fundamentalmente para a reprodução da ordem social: a integração ‘lógica’ é a condição de integração moral.”<sup>14</sup>.

De maneira mais ampla, o espaço das posições sociais reflete-se em um espaço de escolhas e ações, mediado pelo espaço das disposições (ou *habitus*). Para o autor, o *habitus* é um conjunto de disposições duradouras e incorporadas que são adquiridas pelos indivíduos através de sua socialização e interação com o mundo social, que incluem valores, crenças, atitudes, preferências e habilidades que moldam a forma como os indivíduos percebem, pensam e agem no mundo.<sup>15</sup>

O poder simbólico encontra sua legitimidade no Estado e em outras instâncias de poder, o que permite seu desenvolvimento nos diversos “campos” da vida social. Dessa forma, ele gera representações, discursos e práticas que moldam a sociedade.

Entre as expressões recorrentes desse poder e da violência simbólica, destaca-se a dominação masculina. Nas palavras do autor, “trata-se do exemplo por excelência da submissão paradoxal da violência simbólica – violência suave, insensível, invisível – a suas próprias vítimas.”<sup>16</sup>.

A partir da teoria suscitada por PATEMAN<sup>17</sup> (1993), em uma perspectiva política e histórica, a constituição da sociedade moderna se fundamenta na negação da condição humana à

---

13BOURDIEU, Pierre. **O poder simbólico**. Tradução de Fernando Tomaz. Difusão Editorial: Lisboa, 1989, p. 9.

14**Idem, ibidem**, p. 10.

15BOURDIEU, Pierre. **Razões Práticas sobre a Teoria da Ação**. Tradução de Mariza Corrêa. Campinas: Papirus, 2009, p. 20.

16BOURDIEU, Pierre. **A dominação masculina**, 2a ed. Tradução de Maria Helena Kühner. Rio de Janeiro: Bertrand Brasil, 2002, pp.1-2.

17PATEMAN, Carole. **O Contrato Sexual**. Paz e Terra, São Paulo, 1993.

mulher. Segundo a autora, a história do Contrato Social diz respeito à constituição da esfera pública da liberdade civil; a esfera privada, por outro lado, não é considerada politicamente relevante por se constituir do âmbito em que não há negociação política. Nesse sentido, a diferença sexual também se revela como uma questão política de liberdade e submissão. As mulheres, por não terem participado do contrato original através do qual os homens trocam sua liberdade natural pela segurança da liberdade civil, são simplesmente tratadas como objeto desse contrato (a liberdade civil incluindo até mesmo o direito de usufruir das mulheres), sem desempenhar um papel ativo nele, resultando na negação fundamental da categoria de “indivíduo” para as mulheres, histórica e socialmente, levando a restrições tanto sociais quanto jurídicas aos seus direitos de integração na sociedade.

No âmbito público, homens e mulheres ocupam posições antagônicas na hierarquia de valores sociais. Enquanto o homem é frequentemente associado a atributos como honra e virtude, sendo reconhecido por sua atuação destacada e influente na esfera social, com papéis de liderança e prestígio, a mulher, por outro lado, enfrenta estigmas que a colocam à margem. Ela é muitas vezes invisibilizada ou desvalorizada, sem que sua identidade própria seja devidamente reconhecida ou respeitada.

Essa dinâmica reflete uma desigualdade enraizada, que perpetua estereótipos e limita o espaço das mulheres na vida pública.

É fundamental ressaltar que o contrato sexual não está restrito apenas ao campo privado. O patriarcado não se limita à esfera familiar ou à vida privada; ele configura a sociedade civil como um todo. O contrato original estabelece a estrutura patriarcal em toda a sociedade civil. Os homens circulam entre os domínios privado e público, e a lei do direito sexual masculino regula ambos. Embora a sociedade civil seja fragmentada, a unidade da ordem social é, em grande medida, mantida pela rede de relações patriarcais.<sup>18</sup>

Da formulação do conceito de ação política, conforme apresentado por ARENDT<sup>19</sup>, surge sua teoria da esfera pública. De acordo com a filósofa<sup>20</sup>, na experiência da pólis, considerada o mais eloquente dos corpos políticos – e na filosofia política que dela emergiu –, houve uma separação entre ação e discurso, com estes se tornando atividades independentes.

A ênfase passou de agir para discursar, especialmente no contexto da persuasão, em vez de ser uma forma de resposta ou reação ao ocorrido.

---

18PATEMAN, *Op. cit.*, p. 29.

19ARENDT, Hannah. **A condição humana**. Tradução de Roberto Raposo. 13. ed. rev. [Reimpr.]. Rio de Janeiro: Forense Universitária, 2020.

20*Idem, ibidem*, pp. 83-84.

Segundo a autora, a esfera pública consiste na comunidade de indivíduos que habitam a pólis, onde todas as decisões eram tomadas por meio dos significados da vida em comum.

Ser político na pólis significava que as decisões eram tomadas por meio de palavras e persuasão, não por força ou violência. Para os gregos, a violência e a imposição, em vez da persuasão, eram formas pré-políticas de governar, típicas da vida doméstica, onde o chefe da casa governava de maneira autoritária, ou dos impérios bárbaros da Ásia, cujo despotismo era frequentemente comparado à autoridade familiar. A violência era vista como uma ação pré-política, própria da vida fora da pólis, sendo característica apenas do ambiente doméstico, onde o chefe da casa exercia um poder absoluto e despótico. Enquanto na esfera pública o poder é exercido através da persuasão e do discurso, na esfera privada ele é estabelecido pela estrutura familiar, baseado na satisfação dos desejos pessoais do homem.

Tais padrões de comportamento não são inatos, e tanto a masculinidade quanto a feminilidade são conceitos dinâmicos e historicamente construídos, variando de acordo com o contexto cultural e temporal. Dessa forma, o gênero é compreendido como um fenômeno social que não pode ser dissociado de seu contexto histórico e conceitual.

Essas identidades (homem/masculino, mulher/feminino) ganham significado através da linguagem e dos sistemas simbólicos pelos quais são representadas.

A representação desempenha um papel simbólico na classificação do mundo e de nossas relações dentro dele.

Atualmente, as mulheres desempenham diversas funções na sociedade, ocupando posições em empresas, liderando famílias, exercendo o direito de voto, buscando educação e, em alguns casos, governando.

Elas também têm a liberdade de escolher ser donas de casa e mães, caso assim desejem. No entanto, a conquista dessas oportunidades e da autonomia para determinar seu próprio destino foi resultado de um longo processo histórico e cultural de transformação do papel social da mulher. Esse processo envolveu a internalização e reprodução de normas, crenças e costumes, de modo que as realidades vivenciadas fossem vistas como naturais.

Na perspectiva fenomenológica de MERLEAU-PONTY<sup>21</sup> sobre a percepção e a relação entre o sujeito e o mundo, o “real” é visto como uma entidade coesa e contínua, independente das interpretações individuais, sugerindo que o mundo existe objetivamente.

---

<sup>21</sup>MERLEAU-PONTY, Maurice. **Fenomenologia da percepção**, 2a ed. Tradução de Carlos Alberto Ribeiro de Moura. São Paulo: Martins Fontes, 1999.

O mundo se apresenta como mais do que um objeto externo; ele é também o ambiente natural e o contexto no qual nossos pensamentos e percepções ocorrem. Não temos controle total sobre sua constituição. Em vez disso, estamos inseridos nele, e essa inserção é o que define nossa compreensão da realidade. Assim, nas palavras do pensador: “Buscar a essência do mundo não é buscar aquilo que ele é em ideia, uma vez que o tenhamos reduzido a tema de discurso; é buscar aquilo que de fato ele é para nós antes de qualquer tematização.”<sup>22</sup>

Contudo, é importante assinalar que a relação entre percepção e mundo da vida não é uniforme para todos os grupos sociais, principalmente para as mulheres, que enfrentam desafios específicos devido a estereótipos de gênero, discriminação e limitações sociais. A hegemonia masculina é construída por meio de diversas representações e mitologias<sup>23</sup>, as quais influenciam a percepção dos corpos na esfera social.

Os discursos desempenham papel determinante na formação e perpetuação dessa dominação, moldando a estrutura da sexualidade humana de acordo com as normas culturais estabelecidas, o que resulta na criação de uma hierarquia entre os sexos, onde o masculino é privilegiado em relação ao feminino. Essa perspectiva, enraizada no patriarcado e centrada no homem, cria uma divisão artificial do mundo como oposição entre o feminino e o masculino.

No âmbito do Direito, de acordo com COLLING<sup>24</sup>, a mulher era afastada do universo do pensamento e do conhecimento na Grécia clássica, áreas altamente valorizadas por essa sociedade. Já entre os romanos, o sistema de normas institucionalizou a discriminação contra a mulher por meio da figura do *paterfamilias*, que concentrava todo o poder nas mãos do homem, abrangendo o controle sobre a mulher, os filhos, os servos e os escravos.

O Direito, na Roma antiga, se tornou um instrumento para perpetuar essa desigualdade, conferindo legitimidade à subordinação da mulher na estrutura social. Por essa razão, ao buscar resgatar a presença feminina na história, foi necessário elaborar um novo marco, questionando as categorias tidas como universais e, ao mesmo tempo, destacando as particularidades, pluralidades e diferenças. De acordo com a autora, o corpo feminino deve ser entendido como um texto histórico, reescrito de maneiras diversas ao longo do tempo.

Assim como o homem, a mulher é uma construção social e não pode buscar em seu interior uma essência primordial.

---

22MERLEAU-PONTY. *Op. cit.*, p. 13.

23BOURDIEU, 2002, pp. 1-2.

24COLLING, Ana Maria. **Tempos diferentes**, discursos iguais: a construção do corpo feminino na história. Dourados: Ed. UFGD, 2014, p. 23.

Não existe, portanto, uma natureza feminina essencial, mas uma construção cultural em que, ao longo de séculos, a feminilidade foi vista como atributo natural da mulher. Não há uma “verdadeira mulher”, pois os termos “verdadeira” e “mulher” são conceitos criados, sendo, portanto, aparências e construções superficiais. Além desses conceitos, não há algo que possa ser denominado “mulher”; o que existe são apenas relações de poder e hierarquia, estabelecidas socialmente<sup>25</sup>.

Da mesma forma, não há uma essência masculina intrínseca, mas uma construção histórica e cultural que define a masculinidade como um conjunto de atributos naturalizados. O conceito de “verdadeiro homem” também é uma invenção social, moldada por expectativas e normas que variam conforme o contexto histórico e cultural. Assim, tanto a masculinidade quanto a feminilidade são formadas a partir de normas e expectativas culturais que estabelecem e reproduzem determinados modelos de comportamento e identidade.

Ao examinarmos os discursos que ocultam a história das mulheres, identificamos as razões específicas e generalizadas dessa invisibilidade.

A dominação masculina impôs às mulheres espaços de não visibilidade, silenciamento e confinamento ao âmbito privado e à função reprodutiva. Dessa forma, a história das mulheres se caracteriza principalmente como a história de seus corpos.

As mulheres foram sistematicamente excluídas, silenciadas e mantidas invisíveis, não apenas nos espaços domésticos e privados, mas também nos setores econômico, social e cultural. Frequentemente foram retratadas e descritas por discursos misóginos que negavam seu direito à voz, defesa e liberdade. É essencial explorar esse lado negligenciado da história e compreender como tais discursos contribuíram para uma visão distorcida da realidade.

Uma história das mulheres requer que a narrativa histórica em geral seja reconhecida como o produto de interpretações e representações moldadas por relações de poder subjacentes. Essas representações da mulher, que se estendem ao longo do tempo, fundamentaram o pensamento simbólico sobre a diferença entre os sexos.

As mulheres enfrentam o desafio constante de equilibrar suas responsabilidades na esfera privada e pública. A busca por legitimidade política e pelo reconhecimento como sujeitos de direitos representa uma jornada contínua e complexa para as mulheres, cuja trajetória envolve a construção da própria identidade, a inserção em comunidades, a ocupação de espaços de representação e a atuação política, com o objetivo de defender suas demandas

---

<sup>25</sup>COLLING, *Op. cit.*, p. 27.

específicas e influenciar as agendas coletivas.

A cultura, como construção coletiva, é formada pelos indivíduos, que são, ao mesmo tempo, moldados pelas práticas, leis e convenções que eles mesmos estabelecem e propagam.

O machismo e a misoginia, ao longo da história, contribuíram para a subordinação das mulheres, afetando suas oportunidades, direitos e visão de mundo. Isso torna fundamental considerar esses fatores ao analisar a relação das mulheres com a sociedade, com o intuito de compreender suas experiências e perspectivas singulares de maneira integral.

O corpo feminino foi historicamente moldado e analisado sob diversas perspectivas; filosóficas, médicas, pedagógicas, psicológicas e jurídicas. Esses discursos, embora distintos, se interconectam de forma sutil, contribuindo para definir o que significa ser homem ou ser mulher e quais papéis sociais cada um deve desempenhar.

Dessa maneira, eles delineiam a construção cultural da identidade feminina e a concepção de um corpo sexualmente marcado.

Esse processo parece refletir uma trajetória dominada pela razão masculina, enquanto as mulheres, em contrapartida, demonstram resistência a essa lógica.

As variações no pensamento filosófico e suas conexões com outros campos de conhecimento não são apenas teóricas, mas práticas que moldam a sociedade, estabelecendo padrões específicos para homens, mulheres e suas relações. Aquelas que desafiam essas normas são frequentemente vistas como uma ameaça, sendo acusadas de promover a desestabilização das estruturas familiares e sociais.

Essa perspectiva contribui para a manutenção das desigualdades enfrentadas pelas mulheres em várias áreas, como participação política, acesso a oportunidades educacionais e profissionais, e liberdade de expressão, estando atrelada aos estereótipos de papéis de gênero. Enquanto os homens são frequentemente caracterizados como assertivos, fortes e independentes, as mulheres são associadas a traços como delicadeza, subordinação, passividade e emotividade. Tais expectativas limitam a liberdade individual e restringem o desenvolvimento pessoal, intensificando as disparidades sociais e materiais.

O quadro apresentado não deve ser encarado como algo natural, mas sim como uma construção histórica, passível de desconstrução e superação.

## 1.2 IDENTIDADE E PERSONALIDADE

A identidade, inerente à natureza singular e irrepetível da pessoa humana, configura-se como um projeto em contínuo desenvolvimento, jamais concluído. Mais do que uma construção formal jurídica, ela surge dos sentidos de existência e evolução, refletindo uma dimensão pessoal e única na qual não se trata de simplesmente possuí-la, mas de constituí-la em sua própria essência.

A identidade de um indivíduo é construída a partir de um conjunto de elementos e características que ele assume ou rejeita, definindo seu lugar no mundo, mesmo que de forma transitória. Diante disso, os direitos da personalidade desempenham um papel essencial, pois garantem que cada pessoa possa expressar sua identidade de maneira livre e digna, respeitando suas particularidades existenciais. Esses direitos não se confundem com outras categorias, como os direitos sociais, do consumidor, trabalhistas, garantias fundamentais, políticas públicas ou questões tributárias.

Embora todos esses também visem proteger a dignidade humana, os direitos da personalidade destacam-se por focar especificamente na realização da singularidade identitária. Eles funcionam como ferramentas que permitem a afirmação das diferenças individuais, assegurando que a identidade única de cada um seja preservada como um caminho para a emancipação pessoal.

Ao visar uma identidade digna, torna-se evidente que a abordagem requer “uma dinâmica de ubiquidade entre igualdade e diferença, entre o direito à igualdade e o direito à diferença.”<sup>26</sup>

Ainda que frequentemente considerados como termos intercambiáveis, personalidade e identidade não devem ser confundidos. A personalidade reflete principalmente “como” uma pessoa geralmente se comporta (serena, agressiva, ativa, passiva, amigável, rude, etc.), enquanto a identidade define “quem” uma pessoa é (ou não é), abrangendo aspectos como nacionalidade, raça, aparência física, nome, orientação sexual, etc.

Mesmo que intimamente relacionadas, personalidade e identidade não são sinônimos; a personalidade influencia a identidade, moldando a pessoa. Juntas, essas duas dimensões definem “quem ela é” e “como ela é”. Resumidamente, a identidade engloba a personalidade, estabelecendo uma relação intrínseca entre ambas.

---

<sup>26</sup>SCHREIBER, Anderson. **Direitos da personalidade**. 3.ed. São Paulo: Atlas, 2014, p. 258.

Seguindo a análise de MOREIRA<sup>27</sup>, os mecanismos de discriminação têm suas raízes nas diferenças de *status* cultural entre grupos sociais, o que justifica tratamentos discriminatórios entre eles.

Nascer com genitais aparentes do sexo feminino e, por isso, ser coercitivamente assinalada com o gênero feminino, estabelece padrões morais de julgamento social. Esses padrões, de maneira institucional, estrutural e intergeracional, geram expectativas sobre o comportamento das mulheres, tendendo a desvalorizá-las e restringi-las ao pleno exercício de seus direitos econômicos, sociais, políticos e culturais, em paridade com os homens. Pessoas com corpos de mulheres e corpos de homens não são pessoas similarmente situadas em consideração ao gênero (baseado na opressão). O gênero não é simplesmente um binário neutro, mas uma hierarquia.

A hierarquia de gênero influencia, entre outras, aquelas questões relacionadas à reprodução, contracepção e direitos reprodutivos, pelas quais as mulheres enfrentam pressões para se conformar a padrões de maternidade acarretando menos controle sobre suas próprias decisões reprodutivas.

A partir dessas reflexões, é possível perceber que o papel da mulher é fortemente condicionado por contextos sociais e históricos complexos, sendo continuamente moldado e ajustado em resposta a esses fatores.

Mulheres são sujeitos que carregam em sua história um longo processo de dominação, a quem são impostas restrições e subjugação da sua própria identidade pelas normas de gênero. Como resultado, frequentemente se tornam as principais vítimas de violência doméstica, agressões sexuais e disparidades salariais.

A discriminação de indivíduos costuma ter como base o desejo de preservar estruturas sociais que asseguram o privilégio de certos grupos, enquanto mantêm outros em posição de subordinação. Os métodos usados para perpetuar essa discriminação são justificados por uma série de estereótipos culturais, que são representações construídas por grupos majoritários – aqueles que detêm o poder simbólico e político para criar e disseminar significados culturais.<sup>28</sup>

A identidade e o comportamento das mulheres na sociedade, portanto, não são rígida e/ou exclusivamente definidos por aspectos biológicos, mas são amplamente influenciados pelas ideologias dominantes e pela cultura em que estão inseridas, manifestando-se através dos estereótipos de gênero e das expectativas que se formam a partir deles.

---

<sup>27</sup>MOREIRA, Adilson José. **Tratado de Direito Antidiscriminatório**. São Paulo: Contracorrente, 2020.

<sup>28</sup>*Idem, ibidem*, p. 314.

Pode-se, assim, perceber que a subjetividade das mulheres é influenciada pelas interações estabelecidas entre os agentes sociais, as instituições existentes e os eventos que afetam ambos.

De tal forma, entendemos que utilizar o conceito de “gênero” em vez de “sexo” implica determinar que a posição das mulheres adota submissão às construções sociais e políticas. Ser homem ou mulher é uma construção simbólica que surge dentro do contexto dos discursos que constroem identidades. É necessário, portanto, questionar e eliminar estereótipos universais e valores tidos como inerentes à feminilidade natural.

Com isso, compreende-se que o gênero é uma ferramenta poderosa, porém não inevitável, de organizar as relações sociais e distribuir poder, inclusive recursos materiais entre os sexos.

O reconhecimento dos direitos individuais ou coletivos está frequentemente vinculado à aceitação de seus sujeitos como participantes legítimos na esfera pública. Todavia, essa participação é moldada por estruturas sociais que sustentam desigualdades de gênero, reforçando processos de hierarquização e diferenciação.

Ao compreender o gênero como um fenômeno social e político, abre-se a possibilidade de questionar as estruturas de poder que organizam as relações entre os sexos, criando condições para a transformação da sociedade em direção à igualdade de direitos e oportunidades.

### 1.3 VULNERABILIDADE EXISTENCIAL

A concepção de vulnerabilidade evoca a ideia de fraqueza ou debilidade de um sujeito em relação a outro em um contexto específico, especialmente quando essa relação envolve uma situação de risco, e decorre da condição de finitude e fragilidade que todos compartilham, uma característica intrínseca à existência. Contudo, essa fragilidade se intensifica quando a vida está continuamente sujeita a riscos que escapam à capacidade de proteção do indivíduo, tornando-o suscetível a danos maiores.

A vulnerabilidade não se limita apenas ao aspecto físico; ela se estende à própria existência, uma vez que o ser humano depende de símbolos e sistemas simbólicos para garantir sua continuidade e bem-estar.

Além dessa vulnerabilidade inerente, alguns indivíduos e grupos enfrentam diretamente circunstâncias adversas, como pobreza, falta de educação, barreiras geográficas, doenças crônicas, violência e outros infortúnios, que os tornam ainda mais vulneráveis.

Nesses casos, essas pessoas são frequentemente incapazes, total ou parcialmente, de proteger seus próprios interesses devido à falta de poder, educação, recursos ou outras habilidades necessárias para fazê-lo.

Identificar o processo pelo qual os vulneráveis se tornam mais suscetíveis a um risco potencial é importantíssimo, a fim de evitar que passem de uma condição de vulnerabilidade para uma situação de estar vulnerável. Isso requer uma larga compreensão das instâncias e fatores envolvidos, incluindo o Estado, a comunidade, os sistemas econômicos e sociais, a cultura e a moralidade vigente. Todos esses elementos podem contribuir para colocar um indivíduo ou grupo de indivíduos em uma situação concreta de vulnerabilidade.

No campo do direito, a vulnerabilidade existencial descreve a situação subjetiva em que uma pessoa ou titular de direitos está mais exposta a prejuízos em aspectos não materiais de sua vida, particularmente aqueles ligados aos direitos da personalidade.

Essa condição exige a aplicação de regras jurídicas específicas e protetivas, destinadas a garantir o cumprimento do princípio da dignidade humana.

Tais vulnerabilidades muitas vezes surgem de estruturas políticas e econômicas influenciadas por trajetórias históricas injustas, que beneficiam alguns grupos enquanto marginalizam outros (vulnerabilidade social). Esse cenário gera desigualdades com impactos significativos nas esferas econômica, ideológica e política, além de serem reforçadas por um contexto cultural que molda nossa visão de mundo e valores (vulnerabilidade moral). Além da posição social, diversos fatores como religião, costumes e arte influenciam essa construção.

Como resultado, as situações de vulnerabilidade moral são mais sutis e difíceis de serem identificadas, pois são alimentadas por convicções humanas e, por vezes, negadas como formas de vulnerabilidade.

BARLETTA & PALMEIRA<sup>29</sup> analisam a condição de vulnerabilidade da mulher em razão do não reconhecimento dos seus direitos sexuais e reprodutivos. Mais que uma aspiração, esse princípio está consolidado na Conferência do Cairo sobre População e Desenvolvimento

---

<sup>29</sup>BARLETTA, Fabiana; PALMEIRA, Carolina S. de Sá. “Vulnerabilidade da Mulher, Autonomia Privada e o Exercício de Direitos Reprodutivos e Sexuais”. **Vulnerabilidades e suas Dimensões Jurídicas**. (Coord.) Fabiana Barletta e Vitor Almeida. Indaiatuba: Foco, 2023, pp. 97-109.

de 1994<sup>30</sup>.

A Conferência Mundial de Beijing de 1995<sup>31</sup> reitera os conceitos de saúde reprodutiva e os princípios consagrados na Conferência do Cairo e destaca que os direitos reprodutivos e sexuais são parte inalienável dos direitos humanos universais e indivisíveis.

Acrescentamos a esse rol as disposições contidas na Convenção de Belém do Pará, ratificada pelo Brasil através da promulgação do Decreto nº 1.973/1996<sup>32</sup>.

Ao tratarem sobre a autonomia privada na perspectiva jurídica, as autoras reconhecem-na como o poder de autodeterminação do indivíduo ou de grupos, permitindo que organizem suas próprias escolhas de vida. Essa tutela da pessoa humana, concebida como uma cláusula geral, projeta-se na dignidade como princípio constitucional fundamental. O Código Civil traduz essa proteção nos direitos da personalidade, construídos como direitos inalienáveis e voltados a garantir o desenvolvimento humano em suas múltiplas dimensões – física, intelectual e psíquica. Nessa perspectiva, dizem as autoras, “a autonomia existencial, como espécie do gênero autonomia privada, é reconhecida como instrumento de realização de liberdades individuais e de interesses não patrimoniais, que rege situações jurídicas existenciais, e que, por sua vez, geram efeitos na esfera jurídica individual ou na esfera jurídica de terceiros”.<sup>33</sup>

Ainda assim, no campo dos direitos reprodutivos e sexuais, há um histórico de vulnerabilidade imposto às mulheres. Tal condição decorre, em grande parte, de sua capacidade biológica de gerar vida – muitas vezes em descompasso com o desejo de reproduzir-se – e da dificuldade de acesso a métodos contraceptivos eficazes. Outrossim, persiste uma visão social que frequentemente as coloca em posição de menor estatura, sem o reconhecimento pleno de sua aptidão psicossocial para decidir livremente sobre o próprio corpo, uma realidade que exige o fortalecimento de medidas que assegurem à mulher sua condição de sujeito pleno de direitos, livre para conduzir sua existência com dignidade e autonomia.

---

30ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Plataforma do Cairo**, 1994. Relatório da Conferência Internacional sobre População e Desenvolvimento. Disponível em: <http://www.unfpa.org.br/Arquivos/relatorio-cairo.pdf>. Acesso em: 8 mar. 2025.

31ORGANIZAÇÃO DAS NAÇÕES UNIDAS. IV Conferência das Nações Unidas sobre a Mulher, 1995. **Declaração e Plataforma de Ação de Pequim**. Disponível em: [https://www.onumulheres.org.br/wp-content/uploads/2013/03/declaracao\\_beijing.pdf](https://www.onumulheres.org.br/wp-content/uploads/2013/03/declaracao_beijing.pdf). Acesso em: 8 mar. 2025.

32BRASIL. Congresso Nacional. **Decreto nº 1.973**, de 1º de agosto de 1996. Promulga a Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher, concluída em Belém do Pará, em 9 de junho de 1994. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decreto/1996/d1973.htm#:~:text=DECRETO%20N%C2%BA%201.973%2C%20DE%201%C2%BA,9%20de%20junho%20de%201994](https://www.planalto.gov.br/ccivil_03/decreto/1996/d1973.htm#:~:text=DECRETO%20N%C2%BA%201.973%2C%20DE%201%C2%BA,9%20de%20junho%20de%201994). Acesso em: 8 mar. 2025.

33BARLETTA & PALMEIRA, *Op. cit.*, p. 103.

É importante salientar que o termo “vulnerabilidade” é frequentemente associado a uma dimensão mais pessoal, referindo-se à maneira como as vulnerabilidades programáticas, sociais e individuais são vivenciadas e experienciadas. Em contraste, o conceito de “risco” se relaciona a uma situação de fragilidade enfrentada por uma população, grupo ou sociedade.

De modo que podemos compreender que a vulnerabilidade humana está vinculada a um risco coletivo, e, por conseguinte, político.

No caso das mulheres, essa vulnerabilidade assume uma dimensão ainda mais ampla, ultrapassando os limites da esfera individual e psíquica, configurando-se, portanto, como um risco coletivo e político. Tal situação se traduz em uma luta contínua pela autonomia e pela liberdade, no enfrentamento das manipulações culturais e sociais.

#### 1.4 MANIPULAÇÃO DIGITAL E VULNERABILIDADE HUMANA

Entender as transformações sócio jurídicas contemporâneas, explorando as alterações do Direito em face de um mundo cada vez mais interconectado e as tensões entre normatividade estatal e a emergência de tecnologias digitais, leva a uma séria reflexão sobre o papel dos algoritmos como elementos promotores de uma deformação social conforme vemos atualmente, trazendo novas práticas discursivas que dão corpo a um quadro de patologias sociais.

Hans Kelsen escreveu, em 1922, “O conceito de Estado e a psicologia social”<sup>34</sup> influenciado pelo livro *Psicologia das Massas e Análise do Eu*<sup>35</sup>. Para determinar se o grupo social, concebido pelo Estado retém as características desse vínculo, a investigação proposta por ele concentrou-se em “Totem e Tabu”<sup>36</sup>, um texto que orienta a história da humanidade por meio de uma narrativa mítica sobre a revolta dos filhos contra o líder da horda primitiva, resultando em parricídio.

O autor defende que o pensamento primitivo era marcado pela predominância do emocional sobre o racional na psique humana. Nesse estágio, as coisas ganhavam existência ao serem investidas de emoções como esperança, temor, desejo ou terror.

O homem primitivo agia impulsionado por essas emoções, tendo o desejo como o

---

34KELSEN, Hans. **A Democracia**. Tradução de Ivone Castilho Benedetti, Jefferson Luiz Camargo, Marcelo Brandão Cipolla, Vera Barkow. São Paulo: Martins Fontes, 2000.

35FREUD, Sigmund. **Psicologia das Massas e Análise do Eu**. Brasil: Lebooks, 2020.

36FREUD, Sigmund. **(1912-1914) - Obras completas volume 11: Totem e tabu, Contribuição à história do movimento psicanalítico e outros textos**. São Paulo: Companhia das Letras, 2012.

elemento central de sua experiência emocional.

O que Sigmund Freud denominou de inconsciente, representando os desejos do “id”, é abordado por KELSEN<sup>37</sup> como uma ausência de consciência do “eu”. Ele observa que, nas sociedades primitivas, a tendência emotiva prevalece sobre o racional. Em sua reflexão, ele argumenta que o homem primitivo não age de forma coletivista apenas por não possuir uma consciência clara do “eu”, mas também porque tende a enxergar sua existência como algo profundamente ligado à realidade do grupo. Dessa forma, segundo o jurista, o indivíduo não separa sua vida pessoal da vida coletiva, rejeitando a ideia de um ser completamente independente ou isolado da comunidade.

Com base nesse raciocínio, ele desenvolve o conceito de democracia, destacando a importância do relativismo ético e do valor da tolerância<sup>38</sup>, elementos fundamentais para a concretização do direito à participação no processo legislativo, visando especialmente à realização do princípio democrático central: a liberdade como autodeterminação.

Byung-Chul Han<sup>39</sup>, ao explorar como os algoritmos configuram novas formas de poder, abre a discussão acerca do risco coletivo e político que advém da manipulação psicológica – que ele denomina de “psicopolítica”. Diferentemente das formas clássicas de controle, a psicopolítica atua diretamente sobre os desejos e a subjetividade dos indivíduos, utilizando os dados gerados pelas interações digitais para prever e influenciar comportamentos.

A nova sociedade digital não atua por meio de proibições ou pelo silenciamento direto, mas sim pelo estímulo contínuo à comunicação, ao compartilhamento de experiências e opiniões, e à exposição voluntária de desejos e histórias pessoais. Através de uma sedução amigável, ela captura e quantifica a psique humana com o auxílio dos *big data*, incentivando o uso de dispositivos que promovem o automonitoramento.

Nesse “pan-óptico digital” do século XXI, no qual a internet e os smartphones assumem papel central, o sofrimento não é imposto por torturas físicas, mas pela compulsão de tuitar, postar e expor. Nesse processo, o sujeito se torna um produtor constante de dados pessoais, transformados em mercadorias e monetizados sem cessar.

HAN aponta para a transição paradigmática que caracteriza o momento presente, mostrando como a liberdade, na lógica contemporânea, assume um aspecto paradoxal, transformando-se em uma forma de restrição.

---

37KELSEN, *Op. cit.*, pp. 303-343.

38KELSEN, *Op. cit.*, pp. 345-357.

39HAN, Byung-Chul. **Psicopolítica e Neoliberalismo**. Belo Horizonte: Ayiné, 2018.

Hoje, acreditamos que somos sujeitos submissos, mas projetos livres, que se esboçam e se reinventam incessantemente. A passagem do sujeito ao projeto é acompanhada pelo sentimento de liberdade. E esse mesmo projeto já não se mostra tanto como uma figura de coerção, mas sim como uma forma mais eficiente de subjetivação e sujeição. O “eu” como projeto, que acreditava ter se libertado das coerções externas e das restrições impostas por outros, submete-se agora a coerções internas, na forma de obrigações de desempenho e otimização.<sup>40</sup>

O estudo conduzido por MIRBABAIE, STIEGLITZ & MARX<sup>41</sup> oferece uma análise detalhada sobre o amplo impacto das inovações tecnológicas, especialmente no âmbito laboral e suas limitações. Os profissionais, sobretudo aqueles que atuam como trabalhadores do conhecimento, mantêm contato constante com dispositivos eletrônicos durante sua jornada de trabalho. Paralelamente, redes sociais e formas de entretenimento digital consomem boa do tempo livre das pessoas. A exposição prolongada às telas pode gerar sérias implicações para o bem-estar, culminando em um fenômeno conhecido como “ansiedade tecnológica”, que surge como efeito adverso, seja de forma direta ou indireta, do uso de dispositivos computacionais.

Ao analisarem e influenciarem as decisões dos indivíduos, os algoritmos exploram as fraquezas psíquicas ligadas aos desejos mais profundos de aceitação e pertencimento, atuando de maneira a distorcer ou moldar os impulsos e escolhas do indivíduo, afetando sua autonomia psíquica. Por conseguinte, a vulnerabilidade humana, ao ser constantemente invadida e direcionada por essas influências externas, coloca o sujeito em uma posição de fragilidade, onde sua capacidade de individuação – ou seja, de ser verdadeiramente si mesmo – fica comprometida.

À medida que os algoritmos se consolidam como instrumentos de comunicação sistemática, a manipulação algorítmica representa uma forma de vulnerabilidade existencial.

A vulnerabilidade existencial, no contexto da psicanálise freudiana, refere-se à condição humana de enfrentar as tensões entre o indivíduo e suas pulsões internas, assim como o ambiente externo que constantemente ameaça o equilíbrio psíquico.

FREUD<sup>42</sup> posicionou a subjetividade no entrelaçamento entre pulsões e repressões culturais. Nesse passo, os algoritmos e as redes sociais assumem o papel de um “superego” digital, impondo normas e expectativas externas às quais os indivíduos se submetem, muitas vezes sem desenvolver uma consciência crítica.

---

40HAN, *Op. cit.*, p. 9.

41MIRBABAIE, M., STIEGLITZ, S. & MARX, J. “Digital Detox”. *Bus Inf Syst Eng* **64**, 239–246 (2022). Disponível em: <https://doi.org/10.1007/s12599-022-00747-x>. Acesso em: 8 mar. 2025.

42FREUD, Sigmund. *Além do princípio de prazer*. Tradução de Renato Zwick. Porto Alegre: L&PM, 2016.

Na psicanálise de Sigmund Freud, o “id”, o “ego” e o “superego” são elementos estruturais da mente, cuja interação molda tanto a personalidade quanto o comportamento humano. Estas instâncias psíquicas se caracterizam por serem interdependentes e dinâmicas, contribuindo para o entendimento das tensões internas e dos processos psíquicos.<sup>43</sup>

A manipulação algorítmica interfere no que o criador da Psicanálise denominou de “princípio de individuação”, dissolvendo os limites entre desejos autênticos e desejos induzidos externamente.

O controle operado pelas plataformas digitais e aliado ao design dos algoritmos que as compõem, enfraquece o equilíbrio entre o “id”, o “ego” e o “superego”, desestabilizando a psique.

A constante comparação social, promovida pelas redes, intensifica sentimentos de inadequação, solidão e exclusão, produzindo angústia. Esse ambiente gera uma constante tensão psíquica intensificando sentimentos de ansiedade, culpa e inadequação.

O “id” refere-se à parte instintiva e primitiva da mente, que abriga desejos e impulsos inconscientes. Guiado pelo “princípio do prazer”, busca atender às necessidades de forma imediata, ignorando restrições sociais ou implicações éticas. O “ego” atua como a instância racional, mediando as demandas do “id” e as imposições do “superego”. Sob a orientação do “princípio da realidade”, o “ego” procura estratégias viáveis e socialmente aceitáveis para lidar com os impulsos e os limites impostos pelo contexto externo. Já o “superego” representa a dimensão moral e crítica da mente, incorporando os valores, as normas culturais e os princípios éticos internalizados, geralmente transmitidos pelas figuras parentais durante a infância. Sua função é regular o comportamento, impondo restrições e gerando sentimentos de culpa quando os padrões internalizados são violados. Essas três instâncias operam de forma contínua e, por vezes, conflituosa: o “id” busca satisfação imediata, o “superego” estabelece limites baseados em valores éticos, e o “ego” trabalha para equilibrar essas forças em resposta às exigências da realidade.

O enfraquecimento do equilíbrio psíquico, portanto, ocorre porque o “id” é constantemente estimulado pelas recompensas instantâneas e pelo fluxo ininterrupto de conteúdos, enquanto o “superego” é sobrecarregado por padrões inalcançáveis de sucesso e comportamento promovidos nas redes, e o “ego”, por sua vez, encontra dificuldades em desempenhar sua função mediadora, já que as demandas excessivas dessas instâncias são

---

43FREUD, Sigmund. (1923-1925) **Obras completas volume 16**: O Eu e o Id, “Autobiografia” e Outros Textos. Tradução de Paulo César de Souza. São Paulo: Companhia das Letras, 2011, pp. 9-43.

amplificadas pela comparação social contínua e pela sensação de insuficiência, dificultando a adaptação saudável às pressões internas e externas.

Para LACAN<sup>44</sup>, o sujeito é constituído na relação com o Outro, mediada pela linguagem e pelo simbólico.

Na psicanálise de Jacques Lacan, o Outro se refere a uma instância simbólica que nos permite existir como sujeitos dentro de uma ordem social, mas também é o que nos impõe limites e nos faz sentir uma “falta”, pois nunca conseguimos nos completar totalmente dentro dessas normas externas. Em termos simples, o Outro representa a estrutura da linguagem, da cultura e das normas sociais que moldam a nossa identidade e a forma como nos entendemos.

É como um grande “depósito” de significados, regras e expectativas que existem fora de nós, mas que influenciam profundamente quem somos e como nos relacionamos com o mundo.

Os algoritmos e as plataformas digitais, ao operarem como mediadores invisíveis, tornam-se novos “Outros” estruturantes, que organizam a experiência subjetiva e a interação social. A lógica algorítmica reconfigura o “Outro” como um sistema de significantes operado por interesses ideológicos e mercadológicos. Os indivíduos, na busca incessante por validação social (gamificação, curtidas, compartilhamentos), são capturados no circuito do desejo, perpetuando um ciclo de falta e insatisfação. Sob essa perspectiva, o design algorítmico intensifica a alienação ao reforçar uma relação especular do sujeito com sua imagem idealizada no campo do olhar do “Outro” digital.

O conceito lacaniano de gozo ajuda a compreender como as plataformas incentivam comportamentos repetitivos e previsíveis. A busca por validação e pertencimento explora um gozo compulsivo, alienante e paradoxalmente insatisfatório, que aprisiona o sujeito em padrões consumistas e submissos.

Assim, a ideia de liberdade se desloca: deixa de ser apenas a ausência de coerção para abarcar também a autonomia psíquica e informacional.

Na sociedade, os indivíduos interagem de acordo com regras que definem expectativas e conferem significado às suas ações. A manipulação algorítmica, então, atua como um desafio à individuação saudável, pois busca obscurecer a consciência crítica do sujeito e reorientar suas escolhas em favor de interesses externos.

---

<sup>44</sup>LACAN, Jacques. **A Identificação: seminário 1961 – 1962**, Tradução de Ivan Corrêa e Marcos Bagno. Recife: Centro de Estudos Freudianos do Recife, 2003.

Ao considerar os impactos psicológicos das redes digitais, o Direito deve proteger os indivíduos contra a exploração algorítmica de suas subjetividades. O objetivo é assegurar que a liberdade seja efetiva, e não meramente formal. Para as mulheres, esse desafio se revela de forma dupla: convivendo com uma estrutura de poder histórica que impõe uma identidade subalterna, ainda enfrentam restrição em sua capacidade de se auto representar.

Embora o Brasil tenha adotado leis específicas para a proteção à mulher no ambiente doméstico e do trabalho, a misoginia como prática ainda corrente entre nós demonstra a ocorrência de uma proteção deficiente, que não é capaz de prevenir ou erradicar a misoginia social como forma de violência específica.

Com vistas a corrigir essa situação, tramita no Senado proposta que criminaliza a misoginia (PL 896/2023<sup>45</sup>), a fim de incluí-la no rol da Lei nº 7.116/1989<sup>46</sup>.

Misoginia é o ódio, a repulsa, o desprezo ou aversão às mulheres. Como forma de violência, a misoginia se expressa através de comportamentos agressivos, deprecições e objetificação da sexualidade da mulher, por palavras, escritos ou gestos. A misoginia está instalada na sociedade como uma prática derivada do machismo e do sexismo, os quais naturalizam a ideia de inferioridade das mulheres em relação aos homens, baseando-se em funções estereotipadas de gênero. A estrutura social, fortemente calcada em um sistema de valores patriarcais, reforça o senso comum de que as mulheres são ‘homens imperfeitos’ e devem ser condicionadas desde a infância a se submeterem à vontade masculina. As violências que atingem as mulheres constituem a forma mais extrema de desigualdades historicamente estabelecidas, perpetuando-se, com leves variações, nos campos social, político, cultural e econômico da grande maioria das sociedades e culturas<sup>47</sup>.

De acordo com o Escritório do Alto Comissário das Nações Unidas para os Direitos Humanos<sup>48</sup>, um estereótipo de gênero é uma opinião ou um preconceito generalizado sobre atributos ou características que homens e mulheres possuem ou deveriam possuir ou das

---

45BRASIL. Senado Federal. **Projeto de Lei nº 896/2023**. Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/03/07/proposta-que-criminaliza-misoginia-comeca-a-tramitar-no-senado>. Acesso em: 8 mar. 2025.

46BRASIL. Congresso Nacional. **Lei nº 7.116**, de 5 de janeiro de 1989. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/17716.htm](https://www.planalto.gov.br/ccivil_03/leis/17716.htm). Acesso em: 8 mar. 2025.

47INSTITUTO PATRÍCIA GALVÃO. “Cultura e Raízes da Violência contra Mulheres”. **Dossiê Violência Contra Mulheres**. Disponível em: <https://dossies.agenciapatriciagalvao.org.br/violencia/violencias/cultura-e-raizes-da-violencia/>. Acesso em: 8 mar. 2025.

48DANTAS, Maria Eduarda Borba. “Dimensões da violência contra mulheres defensoras de direitos humanos no Brasil”. **ONU MULHERES BRASIL**. Disponível em: <http://www.onumulheres.org.br/wp-content/uploads/2021/11/Relatorio-Defensoras-Violencia1.pdf>. Acesso em: 8 mar. 2025.

funções sociais que ambos desempenham ou deveriam desempenhar. Por conseguinte, um estereótipo de gênero é prejudicial ao limitar a capacidade de mulheres para desenvolverem suas aptidões pessoais, terem uma carreira profissional e tomarem decisões sobre suas vidas e projetos existenciais.

A Lei nº 7.116/1989, tipificando os crimes de racismo, definiu as discriminações resultantes de preconceito de raça, cor, etnia, religião ou procedência nacional, valendo pontuar que constituem objetivos fundamentais da República Federativa do Brasil promover o bem de todos, sem preconceito de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação. Conforme atesta o art. 3º, IV, da Constituição Federal de 1988, qualquer ato discriminatório é violação grave a esse princípio basilar que concretiza a salvaguarda dos direitos humanos.

O desprezo à condição de mulher configura uma forma de violência de gênero e também pode ser traduzida como expressão de racismo, compreendido este em sua dimensão social, o qual ajusta-se, por identidade de razão e mediante adequação típica, aos preceitos primários de incriminação definidos na Lei nº 7.716/1989.

Retomando as bases fenomenológicas do pensamento de MERLEAU-PONTY<sup>49</sup>, o mundo social surge da interação entre os esforços competitivos e cooperativos dos membros da sociedade, que conferem sentido às suas vidas e moldam o tecido de significados estabelecidos por leis, costumes e tradições. Desse modo, compreender verdadeiramente o mundo social significa entender a intersubjetividade que o antecede e o torna viável.

A intersubjetividade é uma complexa rede de relações entre diferentes subjetividades.

A fenomenologia de MERLEAU-PONTY está enraizada na consciência, manifestando-se a subjetividade na experiência do próprio corpo, que é apresentada à consciência. A relação com o outro extrapola a simples posição de objeto, transformando-se em comunicação com um modo de ser.

O filósofo descreve o sujeito como um “corpo-consciência”, um “cogito encarnado” ou “corpo-vivido”. Sua visão sobre a subjetividade, nesse sentido, aponta para a experiência sensorial e comportamental, revelando uma série de dualidades como mente e corpo, sujeito e objeto, percebedor e percebido, interno e externo. Nesse passo, ao tratar das vulnerabilidades sociais, é essencial desenvolver uma compreensão detalhada baseada nas experiências cotidianas de indivíduos, grupos e instituições.

---

49MERLEAU-PONTY. *Op. cit.*

A vulnerabilidade e a situação de risco se apresentam como aspectos intrínsecos à existência humana, entrelaçando-se de forma inseparável com a realidade vivida, e dizem respeito à fragilização da subsistência humana, ligada a predisposições e suscetibilidades a consequências negativas das condições sociais.

## CAPÍTULO 2 – PROCESSOS DE ELABORAÇÃO, DESENVOLVIMENTO E TREINAMENTO DE IA

A ideia de uma máquina que “pensa” nasceu com o matemático e criptógrafo Alan Turing, em 1950, que pretendia descobrir se um computador conseguiria demonstrar a mesma inteligência de uma pessoa. No entanto, o termo “Inteligência Artificial” apareceu apenas em 1956 com John McCarthy, quando o cientista estadunidense inventou a linguagem de programação Lisp<sup>50</sup> – que abriria as portas para o desenvolvimento da IA. Inteligência Artificial refere-se à capacidade de um sistema ou máquina de imitar ou simular a inteligência humana. É um campo multidisciplinar da ciência da computação que se concentra no desenvolvimento de algoritmos e técnicas que permitem que os sistemas aprendam, raciocinem, tomem decisões e ajam de forma autônoma. Seu desenvolvimento ocorre por meio da elaboração de modelos matemáticos e da utilização de técnicas de aprendizado de máquina, permitindo a adaptação desses sistemas a novos dados e cenários. No entanto, à medida que os algoritmos se tornam mais sofisticados, desafios éticos e técnicos emergem, especialmente no que tange à transparência, ao controle e às consequências do uso dessas tecnologias. Embora a lógica procedimental tenha sido amplamente utilizada na formulação inicial da IA, sua limitação em lidar com problemas complexos e contextos de incerteza levou ao avanço das abordagens baseadas em aprendizado de máquina. Contudo, os sistemas algoritmos não estão imunes a distorções cognitivas e vieses, uma vez que são treinados com dados que refletem padrões preexistentes na sociedade.

Considerando que modelos de IA generativa são frequentemente complexos e baseados em redes neurais profundas<sup>51</sup>, compreender exatamente como esse sistema chega a determinada saída pode ser uma tarefa difícil, o que levanta preocupações sobre a falta de transparência nos processos de tomada de decisão das IAs generativas, podendo dificultar a

---

<sup>50</sup>Lisp é uma linguagem dinâmica, cujos programas são constituídos por pequenos módulos, de funcionalidade genérica e que cumprem um objectivo muito simples. Seu autor mostrou que era possível usar exclusivamente funções matemáticas como estruturas de dados elementares. A linguagem Lisp foi projetada primariamente para o processamento de dados simbólicos.

<sup>51</sup>Redes Neurais Profundas, também conhecidas como *Deep Learning*, são um tipo de arquitetura de rede neural artificial composta por múltiplas camadas de neurônios artificiais. Essas redes são chamadas de “profundas” porque têm muitas camadas intermediárias (também chamadas de camadas ocultas) entre a camada de entrada e a camada de saída. Cada camada em uma rede neural profunda consiste em um conjunto de neurônios, que são unidades computacionais básicas que realizam cálculos sobre os dados de entrada. Essas camadas são conectadas entre si por meio de conexões ponderadas, e cada neurônio em uma camada recebe entradas das saídas da camada anterior, realiza cálculos com essas entradas e passa o resultado para a próxima camada. (In AMAZON WEB SERVICES (AWS). “O que é uma rede neural?”. Disponível em: <https://aws.amazon.com/pt/what-is/neural-network/>. Acesso em 8 mar. 2025.)

detecção de possíveis vieses ou discriminações nos resultados gerados. Dessa forma, torna-se essencial compreender os vieses algorítmicos e as heurísticas intuitivas que influenciam a operação desses sistemas, impactando a equidade e a justiça das decisões automatizadas.

A IA é baseada na ideia de construir computadores e sistemas capazes de executar tarefas que normalmente requerem inteligência humana, como reconhecimento de fala, visão computacional, processamento de linguagem natural, tomada de decisões, resolução de problemas e aprendizado.

A lógica procedimental é uma abordagem da Inteligência Artificial que utiliza a lógica simbólica e algoritmos para representar e manipular o conhecimento. Essa abordagem é baseada em regras formais e procedimentos explícitos para realizar inferências lógicas e resolver problemas. Os símbolos podem representar objetos, relações, propriedades, ações ou qualquer outra informação relevante para o problema em questão. As regras de inferência definem como esses símbolos podem ser combinados e manipulados para chegar a conclusões, permitindo que o sistema de IA resolva problemas, faça deduções e tome decisões com base na lógica formal.

Para ilustrar o funcionamento da lógica procedimental, considere o exemplo de um sistema de IA utilizado para determinar se uma pessoa é apta a receber um seguro de saúde. O conhecimento seria representado por símbolos, como “idade”, “doenças preexistentes”, “histórico médico” e “ocupação”. As regras de inferência seriam estabelecidas para combinar esses símbolos e tomar decisões. Uma regra poderia ser definida da seguinte forma: “Se a pessoa tem menos de 30 anos e não possui doenças preexistentes, ela pode ser aprovada para o seguro de saúde.” O mecanismo de inferência, então, verificaria as informações disponíveis sobre a pessoa. Se ela tiver menos de 30 anos e não apresentar doenças preexistentes, o sistema inferiria que ela é apta a receber o seguro. A lógica procedimental também pode tratar de cenários mais complexos ao encadear diversas regras. Suponha que exista uma regra adicional: “Se a pessoa tem mais de 60 anos e fuma, então ela precisa passar por uma avaliação médica antes de ser aprovada para o seguro.” Nesse caso, o mecanismo de inferência verificaria se a pessoa tem mais de 60 anos e se ela fuma. Se ambas as condições forem atendidas, o sistema acionaria a regra para submeter a pessoa à avaliação médica antes de considerar sua elegibilidade para o seguro.

Essa abordagem baseada em regras e procedimentos é muito útil em domínios nos quais o conhecimento é bem definido e estruturado. No entanto, a lógica procedimental pode encontrar

dificuldades quando lidar com problemas mais complexos ou situações de incerteza, nos quais o conhecimento é incompleto, vago ou sujeito a exceções. Por esse motivo, a lógica procedimental muitas vezes é combinada com técnicas de aprendizado de máquina, permitindo que o sistema de IA aprenda a partir de dados e reconheça padrões complexos.

Essa combinação de abordagens, conhecida como “abordagem híbrida”, busca aproveitar as vantagens da lógica procedimental em termos de raciocínio lógico explícito, somado às capacidades de aprendizado e generalização das técnicas de aprendizado de máquina. Uma abordagem conhecida como “lógica indutiva” combina a lógica simbólica com técnicas de aprendizado de máquina para inferir conhecimento a partir de dados.

Segundo KOWALSKI<sup>52</sup>, a definição mais básica de algoritmos é que eles são “lógica + controle”. Eles são simultaneamente um conjunto de instruções abstratas (lógica) e possibilidades de ação (controle).

Sistemas de Inteligência Artificial baseados em lógica simbólica e regras de inferência são amplamente aplicados em diagnósticos médicos, planejamento de rotas, sistemas especialistas e processamento de linguagem natural. Entretanto, a IA contemporânea muitas vezes combina abordagens diferentes, aproveitando tanto a lógica procedimental quanto as técnicas de *machine learning*, mais flexíveis, adaptáveis e capazes de lidar com problemas complexos e incertos.

Esses sistemas já estão constantemente reproduzindo viés na produção sem que os os cientistas saibam como. Um desses exemplos foi descoberto em um estudo pioneiro de 2018 intitulado *Gender Shades*<sup>53</sup>. Nele, os pesquisadores descobriram que os sistemas de reconhecimento facial populares detectaram com mais precisão os homens com pele mais clara e tiveram os maiores erros ao detectar as mulheres com pele mais escura. Também podem “adivinhar” a raça de uma pessoa com base em raios-X e tomografias computadorizadas, mas os cientistas não têm ideia do porquê ou como isso acontece.

As decisões mediadas por sistemas computacionais representam uma versão imperfeita da realidade, pois derivam de juízos humanos que, inevitavelmente, influenciam a forma como o mundo é percebido e categorizado. Quando tais decisões são interpretadas de

---

52KOWALSKI, R., “Algorithm = Logic + Control”, **Communications of the ACM**, 22(7), 1979, pp. 424-436. Disponível em: <https://www.doc.ic.ac.uk/~rak/papers/algorithm%20=%20logic%20+%20control.pdf>. Acesso em: 8 mar. 2025.

53BUOLAMWINI, Joy; GEBRU, Timnit. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. **Proceedings of Machine Learning Research** 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency. Disponível em: <https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/>. Acesso em: 8 mar. 2025.

maneira acrítica, sem a devida consideração dos vieses subjacentes aos algoritmos, corre-se o risco de consolidar dinâmicas discriminatórias e reforçar desigualdades preexistentes.

O termo viés (ou *bias*, em inglês) refere-se a distorções sistemáticas na coleta, análise, interpretação e apresentação de dados, frequentemente influenciadas por preconceitos, opiniões pessoais ou outros fatores que resultam em uma representação parcial da realidade<sup>54</sup>.

Um exemplo comum é o viés de confirmação, um padrão cognitivo no qual as pessoas tendem a buscar, interpretar e lembrar informações de forma a confirmar suas próprias crenças, opiniões ou expectativas anteriores. Isso significa que os indivíduos têm uma tendência a favorecer informações que sustentam suas perspectivas, enquanto minimizam ou ignoram aquelas que contradizem suas visões.

Um dos principais obstáculos encontrados nas avaliações de Inteligência Artificial é a chamada “opacidade algorítmica”. Em algumas formas de IA, os desenvolvedores têm controle total sobre como os algoritmos produzem seus resultados, como é o caso de sistemas de conhecimento simples. Porém, em situações em que há menos intervenção humana no funcionamento da IA, como em aprendizado de máquina (*machine learning*) não-supervisionado ou redes neurais (formas mais complexas de IA), compreender os motivos por trás dos resultados torna-se mais complexo devido à falta de transparência dos parâmetros internos utilizados. Por essa razão, os algoritmos podem ser considerados “opacos”.

Essa opacidade constitui um dos maiores desafios para o desenvolvimento e uso ético da Inteligência Artificial, pois dificulta a compreensão dos fundamentos por trás de resultados específicos, incluindo potenciais casos de discriminação. Conseqüentemente, torna-se mais complexa a tarefa de explicar as decisões tomadas por algoritmos.

Dada a importância da explicabilidade, especialmente em relação à legislação de proteção de dados atual, uma tendência é o desenvolvimento de IA explicável<sup>55</sup>.

No entanto, nem sempre a IA em questão será do tipo explicável; geralmente ela será mais ou menos opaca. Portanto, é essencial que a metodologia utilizada seja aplicável em diferentes níveis de opacidade algorítmica.

---

54SPENCER, E.A.; HENEGHAN C.; “Confirmation bias”. **Catalogue Of Bias** 2018. Disponível em: [www.catalogueofbiases.org/biases/confirmationbias](http://www.catalogueofbiases.org/biases/confirmationbias). Acesso em: 8 mar. 2025.

55NUNES, José Coelho D.; MORATO DE ANDRADE, O. “O Uso da Inteligência Artificial Explicável enquanto ferramenta para compreender decisões automatizadas: Possível Caminho para aumentar a legitimidade e confiabilidade dos modelos algorítmicos?”. **Revista Eletrônica do Curso de Direito da UFSM**, [S. l.], v. 18, n. 1, p. e69329, 2023. DOI: 10.5902/1981369469329. Disponível em: <http://periodicos.ufsm.br/revistadireito/article/view/69329>. Acesso em: 8 mar. 2025.

Algoritmos de aprendizado de máquina demandam retorno (*feedback*), algo que os diga que se desviaram do caminho, e os desenvolvedores usam erros para treinar seus modelos e fazê-los mais inteligentes. Caso fique claro que sistemas automatizados estão errando de modo vergonhoso e sistemático, os programadores voltarão e ajustarão os algoritmos, mas, se, por meio de uma correlação defeituosa através de um algoritmo de avaliação de desempenho, este algoritmo passasse a fazer avaliações enviesadas, sem *feedback* (retorno) humano o sistema continuaria a realizar avaliações prejudiciais, sem nunca aprender com seus erros.

Quando se cria um modelo, faz-se escolhas sobre quais elementos merecem ser integrados, tornando mais clara a complexidade do mundo em uma representação mais acessível que seja facilmente compreensível. A partir dessa simplificação pode-se inferir dados e tomar decisões.

Algoritmos de avaliação exercem um papel fundamental na modelagem de comportamentos, analisando grandes volumes de dados e atribuindo pontuações que indicam a compatibilidade de um indivíduo com determinada função. Essas pontuações, baseadas em probabilidades, categorizam pessoas como “aptas” ou “inaptas”, “adequadas” ou “inadequadas”, conforme critérios previamente estabelecidos. No ambiente corporativo, a adoção de um sistema algorítmico para avaliar o desempenho dos funcionários pode estabelecer parâmetros para promoções e aumentos salariais, incentivando a competitividade e a adaptação às exigências organizacionais. Dessa forma, os trabalhadores são estimulados a ajustar suas condutas para corresponder às métricas estabelecidas, seja para alcançar melhores oportunidades, seja para evitar prejuízos em sua posição profissional.

Contudo, se considerarmos frequência, pontualidade e produtividade como alguns dos critérios adotados por esse sistema de avaliação, as mulheres seriam, certamente, as mais prejudicadas, pela sobrecarga com as atividades domésticas e de cuidados com os filhos<sup>56</sup>.

O ponto nodal da questão está em que os desenvolvedores desses algoritmos, a partir de sua própria concepção de valores e visão de mundo, poderão definir nos algoritmos que criarem sua própria realidade, usando-a para justificar seus resultados. Esse tipo de modelo se autoperpetua e é altamente nocivo, sendo um exemplo de um ciclo de *feedback*, ou retroalimentação.

---

56SALVAGNI, Julice; *et al.* “Maternidade e Mercado de Trabalho: A trajetória das mulheres no desenvolvimento de carreiras”. **Confluências Revista Interdisciplinar de Sociologia e Direito**, Volume 25, Número 1, abril de 2023, pp. 18-42. Disponível em: <https://periodicos.uff.br/confluencias/article/view/57084>. Acesso em: 8 mar. 2025.

Realizar ações – seja uma empresa comercializando produtos, um indivíduo adquirindo uma casa ou uma estratégia de desenvolvimento governamental – envolve certas competências, tais como consciência estratégica, planejamento e a capacidade de influenciar pessoas através de persuasão. Para que essas ações sejam eficazes, é necessário que os objetivos das partes envolvidas estejam devidamente alinhados, garantindo que os esforços realizados conduzam ao resultado desejado.

O termo “alinhado” é frequentemente usado para se referir aos objetivos de um sistema, no sentido de que os objetivos de uma IA estão alinhados se produzirem as mesmas ações da IA que ocorreriam se a IA compartilhasse os objetivos de alguma outra entidade (por exemplo, seu usuário ou operador).

O conceito de “alinhamento” da Inteligência Artificial possui diferentes definições na literatura, variando conforme a relação entre os objetivos da IA e os interesses humanos. Uma IA pode ser considerada alinhada quando suas decisões maximizam benefícios para um agente principal, seguem as intenções humanas ou evitam impactos negativos. A IA “alinhada à intenção” busca atender aos objetivos do operador, enquanto a IA “alinhada ao impacto” previne ações prejudiciais. O alinhamento pleno ocorre quando não há comportamentos indesejáveis decorrentes de falhas nos objetivos da IA.

No entanto, alguns sistemas podem, por padrão, apresentar desalinhamento, desenvolvendo planos que colocam em risco a capacidade da humanidade de influenciar o mundo, mesmo sem que essa perda de controle seja intencional.

Problemas de desalinhamento em sistemas robóticos podem surgir devido a diversas dificuldades na especificação de suas tarefas. Retornos ambíguos podem levar a interpretações equivocadas das instruções humanas, resultando em comportamentos inesperados. A segurança pode ser comprometida quando a busca por eficiência entra em conflito com a integridade física das pessoas. Ambientes dinâmicos podem tornar as diretrizes iniciais inadequadas, assim como o viés nos dados de treinamento pode limitar a capacidade do robô de lidar com situações reais.

Diferenças na interpretação humana e robótica podem gerar falhas na comunicação, enquanto a rigidez na adaptação a novos contextos prejudica a flexibilidade operacional. Além disso, sistemas que aprendem continuamente podem desenvolver comportamentos que se afastam gradualmente das intenções originais dos humanos.<sup>57</sup>

---

<sup>57</sup>CARLSMITH, Joseph. “Is Power-Seeking AI an Existential Risk?”. *Open Philanthropy*, 2022, pp. 22-25. Disponível em: <https://arxiv.org/pdf/2206.13353>. Acesso em: 8 mar. 2025.

Jogos de especificação são cenários nos quais os engenheiros e técnicos tentam ensinar aos robôs comportamentos desejados, muitas vezes usando retorno (*feedback*) ou recompensas. No entanto, os problemas de desalinhamento podem ocorrer quando há uma diferença entre o que os humanos querem que o robô faça e o que o robô realmente faz, o que pode acarretar resultados indesejados ou até mesmo perigosos.

A Google DeepMind documentou exemplos de jogos de especificação<sup>58</sup>: uma IA indo de acordo com sua função de recompensa especificada (que codifica nossas intenções para o sistema), mas não fazendo o que os pesquisadores pretendiam.

Em um exemplo, um braço de robô foi solicitado a segurar uma bola, mas a recompensa foi especificada em termos de se os humanos achavam que o robô tinha sido bem-sucedido. Como resultado, o braço aprendeu a pairar entre a bola e a câmera, enganando os humanos e fazendo-os pensar que havia agarrado a bola.

Um caso que ilustra como um sistema de IA pode ser influenciado negativamente pelo comportamento das pessoas com as quais interage ocorreu em 2016, quando a Microsoft lançou um *chatbot* chamado Tay no Twitter que rapidamente se tornou racista e ofensivo após interações com outros usuários<sup>59</sup>.

A Tesla, fabricante de veículos elétricos, tem enfrentado incidentes em que seu sistema de assistência ao motorista, o Autopilot, foi mal compreendido ou mal utilizado, resultando em acidentes. Desde 2019, foram 736 acidentes envolvendo esses veículos nos Estados Unidos<sup>60</sup>.

Um estudante do estado de Michigan, nos Estados Unidos, relatou que o *chatbot* Gemini, da Google, enviou uma mensagem de teor ameaçador. O incidente ocorreu após o estudante questionar o *chatbot* sobre os desafios enfrentados por adultos mais velhos na administração de seus rendimentos após a aposentadoria.

A resposta gerada pela IA continha frases ofensivas e hostis dirigidas diretamente ao

---

58KRAKOVNA, Victoria; UESATO, Jonathan; MIKULIK, Vladimir; RAHTZ, Matthew; EVERITT, Tom; KUMAR, Ramana; KENTON, Zac; LEIKE, Jan; LEGG, Shane. “Specification gaming: the flip side of AI ingenuity”. **DeepMind Safety Research**: 2020. Disponível em:

<https://deepmindsafetyresearch.medium.com/specification-gaming-the-flip-side-of-ai-ingenuity-c85bdb0deeb4>.

Acesso em: 8 mar. 2025.

59MÜLLER, Léo. “Tay: Twitter conseguiu corromper a IA da Microsoft em menos de 24 horas”. **TECMUNDO**. Publicado em 24/03/2016. Disponível em:

<https://www.tecmundo.com.br/inteligencia-artificial/102782-tay-twitter-conseguiu-corromper-ia-microsoft-24-horas.htm>. Acesso em: 8 mar. 2025.

60VENTURA, Layse. “Número de acidentes com piloto-automático da Tesla é maior do que se sabia”. **OLHAR DIGITAL**, Publicado em 10/06/2023. Disponível em: <https://olhardigital.com.br/2023/06/10/carros-e-tecnologia/numero-de-acidentes-com-piloto-automatado-da-tesla-e-maior-do-que-se-sabia/>. Acesso em: 8 mar. 2025.

usuário<sup>61</sup>.

A resposta ameaçadora e desconexa do Gemini ao estudante sugere falhas graves no processamento de linguagem e nos mecanismos de segurança do sistema, sugerindo erro de contextualização, falta de alinhamento semântico e *bias* no modelo. Portanto, sabemos que é possível criar um sistema de IA desalinhado.

O termo Inteligência Artificial abrange sistemas de duas categorias principais: (a) sistemas baseados em conhecimento, nos quais são estabelecidas e mantidas regras conhecidas de uma determinada área de conhecimento, e (b) sistemas de aprendizado, nos quais os algoritmos são capazes de se ajustar com base em seus próprios resultados, visando otimizar seu desempenho por meio da experiência adquirida.

Para utilizar um corpo de conhecimento em uma máquina, é necessário definir uma maneira de representá-lo. Dentro de qualquer programa de computador, está embutido o conhecimento relacionado a um problema específico que precisa ser resolvido. O conhecimento, então, se manifesta nos algoritmos que o programa utiliza e nas regras de decisão que determinam qual algoritmo aplicar em circunstâncias particulares.

Quando um programa é carregado em um computador, pode-se dizer que o computador “adquire” esse conhecimento, embora na maioria dos programas, essas informações não sejam representadas explicitamente e, portanto, não possam ser facilmente atualizadas ou manipuladas. Assim, na IA, o termo “conhecimento” refere-se à informação necessária para que um programa de computador seja capaz de agir de maneira inteligente.

Uma das características fundamentais dos programas de Inteligência Artificial é sua estrutura que separa claramente o código executável dos dados ou conhecimento do sistema. É exatamente por esse motivo que os algoritmos se tornam opacos.

A opacidade dos algoritmos refere-se à dificuldade de compreender como eles funcionam internamente, especialmente em sistemas de Inteligência Artificial. Essa característica está relacionada à separação entre o código do programa, que contém as instruções que ele segue, e os dados, que são as informações que ele processa.

Como dissemos, o conhecimento em IA diz respeito às informações utilizadas pelo programa para executar tarefas específicas, mas essas informações não são diretamente visíveis ou acessíveis no código. Em vez disso, os algoritmos operam de forma que seus processos

---

<sup>61</sup>MOZELLI, Rodrigo. “IA do mal? Gemini ofende usuário e pede que ele morra”. **OLHAR DIGITAL**, publicado em 20/11/2024. Disponível em: <https://olhardigital.com.br/2024/11/20/pro/ia-do-mal-gemini-ofende-usuario-e-pede-que-ele-morra/>. Acesso em 12 mar. 2025.

internos e as regras que determinam as decisões tomadas não são explicitamente apresentadas.

Nos modelos mais complexos, como os que utilizam aprendizado profundo (*deep learning*), o algoritmo não segue regras pré-programadas; ele aprende a partir dos dados fornecidos, ajustando automaticamente seus parâmetros para melhorar o desempenho, o que resulta em um sistema cujo caminho até a decisão final não é transparente, criando o que se chama de “caixa-preta”.

Essa expressão descreve sistemas cuja lógica interna é desconhecida ou muito difícil de compreender, já que, após o treinamento, o algoritmo realiza suas operações sem que seja possível explicar detalhadamente como cada decisão foi tomada.

A opacidade algorítmica representa uma questão importante, sobretudo em áreas como medicina, direito ou segurança, onde é essencial compreender como e por que uma decisão foi tomada. Assim, a dificuldade em rastrear os processos internos de algoritmos complexos deriva da forma como eles aprendem e ajustam suas operações de maneira independente, o que compromete a transparência e a explicabilidade do sistema.

O funcionamento das inteligências artificiais é baseado em algoritmos complexos de aprendizado de máquina, como redes neurais convolucionais (CNNs)<sup>62</sup> ou redes adversárias generativas (GANs)<sup>63</sup>.

Esses algoritmos são treinados com grandes volumes de dados para reconhecer padrões e características presentes nesses conjuntos. Após o treinamento, as IAs generativas, como as GANs, são capazes de gerar novas amostras que seguem a distribuição dos dados originais, criando variações que se assemelham aos dados usados no treinamento.

Sistemas de Inteligência Artificial (IA) podem ser classificados de diferentes formas, dependendo de critérios como capacidades, tipos de aprendizado e aplicações<sup>64</sup>:

---

62Uma Rede Neural Convolucional (ConvNet/Convolutional Neural Network/CNN) é um algoritmo de aprendizado profundo que pode captar uma imagem de entrada, atribuir importância (pesos e vieses que podem ser aprendidos) a vários aspectos/objetos da imagem e ser capaz de diferenciar um do outro. O pré-processamento exigido em uma ConvNet é muito menor em comparação com outros algoritmos de classificação. A arquitetura de uma ConvNet é análoga àquela do padrão de conectividade de neurônios no cérebro humano e foi inspirada na organização do Visual Cortex. Os neurônios individuais respondem a estímulos apenas em uma região restrita do campo visual conhecida como Campo Receptivo. (In DEEP LEARNING BOOK. “Introdução às Redes Neurais Convolucionais” – Capítulo 40. **Data Science Academy**. Disponível em <https://www.deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/>. Acesso em: 8 mar. 2025).

63Redes Adversárias Generativas (GANs) são arquiteturas de redes neurais profundas compostas por duas redes colocadas uma contra a outra (daí o nome “adversárias”). (In DEEP LEARNING BOOK. “Introdução às Redes Neurais Convolucionais” – Capítulo 40. **Data Science Academy**. Disponível em <https://www.deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/>. Acesso em: 8 mar. 2025).

64RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. Tradução de Regina Célia Simille. Rio de Janeiro: Elsevier, 2013, pp. 284-329.

### 1. Classificação pela Capacidade:

- IA Fraca: Focada em tarefas específicas, como reconhecimento de voz ou recomendação de produtos. Não possui consciência ou entendimento além da tarefa programada. Exemplos: Assistentes virtuais (Alexa, Siri), sistemas de tradução automática;
- IA Geral: Capaz de realizar qualquer tarefa intelectual que um humano poderia executar, com raciocínio, aprendizado e adaptação em múltiplos contextos. Atualmente, é mais uma meta do que uma realidade prática;
- Superinteligência: IA hipotética que supera a inteligência humana em todos os aspectos, incluindo criatividade e habilidades de resolução de problemas. Ainda está no campo da especulação.

### 2. Classificação pelo Tipo de Aprendizado:

- Aprendizado Supervisionado: A IA é treinada com dados rotulados (*input* e *output* conhecidos), permitindo prever saídas para novos dados. Exemplos: Classificadores de *spam*, detecção de fraudes;
- Aprendizado Não-Supervisionado: Trabalha com dados sem rótulos, identificando padrões ocultos ou agrupamentos. Exemplos: Análise de comportamento de clientes, algoritmos de clusterização<sup>65</sup>;
- Aprendizado por Reforço: Baseia-se em recompensas e punições para aprender a tomar decisões ótimas em um ambiente. Exemplos: Robótica, jogos (AlphaGo, DeepMind).

### 3. Classificação pela Arquitetura ou Tecnologia:

- Sistemas Baseados em Regras: Operam com base em conjuntos de regras predefinidas, geralmente utilizadas em sistemas especialistas. Exemplo: Sistemas de diagnóstico médico baseados em regras;
- Redes Neurais Artificiais: Modelos inspirados no cérebro humano que processam informações em camadas. Exemplo: Reconhecimento facial;
- Modelos de IA Generativa: Sistemas capazes de criar novos conteúdos, como textos, imagens ou música. Exemplos: GPT, DALL-E.

### 4. Classificação pela Aplicação:

---

<sup>65</sup>Algoritmos de clusterização são técnicas de aprendizado de máquina não supervisionado utilizadas para agrupar dados em conjuntos ou *clusters*, de maneira que os dados dentro de cada *cluster* sejam mais semelhantes entre si do que com os dados de outros *clusters*. O objetivo é identificar estruturas ou padrões nos dados sem a necessidade de rótulos ou categorias predefinidas. Esses algoritmos ajudam a segmentar grandes volumes de dados de maneira eficiente e são amplamente utilizados em várias áreas, como análise de mercado, segmentação de clientes, compressão de imagens, biologia computacional, entre outras.

- IA Conversacional: Projetada para interagir com humanos por meio de linguagem natural. Exemplos: *Chatbots*, assistentes virtuais;
- IA Preditiva: Focada em prever resultados futuros com base em dados históricos. Exemplos: Previsão do tempo, análise de mercado;
- IA de Percepção: Envolve habilidades sensoriais, como visão e audição. Exemplos: Reconhecimento de imagens, sistemas de vigilância;

#### 5. Classificação pela Dominação Cognitiva:

- IA Reativa: Toma decisões com base no estado atual, sem armazenar experiências passadas. Exemplo: Deep Blue, o sistema que venceu Garry Kasparov<sup>66</sup> no xadrez;
- IA com Memória Limitada: Considera dados históricos para tomar decisões, mas não os armazena permanentemente. Exemplo: Veículos autônomos;
- IA com Teoria da Mente (Teórica): Planejada para compreender emoções humanas e interagir socialmente. Exemplo: Protótipos de IA social;
- IA Autoconsciente (Hipotética): Capaz de entender sua própria existência e tomar decisões com base nessa percepção.

A fim de prover uma compreensão elementar sobre o processo de aprendizado de máquina, segundo FELTRIN<sup>67</sup>, a criação dos modelos de algoritmos para análise e interpretação dos dados são baseados em três tipos diferentes e independentes:

- Aprendizado Supervisionado: Consistente na introdução manual de dados de entrada e suas respectivas saídas durante o treinamento de uma rede neural artificial. Neste processo, a máquina é guiada a reconhecer e aprender os padrões que resultam em uma saída específica. Em essência, trata-se de um modelo em que se dispõe das entradas e das soluções para os problemas, e o treinamento da rede é configurado para ensinar como alcançar a resposta correta. Dessa forma, a rede aprende o padrão a ser aplicado em situações futuras com novos dados;
- Aprendizado Não-Supervisionado: Neste método, a rede neural processa dados acumulados contendo diversas informações sobre as amostras. Sem conhecimento prévio sobre as respostas ou classificações corretas, a rede identifica e organiza os dados com

---

<sup>66</sup>Kasparov se tornou campeão mundial de xadrez em 1985, aos 22 anos, ao derrotar Anatoly Karpov, e manteve o título até 2000, e também é reconhecido pelo seu confronto histórico com o supercomputador de xadrez Deep Blue, da IBM. Em 1997, após uma série de partidas, Deep Blue venceu Kasparov, o que foi um marco na história da Inteligência Artificial e da computação.

<sup>67</sup>FELTRIN, Fernando (aut.). **Ciência de Dados e Aprendizado de Máquina**: Uma Introdução às Redes Neurais Artificiais. São Paulo-SP: Amazon, 2019. E-book (296p.) (eBook Kindle). Disponível em: <https://a.co/d/8RSfHer>. Acesso em: 8 mar. 2025.

- base em similaridades, padrões ou categorias identificadas de forma autônoma. O sistema reconhece características que diferenciam ou conectam as amostras, agrupando-as de acordo com suas semelhanças. Essa abordagem se limita à classificação, sem tomar decisões adicionais além de categorizar novas amostras conforme os padrões reconhecidos;
- **Aprendizado por Reforço:** Nesse método, a rede neural opera a partir de seu estado atual e do ambiente ao redor, buscando alcançar um objetivo específico, como encontrar o caminho mais eficiente entre dois pontos. Por meio de inúmeras tentativas e erros, a rede ajusta suas ações até atingir o objetivo definido. Durante esse processo, o sistema aprende ao identificar seus acertos e erros, sendo recompensado por escolhas corretas e penalizado por escolhas inadequadas. Essas recompensas e penalidades auxiliam a rede a distinguir ações eficientes das ineficientes, armazenando os padrões aprendidos em uma memória de experiência para aplicá-los em situações futuras.

As redes neurais, também denominadas como redes neurais artificiais (ANNs, na sigla em inglês) ou redes neurais simuladas (SNNs), constituem um elemento essencial no domínio do aprendizado de máquina, ocupando uma posição de destaque no desenvolvimento e aplicação de algoritmos de aprendizado profundo (*deep learning*). Elas foram concebidas a partir da inspiração no funcionamento do cérebro humano, buscando imitar como os neurônios biológicos se comunicam uns com os outros<sup>68</sup>.

O aprendizado profundo acontece por meio da construção de modelos de redes neurais, que processam múltiplas entradas, cada uma associada a pesos específicos, atravessando diversas camadas intermediárias, em um processo que envolve fases que podem ser supervisionadas ou não, além de etapas de propagação direta (*feedforward*<sup>69</sup>) e retropropagação (*backpropagation*<sup>70</sup>).

Redes neurais artificiais (ANNs) consistem em camadas de unidades chamadas “nós”, que incluem uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada “nó”, também denominado “neurônio artificial”, está interconectado com outros “nós” e possui atributos como peso e limiar. Quando a saída de um “nó” individual excede um valor

---

68IBM. “O que são redes neurais”. Disponível em: <https://www.ibm.com/br-pt/topics/neural-networks>. Acesso em: 8 mar. 2025.

69*Feedforward* pode ser traduzido como “propagação direta” ou “alimentação direta”. Refere-se ao fluxo de dados que ocorre na rede neural, onde os dados são transmitidos das camadas de entrada para as camadas de saída sem retroalimentação. (In FELTRIN. *Op. cit.*)

70*Backpropagation* pode ser traduzido como “retropropagação” ou “retropropagação do erro”. Envolve o cálculo do gradiente do erro em relação aos pesos da rede, começando pelas camadas de saída e retrocedendo até as camadas de entrada. Esse gradiente é usado para ajustar os pesos da rede e melhorar seu desempenho ao longo do tempo. (*Idem, ibidem*)

de limiar predefinido, ele é ativado, transmitindo dados para a próxima camada da rede. Caso contrário, nenhum dado é transferido para a próxima camada. Finalmente, após o processamento, buscam-se ajustes precisos nos parâmetros. O objetivo é permitir a identificação e o reconhecimento de padrões específicos.

Essas redes neurais dependem do uso de dados de treinamento para aprender e melhorar sua precisão ao longo do tempo. No entanto, assim que esses algoritmos de aprendizado são ajustados para aumentar a precisão, eles se transformam em ferramentas poderosas no campo da ciência da computação e Inteligência Artificial, propiciando a classificação e o agrupamento de dados a uma velocidade assombrosa, tornando tarefas como reconhecimento de fala ou identificação de imagens muito mais eficientes em comparação com a identificação manual realizada por especialistas humanos. Um exemplo de rede neural é o algoritmo de busca do Google.

Na literatura, esse conceito é frequentemente mencionado como “redes neurais densas” ou “redes neurais profundas”, referindo-se à representação visual dessas redes, que contêm um maior número de “neurônios” e camadas, ampliando sua capacidade de análise.

Como ilustração desse processo, imagine que estamos ensinando um computador a reconhecer sons de animais, como o latido de um cachorro.

Primeiro, criamos algo semelhante a um cérebro artificial, chamado “rede neural”. Pense nisso como um complexo de várias unidades, cada uma com uma função específica. Em seguida, tocamos muitos sons de latidos de cachorro para essa “máquina cerebral”. Ela analisa aspectos importantes dos sons, como frequência e ritmo, e tenta entender como esses elementos se combinam para formar o som de um latido.

À medida que a máquina escuta mais gravações e pratica ela começa a identificar o som de um latido mais precisamente. Às vezes, informa-se se ela acertou ou errou, para que possa aprender e melhorar mais rapidamente. Aqui entram os dois conceitos mencionados: “propagação direta” (*feedforward*) é como a máquina aprende captando os sons, e “retropropagação do erro” (*backpropagation*) é como ela ajusta suas ideias com base nos erros que comete. Depois que a máquina aprende bastante, é testada para ver se consegue reconhecer latidos de cães que ela nunca ouviu. Se ela cometer erros, ajusta-se o que ela aprendeu para melhorar.

## 2.1 VIESES ALGORÍTMICOS E HEURÍSTICAS INTUITIVAS

Vieses algorítmicos podem ser caracterizados como padrões sistemáticos e não intencionais que se manifestam nos algoritmos, os quais podem levar a resultados injustos, discriminatórios, ou imprecisos.<sup>71</sup>

Dentre os tipos identificados, alguns se destacam, incluindo:

- Vieses de dados: Algoritmos frequentemente aprendem a partir de dados históricos, podendo perpetuar preconceitos presentes nesses dados, como os de natureza racial, de gênero, socioeconômica, entre outros;
- Vieses de seleção de amostras: Se os dados utilizados para treinar um algoritmo não forem representativos da população-alvo, isso pode distorcer os resultados obtidos;
- Vieses de atribuição: Algoritmos podem incorretamente atribuir causas a determinados efeitos, levando a conclusões equivocadas ou injustas;
- Vieses de modelagem: A escolha das características ou variáveis incluídas em um modelo de aprendizado de máquina pode introduzir vieses, especialmente se características relevantes forem omitidas ou características irrelevantes ou discriminatórias forem incluídas;
- Vieses de otimização: Algoritmos de otimização utilizados para treinar modelos podem priorizar resultados que não são justos, como maximizar o lucro em detrimento da equidade;
- Vieses de retroalimentação: Sistemas de recomendação ou classificação podem criar bolhas de filtro ao maximizar o envolvimento do usuário, expondo as pessoas apenas a informações que confirmam suas crenças pré-existentes;
- Vieses contextuais: Algoritmos podem tomar decisões distintas com base em informações contextuais, podendo resultar em discriminação se o contexto incluir informações sensíveis;
- Vieses de saída: Respostas geradas por algoritmos, como *chatbots* ou sistemas de reconhecimento de fala, podem conter vieses se o treinamento dos modelos não equitativo ou se os dados de treinamento contiverem vieses;

---

71COZMAN, Fabio Gagliardi; KAUFMAN, Dora. “Viés no aprendizado de máquina em sistemas de Inteligência Artificial: a diversidade de origens e os caminhos de mitigação”. **Revista USP**. São Paulo, n. 135, p. 195-210, outubro/novembro/dezembro 2022, pp. 195-210.

- Vieses de reforço: Algoritmos de aprendizado por reforço podem aprender a tomar decisões tendenciosas quando recompensados de maneira enviesada;
- Vieses de *feedback* humano: Se algoritmos interagem com humanos e recebem retorno enviesado, isso pode criar um ciclo de retroalimentação viciosa, onde os vieses são aprendidos e perpetuados.

Heurística é uma abordagem ou técnica que visa encontrar soluções aproximadas ou sub ótimas para problemas complexos, especialmente quando não é possível ou prático encontrar soluções exatas em um tempo razoável. Em outras palavras, uma heurística é uma estratégia geralmente baseada em experiência, intuição ou regras práticas que guiam a busca por soluções aceitáveis, mesmo que não garantam a melhor solução possível. São frequentemente usadas em problemas de otimização, onde o objetivo é encontrar a melhor solução possível entre um conjunto de alternativas. Embora as soluções obtidas por meio de heurísticas possam não ser ótimas, elas são valiosas quando o tempo ou os recursos são limitados, ou quando o problema é intratável e uma solução exata é impraticável.

Exemplos comuns de heurísticas incluem algoritmos genéticos, busca local, “busca gulosa”<sup>72</sup> e algoritmos de vizinhança variável.

Essas técnicas são amplamente utilizadas em uma variedade de campos, incluindo Inteligência Artificial, ciência da computação, engenharia, economia e ciências sociais, para resolver problemas complexos de forma eficiente. Ao proporem soluções aproximadas para problemas complexos, devem ser aplicadas com extrema cautela. Embora frequentemente eficazes na busca por soluções rápidas, essas técnicas podem ser terreno fértil para a ampliação de vieses se não forem adequadamente moldadas.

Um estudo acadêmico intitulado *Gender Stereotype Reinforcement: Measuring the Gender Bias Conveyed by Ranking Algorithms*<sup>73</sup>, analisou como sistemas de ordenação podem reproduzir preconceitos de gênero por meio de uma métrica chamada “Reforço de Estereótipos de Gênero”, projetada para mensurar a tendência de motores de busca em replicar esses estereótipos. Para isso, os autores do estudo utilizaram informações relacionadas ao gênero codificadas em representações matemáticas conhecidas como *embeddings* (representações vetoriais) de palavras, as quais capturam associações semânticas

---

<sup>72</sup>A busca gulosa é uma técnica de otimização que seleciona a solução localmente ótima em cada etapa, sem considerar as implicações globais.

<sup>73</sup>FABRIS et al. “Gender Stereotype Reinforcement: Measuring the Gender Bias Conveyed by Ranking Algorithms”. **Department of Information Engineering, University of Padua, Italy**. Disponível em: <https://arxiv.org/pdf/2009.01334>. Acesso em: 8 mar. 2025.

extraídas de grandes conjuntos de dados textuais. A análise mostrou que algoritmos de ranqueamento frequentemente amplificam vieses preexistentes, reproduzindo associações estereotipadas entre termos relacionados a gênero e profissões, comportamentos ou características específicas.

Muitos dos vieses algorítmicos que abordamos nesta pesquisa adotam o modelo de valor agregado<sup>74</sup> como forma de tratamento dos dados coletados, revelando-se extremamente danoso, principalmente em sua utilização para algoritmos de reconhecimento facial.

Os modelos estatísticos de valor agregado usados em sistemas de Inteligência Artificial têm uma tendência a perpetuar vieses por causa de suas limitações estruturais e metodológicas. Eles frequentemente dependem de bases de dados históricas que refletem desigualdades sistêmicas, como diferenças salariais e falta de acesso a oportunidades, o que faz com que esses algoritmos reproduzam e até amplifiquem essas disparidades em seus resultados. Além disso, as métricas usadas para medir o desempenho desses modelos geralmente não são ajustadas para garantir neutralidade, incorporando, assim, preconceitos culturais e normativos que ignoram as diferenças contextuais entre grupos de gênero. Isso faz com que os algoritmos valorizem mais características associadas aos homens e subestimem as relacionadas às mulheres, reforçando estereótipos existentes.

Outro problema é a falta de diversidade nas equipes que desenvolvem esses sistemas. Quando as equipes são homogêneas, faltam perspectivas críticas e interdisciplinares que poderiam ajudar a reduzir os vieses. Por outro lado, quando a análise de gênero é feita, muitas vezes ela é tratada de forma isolada, sem considerar fatores como raça, classe social ou deficiência, o que aumenta as desigualdades, especialmente para mulheres de grupos mais vulneráveis. Por fim, a retroalimentação gerada por esses modelos reforça vieses preexistentes ao influenciar etapas subsequentes de coleta e tratamento de dados, instaurando ciclos de discriminação cumulativa que se tornam progressivamente mais complexos de interromper.

Como uma prática humana, a comunicação possibilita a criação de experiências e a interação entre indivíduos, caracterizada pela reflexividade, desempenhando um importante papel na formação do imaginário e na prática cotidiana. Ela também desempenha função determinante na construção de representações estéticas, identidades, restrições e afetos.

---

<sup>74</sup>O modelo de valor agregado é uma estratégia conceitual que explica como uma empresa pode aumentar a percepção de valor de seus produtos ou serviços para os clientes. Em essência, é uma forma de medir o benefício percebido pelos clientes em relação ao custo total do produto ou serviço.

Segundo THOMPSON<sup>75</sup>, ao optar pelas estratégias de linguagem para alcançar objetivos específicos em diversas situações, simultaneamente se produzem três tipos de significados: (1) vinculados às representações de poder e solidariedade, que englobam as atitudes adotadas em relação ao outro e os papéis sociais exercidos; (2) os significados que dizem respeito à representação da experiência por meio da linguagem, que tratam do conteúdo abordado e de quem age sobre algo ou alguém; e (3) os significados associados à organização da mensagem, que se referem tanto ao que está sendo expresso quanto ao que já foi comunicado anteriormente.

O autor sugere que esses três tipos de significado estão intrinsecamente relacionados com cada uma das metafunções da linguagem: interpessoal, ideacional e textual.

WOJCIECHWSKI & DA ROSA<sup>76</sup> analisam a incidência de vieses cognitivos na prática judicial-penal e sua influência sobre o processo decisório, o que bem nos cabe reproduzir neste estudo, uma vez que demonstra o conhecimento, na prática jurídica, da ocorrência de fatores prejudiciais externos aos elementos normativos que resultam em uma representação tendenciosa da realidade.

Assim, neste Título nos dedicaremos ao estudo das ciências cognitivas para focar nas heurísticas intuitivas e os vieses delas decorrentes, conferindo maior destaque àquelas usualmente presentes nas tomadas de decisão. Mais adiante, abordaremos cada uma das heurísticas e vieses apontados pelos autores a fim de contextualizá-los sob a perspectiva de gênero.

1. heurística da disponibilidade (*availability heuristic*);
2. heurística da referência ou da ancoragem (*anchoring*);
3. heurística do afeto (*affect heuristic*);
4. heurística da correlação ilusória (*illusory correlations/magical thinking*);
5. viés egocêntrico (*egocentric bias*) e excesso de confiança (*overconfidence*);
6. viés confirmatório (*confirmation bias*);
7. viés retrospectivo (*hindsight bias*);
8. erro fundamental de atribuição ou viés de correspondência (*fundamental attribution error or correspondence bias*);
9. viés do ator-observador (*actor-observer bias*); e, finalmente, o
10. viés da perspectiva de câmera (*camera perspective bias*)<sup>77</sup>

Os vieses presentes na Inteligência Artificial (IA) podem surgir em diferentes etapas do desenvolvimento, como na coleta de dados, no treinamento dos modelos, na seleção de

<sup>75</sup>THOMPSON, G. **Introducing functional grammar**. London: Arnold, 1996, p. 8.

<sup>76</sup>WOJCIECHWSKI, Paola Bianchi; DA ROSA, Alexandre Morais. **Vieses da Justiça**: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva. 2ª ed. Florianópolis: Emais, 2021.

<sup>77</sup>*Idem, ibidem*, pp. 18-19.

atributos dos algoritmos e na tomada de decisões. Eles não apenas refletem as desigualdades estruturais, mas também contribuem para sua perpetuação e amplificação ao longo do ciclo de desenvolvimento.

Quando alimentados com dados marcados por preconceitos e estereótipos, os sistemas de IA reproduzem padrões desiguais, reforçando disparidades sociais e de gênero.

Conjuntos de dados que retratam majoritariamente homens em funções de autoridade ou liderança levam a respostas que reafirmam tais representações. Da mesma forma, a geração de textos ou imagens que associam mulheres a papéis sexualizados, de cuidadoras ou submissas, e homens a posições de poder e autonomia, consolida visões distorcidas, impactando espaços de trabalho e perpetuando estruturas desiguais.

A reprodução de padrões também se manifesta na reinterpretação do passado. Ao processar dados históricos sem considerar os contextos culturais, sociais e estruturais que os moldaram, os algoritmos replicam e amplificam visões reducionistas, contribuindo para a manutenção de dinâmicas discriminatórias.

### **2.1.1 Heurística da Disponibilidade**

A heurística da disponibilidade explica a tendência de avaliar a probabilidade de um evento com base na facilidade de acesso a informações relacionadas. Ao confiar naquilo que está mais prontamente disponível em nossa mente, ignoramos informações menos evidentes, mas potencialmente mais relevantes. Esse viés, amplamente estudado no comportamento humano, também encontra implicações no desenvolvimento e aplicação de sistemas de inteligência artificial, especialmente em contextos que envolvem aprendizado de máquina e IA generativa, como os utilizados para criar textos, imagens ou vídeos.

A heurística da disponibilidade pode amplificar vieses de gênero, sobretudo se os dados de treinamento refletirem estereótipos ou desigualdades históricas. Ao treinar um modelo de IA com dados que são predominantemente representações masculinas em posições de liderança ou em áreas como ciência e tecnologia, o sistema tende a associar essas funções a um gênero específico. Isso acontece porque a IA, ao aplicar um processo análogo à heurística da disponibilidade, “acessa” com maior frequência imagens e palavras que reforçam essas associações históricas, perpetuando representações preconceituosas. Um exemplo concreto dessa dinâmica pode ser observado em modelos de geração de texto que,

ao descrever personagens em narrativas, frequentemente reproduzem estereótipos de gênero.

Esses modelos podem associar o papel de “cientista” a figuras masculinas e o de “enfermeira” a figuras femininas, por exemplo, simplesmente porque tais associações são predominantes nos dados de treinamento. De tal forma, os sistemas de IA refletem uma visão histórica e socialmente construída, em vez de uma representação neutra ou equitativa da realidade.

Noticiado pelo portal Carta Capital em outubro de 2023<sup>78</sup>, a Deputada Estadual Renata Souza (PSOL-RJ) denunciou nas redes sociais ter sido vítima de “racismo algorítmico” ao usar uma ferramenta de Inteligência Artificial. Segundo ela, ao criar uma arte inspirada nos pôsteres da Disney, se deparou com uma imagem gerada por Inteligência Artificial que a retratava como uma mulher negra segurando uma arma.

A solicitação original era para a criação de uma imagem de uma mulher negra, com cabelo afro, vestindo roupas típicas. No entanto, o resultado gerado apresentou uma representação distorcida, associando um estereótipo de violência a uma mulher negra em um contexto periférico, refletindo preconceitos implícitos nos dados utilizados para o treinamento do sistema.

O exemplo que representa uma mulher negra portando uma arma oferece um panorama claro sobre o impacto da heurística da disponibilidade nos sistemas algorítmicos. Ao processarem informações e gerarem imagens, esses sistemas não se limitam a replicar os dados de maneira imparcial; em vez disso, reorganizam e projetam as informações que são mais frequentes ou acessíveis nos conjuntos de dados utilizados para seu treinamento. Quando algoritmos são alimentados com um repertório histórico de representações de mulheres negras, frequentemente ligadas a estereótipos de violência ou exclusão social, há uma tendência maior de que essas associações preconceituosas sejam reproduzidas.

No caso citado, a imagem de uma mulher negra armada evidencia uma falha algorítmica: um reflexo de estereótipos históricos amplamente disponíveis, que prevalecem sobre representações mais justas, equânimes e equitativas.

As razões para esse fenômeno são variadas. Em primeiro lugar, os dados utilizados no treinamento muitas vezes provêm de fontes que já possuem distorções ou desigualdades estruturais.

---

78CARTA CAPITAL. “Deputada denuncia ‘racismo algorítmico’ após IA gerar imagem com arma em uma favela”. Publ. em 26/10/2023. Disponível em: <https://www.cartacapital.com.br/cartaexpressa/deputada-denuncia-racismo-algoritmico-apos-ia-gerar-imagem-com-arma-em-uma-favela/>. Acesso em: 8 mar. 2025.

Esses dados não apenas perpetuam preconceitos existentes, mas também os ampliam ao serem incorporados nos sistemas de IA. Em segundo lugar, a própria heurística da disponibilidade nos sistemas de Inteligência Artificial privilegia informações mais acessíveis, reforçando estereótipos amplamente representados devido à predominância dessas imagens nos dados. Ademais, a ausência de diversidade nas bases de treinamento torna mais grave o problema.

Ao serem desenvolvidos com dados limitados ou enviesados sobre grupos sociais, como mulheres negras associadas a cenários de criminalidade ou violência, o sistema segue operando com os mesmos parâmetros tendenciosos e restritos. O resultado é uma reprodução desses estereótipos, que são mais acessíveis ou “disponíveis” no histórico do sistema, perpetuando uma visão distorcida e injusta da realidade. Por conseguinte, o que aqui se observa é uma evidência clara de como o racismo algorítmico não é um fenômeno isolado, mas sim um reflexo das falhas estruturais presentes nos próprios dados e nos processos algorítmicos.

A heurística da disponibilidade, ao permitir que o sistema se baseie nas representações mais facilmente acessíveis, acaba reforçando a normalização de estereótipos prejudiciais, colocando em risco a capacidade dos algoritmos de produzir resultados consoantes com os critérios de verdade, justiça e respeito à dignidade humana.

### **2.1.2 Heurística da Referência ou da Ancoragem**

A heurística da ancoragem, ou da referência, representa um fenômeno cognitivo no qual as decisões ou estimativas de um sujeito são influenciadas por informações previamente apresentadas que funcionam como “âncoras”, que orientam e limitam o processo de julgamento subsequente. Esse conceito, bem estabelecido no campo da psicologia cognitiva, adquire peso quando transportado ao contexto dos sistemas de IA.

Quando se busca uma imagem de mulheres em cargos de poder ou liderança, a IA tende a reproduzir as representações mais tradicionais, nas quais as mulheres são associadas, com frequência, a papéis subalternos ou domésticos, uma vez que, historicamente, essas foram as “âncoras” que prevaleceram em grande parte dos dados utilizados para o treinamento. As causas desse fenômeno são claras e encontram-se nas próprias bases de dados sobre as quais a IA se desenvolve.

Ao processar informações oriundas de fontes que refletem, de maneira desigual, as representações de homens e mulheres nas diversas esferas da sociedade, o sistema acaba por “ancorar-se” nesses padrões, resultando em uma reprodução cega dos estereótipos de gênero que já se encontram profundamente enraizados na estrutura social.

A IA, ao ancorar-se em informações falhas ou desiguais, torna-se uma verdadeira armadilha, criando um ciclo vicioso no qual os padrões históricos de desigualdade de gênero são, não apenas repetidos, mas intensificados. E é exatamente por essa razão que se impõe a necessidade de revisão crítica das bases de dados utilizadas para alimentar esses sistemas.

O que está em jogo é a capacidade da Inteligência Artificial de produzir representações livres de preconceitos estruturais que por muito tempo dominaram as narrativas sobre as distinções entre os sexos.

### **2.1.3 Heurística do Afeto**

A heurística do afeto, ao ser aplicada aos sistemas de IA, guarda estreita relação com os fenômenos da misoginia e do machismo, que, ao longo da história, se apresentam como elementos estruturais profundamente arraigados na sociedade.

A misoginia, enquanto expressão do ódio ou desprezo pelas mulheres, e o machismo, como ideologia que sustenta a subordinação feminina e a supremacia masculina, não se limitam ao campo das ações humanas, mas se refletem, de forma insidiosa, no processo comunicacional e nas formas que adotam para reproduzir a realidade social.

Ao longo do tempo, a sociedade construiu e naturalizou uma visão estereotipada da mulher, associando-a, frequentemente, a representações sexualizadas, em papéis subalternos, de fragilidade ou dependência emocional.

Internalizado de forma afetiva pela sociedade, esse preconceito é refletido nos dados que alimentam a IA, de modo que, ao gerar imagens ou conteúdos textuais, o sistema acaba por produzir representações enviesadas, nas quais as mulheres são colocadas em papéis de vulnerabilidade ou de hipersexualização, como se sua existência social e pessoal estivesse atrelada, inescapavelmente, a esses arquétipos negativos. Ademais, o próprio mecanismo de âncora emocional que rege a heurística do afeto é fortemente influenciado por uma sociedade que, historicamente, estabeleceu uma hierarquia de poder entre os sexos.

Esse padrão de representações, marcado pela carga emocional negativa direcionada às mulheres, não apenas reflete, mas reforça as estruturas machistas que ainda dominam diversos campos sociais, como na política, no mercado de trabalho e na própria dinâmica familiar.

Longe de ser neutro ou acidental, tal processo reflete a ideologia machista que, ao longo do tempo, formou tais representações. A IA, sem capacidade para discernir a validade dessas construções, acaba por replicar e ampliar esses preconceitos.

#### **2.1.4 Heurística da Correlação Ilusória**

A heurística da correlação ilusória refere-se ao processo cognitivo de atribuir relações causais ou conexões entre variáveis que, na realidade, são apenas aparentemente relacionadas. Trata-se da tendência humana de perceber padrões ou correlações onde, objetivamente, eles não existem.

Nos sistemas inteligentes, esse fenômeno se manifesta quando a máquina, ao ser alimentada com dados que contêm preconceitos de gênero, associa erroneamente características ou comportamentos que, embora culturalmente interpretados como conectados, não têm relação objetiva entre si. Essa falácia cognitiva, ao ser aplicada em algoritmos, pode gerar consequências prejudiciais, especialmente em relação aos vieses de gênero, que historicamente subalternizam as mulheres e as associam a estereótipos distorcidos.

A correlação ilusória, nesse contexto, leva à percepção equivocada de que certas características femininas são “naturais” ou inevitáveis, perpetuando a ideia de que essas correlações entre atributos de gênero e comportamentos têm uma base real. Ao absorver essas concepções, a máquina não apenas reflete, mas amplifica estereótipos de gênero, criando a falsa impressão de que tais correlações são naturais.

O viés da correlação ilusória reforça uma visão distorcida do papel das mulheres na sociedade, como se sua presença em certos campos ou seu desempenho em determinadas funções fosse explicado por características essenciais e imutáveis, quando, na realidade, essas correlações são construções sociais e culturais que podem ser desconstruídas.

#### **2.1.5 Viés Egocêntrico e Excesso de Confiança**

O viés egocêntrico refere-se à tendência de perceber o mundo a partir de uma perspectiva

centrada no próprio eu, o que distorce as percepções sobre os outros.

Nos sistemas de IA, esse viés se manifesta nos algoritmos quando as escolhas dos desenvolvedores, ao ignorar ou subestimar a pluralidade de vivências e realidades, resultam na criação de modelos que reforçam estereótipos e desigualdades de gênero. Ao refletir a experiência de um sujeito masculino ou de um grupo limitado, esses algoritmos perpetuam a invisibilidade da mulher, seja no mercado de trabalho, seja na representação midiática.

Em uma sociedade marcada por estruturas de poder desigual, as referências masculinas, predominantes e dominantes, moldam, muitas vezes sem reflexão crítica, a maneira como a IA aborda questões de gênero.

O viés egocêntrico não apenas limita a visão da realidade, mas reforça a subordinação feminina, ao retratar as mulheres em papéis restritos, como se tivessem uma única identidade ou fossem capazes de desempenhar apenas funções específicas. Por outro lado, o excesso de confiança também está relacionado à certeza infundada que os desenvolvedores podem ter sobre a neutralidade e imparcialidade dos algoritmos. Quando projetados sem a devida cautela ou reflexão crítica sobre os dados utilizados, esses sistemas podem amplificar vieses estruturais.

Programadores, confiantes na eficácia de suas ferramentas, podem negligenciar os impactos sociais e as distorções geradas pelos sistemas acreditando erroneamente na objetividade e isenção dos algoritmos. Assim, o excesso de confiança contribui para a perpetuação das desigualdades, já que as falhas nos modelos algorítmicos, alimentados por dados enviesados, são frequentemente desconsideradas.

A confiança em tais sistemas deve sempre ser acompanhada pela consciência de que a neutralidade não existe e que é necessária uma vigilância constante para evitar que as distorções e iniquidades do passado sejam disfarçadamente perpetuadas.

### **2.1.6 Viés de Confirmação**

O viés de confirmação é a tendência de valorizar informações que confirmam crenças preexistentes, ignorando ou minimizando as que as contradizem.

Quando alimentados por dados históricos e sociais impregnados de preconceitos, os algoritmos não apenas refletem tais vieses, mas também os reforçam e amplificam, consolidando padrões de comportamento e representação discriminatórios e excludentes. O viés de confirmação exerce influência particularmente insidiosa nos sistemas algorítmicos.

Ao serem treinados com dados que reproduzem estereótipos de gênero, os algoritmos perpetuam crenças previamente estabelecidas sobre mulheres e homens. Este processo é especialmente evidente quando os dados utilizados provêm de fontes históricas e culturais marcadas por desigualdades. No mercado de trabalho e nas representações midiáticas, as mulheres são frequentemente retratadas em papéis sexualizados ou subalternos enquanto os homens são associados a posições de liderança e autoridade. Esses padrões, consolidados ao longo do tempo, constituem a base sobre a qual os algoritmos operam, perpetuando uma visão enviesada e excludente do papel da mulher na sociedade.

Ao naturalizar tais representações, o viés de confirmação gera um ciclo vicioso de retroalimentação: os sistemas de IA, ao sugerirem ou gerarem resultados com base em dados enviesados, não apenas confirmam os estereótipos existentes, mas também os reforçam, ampliando a discriminação algorítmica.

O viés de confirmação também se manifesta na seleção de dados pelas equipes de desenvolvimento que, ao criar modelos de aprendizado de máquina, acabam por escolher informações que corroboram suas próprias visões de mundo.

Essa escolha seletiva reflete uma falta de autocritica e de atenção às múltiplas dimensões do gênero, resultando em sistemas que não corrigem os preconceitos, mas os perpetuam. Com isso, os dados enviesados alimentam os algoritmos, que, por sua vez, reforçam e replicam esses preconceitos, criando um ciclo contínuo de desigualdade e discriminação.

A utilização de algoritmos sem uma revisão crítica dos dados e das suposições subjacentes contribui para a consolidação de modelos que distorcem ativamente as realidades de gênero, com consequências que vão desde a relativização das capacidades das mulheres em determinados setores até a manutenção das disparidades salariais e de poder.

### **2.1.7 Viés Retrospectivo**

O viés retrospectivo, também conhecido como *hindsight bias*, ou efeito ‘eu sabia o tempo todo’, é uma tendência cognitiva que leva as pessoas a acreditarem, após um evento ter ocorrido, que elas já previam ou sabiam o resultado com antecedência, mesmo que isso não fosse verdade. Ele distorce a memória e a percepção, fazendo com que os indivíduos superestimem sua capacidade de previsão e subestimem a incerteza que existia antes do evento

acontecer. Tal tendência pode influenciar a tomada de decisões futuras, pois as pessoas podem se tornar excessivamente confiantes em sua capacidade de prever resultados, subestimando riscos e incertezas. Resumidamente, o viés retrospectivo é uma distorção que nos faz enxergar o passado como mais previsível do que realmente foi, afetando nossa memória, julgamentos e ações posteriores.

Nos sistemas algorítmicos, o viés retrospectivo reforça estereótipos de gênero ao interpretar as ações de indivíduos com base em expectativas sociais preestabelecidas, em vez de considerar fatores contextuais ou individuais. Isso perpetua papéis tradicionais de gênero, ignora barreiras estruturais e limita as escolhas e oportunidades de mulheres, contribuindo para a reprodução de desigualdades e preconceitos. Portanto, o viés retrospectivo não só reflete o que é dado, mas também confirma a desigualdade, quando, ao buscar padrões históricos, perpetua estereótipos relacionados ao papel das mulheres.

### **2.1.8 Erro Fundamental de Atribuição ou Viés de Correspondência**

Refere-se a uma tendência psicológica humana de atribuir os comportamentos de outras pessoas a características intrínsecas (como personalidade ou intenções), ignorando fatores contextuais ou situacionais. Quando aplicado ao contexto de sistemas de Inteligência Artificial (IA), especialmente em questões relacionadas ao gênero, o conceito pode ser ampliado para analisar como as decisões ou respostas de tais sistemas são interpretadas e influenciadas por pressupostos enviesados.

O viés de correspondência pode ser observado quando atribuímos características humanas, como intenções ou valores morais, às respostas de algoritmos.

No contexto de gênero, isso ocorre, por exemplo, quando assistentes virtuais representando imagens femininas, ou com vozes femininas, são percebidos como mais adequados para funções de atendimento, reforçando estereótipos culturais de feminilidade associada à submissão ou à docilidade.

Ana Paula Sciammarella, Professora da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e do Programa de Pós-graduação em Direito e Políticas Públicas da UNIRIO (PPGD/UNIRIO), em sua obra *Magistratura e gênero: uma análise da condição profissional feminina no Judiciário Fluminense*<sup>79</sup>, focalizou o processo de feminização da

---

<sup>79</sup>SCIAMMARELLA, Ana Paula. **Magistratura e gênero: uma análise da condição profissional feminina no Judiciário Fluminense**. Rio de Janeiro: Autografia, 2020. [recurso eletrônico EPUB].

magistratura fluminense, explorando percepções, significados e discursos das magistradas sobre a interseção entre profissão e gênero. Segundo a autora:

O impacto da carreira na vida privada (e vice-versa), evidenciado pela dificuldade de articulação entre os papéis profissionais e familiares, sugere a persistência de um modelo de família no qual as responsabilidades domésticas cabem tão somente às mulheres. Assim, a disponibilidade das mulheres para o investimento na carreira depende de uma complexa combinação de características pessoais, dentre elas o estado civil e a maternidade.

Quanto à hierarquia profissional, mesmo com o processo de feminização, a magistratura está inserida em um sistema de gênero que reflete um paradoxo: magistradas frequentemente adotam estratégias de apagamento de identidade de gênero, embora suas narrativas revelem a existência de discriminação. Afora isso, há dificuldades e diferenciações na ocupação de cargos de poder e nos processos de promoção nas carreiras, indicando a prevalência de um paradigma masculino no exercício profissional. Esse contexto exige das magistradas posturas profissionais mais rígidas e a constante necessidade de afirmação de competência para o desempenho de suas funções.

Embora muitas mulheres não possuam um emprego satisfatório, a ocupação da esfera pública por meio da inserção profissional, rompendo com a rotina de isolamento e trabalho doméstico, é vista como um bem.

De acordo com a pesquisadora, é essencial analisar criticamente as dinâmicas de poder em espaços tradicionalmente vistos como não públicos ou alheios à política. Essa análise é fundamental para compreender as implicações políticas dos acordos privados. Sem analisar essas relações, fica difícil entender como interações aparentemente livres e espontâneas reforçam hierarquias de poder e perpetuam a submissão das mulheres, limitando diretamente sua autonomia tanto no âmbito público quanto no privado.

O *chatbot* do Tribunal de Justiça do Estado do Rio de Janeiro (TJRJ) para o sistema PJe recebeu o nome de IETE. Conforme o próprio Tribunal, IETE seria “uma homenagem à MM. Juíza e Desembargadora Iete Bomilcar Ribeiro de Souza Passarella, primeira magistrada mulher do TJRJ”, e integra o Projeto Justiça 4.0 implementado pelo Conselho Nacional de Justiça.

Sobre a desembargadora “homenageada” pelo TJRJ, a autora registra:

Os dados coletados nos contam ainda que Iete ingressou na faculdade de Direito em 1935 e se casou em 1937. Teve três filhos, mas não desistiu da ideia de seguir a carreira de magistrada. Segundo as informações contidas no acervo, desejava ser juíza, como o pai. Em 1945, tentou o concurso que seria realizado no Rio de Janeiro, então Distrito Federal. Inscreveu-se, mas a comissão de seleção, tomando conhecimento da inscrição de uma mulher, resolveu impugná-la (9 votos contra 7), sob a argumentação de “falta de idoneidade específica”. Inconformada com a discriminação que sofreu, Iete impetrou um mandado de segurança, para que fosse esclarecido o real motivo da impugnação. O recurso ficou sem julgamento. Em 1951, um novo concurso foi aberto, ela se inscreveu e sua inscrição foi aprovada. Durante o concurso, dois dos três desembargadores/examinadores que compunham sua banca eram declaradamente contra o ingresso de mulheres na magistratura.

Seu pioneirismo e ousadia foram retratados em matéria do jornal *Correio da Manhã*, de 1º de janeiro de 1955, intitulada “Autodeterminação que derrubou séculos de preconceitos sociais”. O texto narrava a história da desembargadora Iete e sua “luta para vencer a obstinação de desembargadores apegados a conceitos arcaicos”, mas ainda a tratava com o “dona” Iete, uma “mulher de médico e mãe de três filhos”<sup>80</sup>.

A trajetória da Desembargadora Iete Bomilcar Passarella ilustra um paradoxo: Por um lado, destaca sua notável determinação ao superar discriminações e preconceitos em sua jornada para se tornar magistrada, enfrentando barreiras de gênero que ainda persistem em diversas áreas profissionais, cuja perseverança é um testemunho de resistência e conquista em um espaço historicamente dominado por homens. Por outro lado, sua história revela uma ironia: embora sua luta mereça reconhecimento e admiração, sua valorização pública parece ter ocorrido não por suas contribuições jurídicas ou competência profissional, mas pela escolha de seu nome para um sistema de inteligência artificial.

Essa reflexão ilustra como as normas de gênero e as expectativas sociais permanecem profundamente enraizadas, reforçando a necessidade de uma conscientização constante acerca das questões de gênero e destacando a urgência de combater estereótipos nocivos que ainda persistem, mesmo em contextos que aparentam avanços progressistas.

A escolha por nomes, vozes e avatares femininas é resultado de decisões de design orientadas por um mercado que reflete e perpetua preconceitos de gênero. No exemplo citado, além do viés de atribuição, outros vieses se evidenciam, como o viés de confirmação e o viés de retrospectiva.

Além desses, também podemos citar o viés de representatividade, cuja utilização se estabelece como paradigma de subordinação aos papéis de gênero na experiência de

---

<sup>80</sup>SCIAMMARELLA, *Op. cit.*

mulheres na magistratura. O viés de negatividade é evidente como discriminação e preconceito, com pouca atenção para os avanços na igualdade de gênero. Por fim, o viés de impacto reflete-se na interpretação de que a homenagem à desembargadora pelo nome do *chatbot* cria um simbolismo negativo desse ato em detrimento de seu reconhecimento histórico.

O erro de atribuição não se limita à interpretação humana das decisões algorítmicas; ele também está presente no treinamento e na implementação dos modelos.

Sistemas de recrutamento baseados em Inteligência Artificial que priorizam currículos historicamente associados a perfis masculinos refletem, não necessariamente uma intenção explícita dos programadores, mas um contexto cultural de discriminação implícita nos dados históricos. A atribuição de tal comportamento à “natureza” do sistema, sem considerar os fatores estruturais que o influenciam, representa o viés de correspondência aplicado às tecnologias.

### **2.1.9 Viés do Ator-Observador**

O viés do ator-observador pode se manifestar na forma como desenvolvedores e usuários explicam as decisões algorítmicas. Desenvolvedores, na posição de “atores”, tendem a justificar vieses ou erros de seus sistemas como produtos de limitações técnicas, restrições nos dados disponíveis ou contextos operacionais. Usuários, por sua vez, na posição de “observadores”, frequentemente atribuem tais falhas ou comportamentos enviesados a características inerentes à própria IA, como uma “tendência natural” à discriminação ou falta de neutralidade.

Essa dinâmica é particularmente evidente em questões de gênero, onde sistemas de IA podem reproduzir e amplificar desigualdades estruturais. Ao adotar soluções tecnológicas sem reflexão crítica, os usuários podem reforçar inadvertidamente os estigmas de gênero perpetuados por esses sistemas. A validação ou aceitação passivamente dessas tecnologias contribuem para a manutenção de desigualdades sistêmicas, mesmo sem intenção explícita.

Assim, a responsabilidade não recai apenas sobre os desenvolvedores, mas também sobre os usuários que, ao não questionarem os vieses presentes, acabam legitimando práticas discriminatórias.

O viés do ator-observador também influencia o design e a aplicação prática da IA. Quando assistentes virtuais recebem vozes e nomes femininos, os desenvolvedores frequentemente justificam a escolha como uma resposta a preferências do mercado (uma variável situacional) e os usuários podem não entender isso como um reflexo intrínseco de sexismo na tecnologia, supondo que é “da natureza da mulher” cumprir essa função.

Essa dicotomia de percepções mascara o papel ativo de contextos socioculturais e decisões de design no reforço de estereótipos de gênero. O viés do ator-observador, desse modo, pode obscurecer a compreensão das responsabilidades compartilhadas na perpetuação das desigualdades. Enquanto os desenvolvedores tendem a minimizar o impacto de suas escolhas de design, atribuindo-o a fatores externos, como limitações de dados, os usuários, por sua vez, podem negligenciar seu papel na legitimação de sistemas discriminatórios ao adotarem soluções sem a devida crítica, que acabam por reforçar estigmas de gênero. A falta de reconhecimento mútuo das responsabilidades contribui para a opacidade dos sistemas de IA, dificultando a implementação de ações corretivas, uma vez que a ausência de *feedback* humano impede o ajuste adequado dos vieses presentes nesses sistemas.

### **2.1.10 Viés da Perspectiva de Câmera**

O viés da perspectiva de câmera refere-se a um fenômeno psicológico em que a forma como uma situação é apresentada visualmente influencia a interpretação do observador. Em vídeos, a escolha do ângulo da câmera pode direcionar empatia ou julgamento em relação a quem está sendo filmado.

No contexto de IA, o viés da perspectiva de câmera pode surgir em sistemas que dependem de imagens e vídeos - como algoritmos de reconhecimento facial, análise comportamental e monitoramento de ambientes -, cujo funcionamento é profundamente influenciado por como as imagens são capturadas: ângulos, iluminação, resolução e até as posições dos sujeitos nas cenas podem afetar a precisão dos resultados. Em relação ao gênero, esse viés frequentemente se manifesta na forma de desigualdades de desempenho e interpretação entre homens e mulheres, em especial sobre populações racializadas.

Algoritmos de reconhecimento facial são notoriamente menos precisos para identificar mulheres, especialmente mulheres negras, em comparação com homens brancos.

O problema não está apenas relacionado à qualidade dos dados de treinamento, mas também à forma como as imagens são capturadas: câmeras posicionadas de maneira desigual podem favorecer certos tipos de fisionomia, com ângulos e iluminações que destacam características mais comuns em grupos privilegiados. Esse desequilíbrio técnico, quando não abordado, é amplificado pelas decisões algorítmicas e pela interpretação dos usuários, resultando em discriminações que parecem intrínsecas ao sistema, mas que, de fato, têm raízes em escolhas metodológicas e estruturais.

Outro aspecto do viés da perspectiva de câmera está presente em sistemas de vigilância automatizada, que frequentemente tratam os comportamentos de homens e mulheres de maneira distinta. Algoritmos que monitoram padrões de movimentação podem interpretar o mesmo comportamento de forma diferente dependendo de como a cena foi capturada ou da posição do indivíduo no espaço filmado. Outrossim, o viés da perspectiva de câmera não se limita a questões técnicas; ele também reflete escolhas de design e valores culturais.

Em muitos sistemas de IA, as representações visuais de mulheres são frequentemente sexualizadas ou limitadas a estereótipos estéticos. Isso ocorre em tecnologias como avatares digitais ou sistemas de recomendação baseados em imagens, que utilizam padrões de gênero normativos como referência para a personalização e interação.

A ampla implementação de sistemas de reconhecimento facial em diversas áreas, como segurança pública, transporte urbano, instituições educacionais, programas de assistência social, fiscalização aduaneira e verificação de identidade, supostamente, teria como propósito a atuação integrada das forças de segurança pública. As autoridades públicas têm aumentado significativamente o uso de tecnologias de reconhecimento facial, visando aprimorar a eficácia administrativa.

Essa é a justificativa apresentada para a Plataforma de Monitoramento CórteX<sup>81</sup>. Criada pelo Ministério da Justiça e Segurança Pública, por meio da Secretaria de Operações Integradas (SIOP), a Plataforma CórteX foi alvo de uma reportagem feita pelo Intercept Brasil<sup>82</sup>, segundo o qual o sistema revela um avanço preocupante na capacidade do governo

---

81BRASIL. Ministério da Justiça e Segurança Pública e da Secretaria de Operações Integradas (SIOP). “Plataforma de Monitoramento CórteX”. Disponível em: <https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/operacoes-integradas/destaques/plataforma-de-monitoramento-cortex>. Acesso em: 8 mar. 2025.

82REBELLO, Aiuri. “Da placa de carro ao CPF: Conheça o CórteX, sistema de vigilância do governo que integra de placa de carro a dados de emprego”. **Intercept Brasil**. Disponível em: <https://www.intercept.com.br/2020/09/21/governo-vigilancia-cortex/>. Acesso em: 8 mar. 2025.

de vigiar a população.

Instalado por meio de uma extensa rede de câmeras em rodovias, ruas, pontes e túneis, esse mecanismo tecnológico utiliza a leitura de placas veiculares e integra dados de diversas fontes, chegando até mesmo a acessar informações sensíveis da RAIS, do Ministério da Economia. Embora oficialmente justificado como um instrumento de combate ao crime, o CórteX expande seu alcance para a construção de perfis detalhados dos cidadãos, cruzando dados de localização com informações pessoais e profissionais, o que levanta sérias suspeitas de uso abusivo e invasão da privacidade.

Conforme divulgado pela Agência Brasil, sistemas de reconhecimento facial estão sendo usados em 37 cidades no país<sup>83</sup>.

O funcionamento dessa tecnologia é, portanto, delicado, principalmente devido à natureza dos dados envolvidos. Bancos de dados usados para treinamento, frequentemente contêm representações desproporcionais de homens em relação a mulheres ou de pessoas de determinadas raças e idades, o que resulta em um desempenho inferior do sistema para grupos sub-representados.

No caso específico de gênero, as características faciais das mulheres podem ser subavaliadas devido à menor quantidade de amostras ou à padronização de traços associados predominantemente aos homens, gerando taxas mais altas de erros, como falsas rejeições ou falsos reconhecimentos. A tecnologia de reconhecimento facial, ao capturar a imagem de um rosto, reconhece características específicas da pessoa, como a distância entre os olhos, a largura do queixo e o comprimento da boca. Utilizando esses dados, o *software* elabora uma fórmula única, conhecida como “assinatura facial”, que serve como a chave para identificação da pessoa.

Uma vez que os dados utilizados para compor a assinatura facial estão associados a características físicas únicas da pessoa, são considerados dados biométricos<sup>84</sup>.

Essa assinatura é confrontada com outras já registradas em um banco de dados contendo imagens de indivíduos procurados. Se as assinaturas faciais são compatíveis, é

---

83VALENTE, Jonas. “Tecnologias de reconhecimento facial são usadas em 37 cidades no país”. **Agência Brasil**. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-09/tecnologias-de-reconhecimento-facial-sao-usadas-em-37-cidades-no-pais>. Acesso em: 8 mar. 2025.

84A LGPD não oferece uma definição precisa para dados biométricos, mas, para o propósito dessa pesquisa, conceituamos dados biométricos como aqueles capazes de identificar uma pessoa de forma única e exclusiva. Esses dados, ao se relacionarem com características físicas ou comportamentais, têm o potencial de individualizar a pessoa, conforme documento da Autoridade Nacional de Proteção de Dados. (CEBRIAN et al. **Biometria e reconhecimento facial**: estudos preliminares. Radar Tecnológico nº 2. ANPD: Brasília, 2024).

viável identificar automaticamente um sujeito. Estando a assinatura facial de um indivíduo registrada no banco de dados de um sistema, essa pessoa não terá a opção de não ser reconhecida, a menos que os dados sejam apagados. Além disso, a ciência da pessoa a ser identificada não é necessária para que essa tecnologia a reconheça, ao contrário da impressão digital, em que o titular está consciente do procedimento de identificação. Ou seja, reconhecer alguém erroneamente num contexto de segurança pública, pode levar a abordagens e apreensões indevidas – como ocorreu no Rio de Janeiro, em que uma mulher foi confundida com outra que havia cometido um crime e, por isso, foi direcionada à delegacia<sup>85</sup>.

Em outra ocasião, uma psicóloga de 35 anos foi abordada por agentes do programa Segurança Presente durante a 5ª Conferência Estadual de Igualdade Racial, realizada no Liceu de Artes e Ofícios, no Rio de Janeiro<sup>86</sup>. O sistema de reconhecimento facial utilizado pelos agentes a identificou erroneamente como uma mulher com mandado de prisão em aberto. Apesar das diferenças físicas evidentes entre a psicóloga e a suspeita, ela só foi liberada após apresentar seu documento de identificação.

Em pesquisa realizada pelo LAPIN<sup>87</sup>, na qual foram levantadas informações sobre o uso da tecnologia de reconhecimento facial pelo setor público brasileiro, o estudo apontou uma série de violações aos direitos humanos pelo uso dessa tecnologia. Entre outras violações, foi apurado que a utilização indiscriminada de tecnologias de reconhecimento facial representa uma ofensa à privacidade, à liberdade de locomoção e à integridade da honra e da imagem das pessoas, uma vez que possibilita a monitorização e identificação das mesmas em espaços públicos.

Esse tipo de tecnologia compromete o princípio da presunção de inocência ao tratar todo indivíduo como suspeito em potencial, sujeito a monitoramento e identificação pelo Estado, e também constitui uma violação do direito à proteção de dados pessoais, estabelecido como um direito fundamental autônomo pelo Supremo Tribunal Federal<sup>88</sup> em maio de 2020.

---

85COELHO, Henrique; NASCIMENTO, Rafael; ALVES, Raoni. “Mulher presa após reconhecimento facial é solta; mandado de prisão já tinha sido cumprido”. **G1Globo**. Publicado em 04/01/2024. Disponível em: <https://g1.globo.com/rj/rio-de-janeiro/noticia/2024/01/04/mulher-presa-apos-reconhecimento-facial-e-solta-mandado-de-prisao-ja-tinha-sido-cumprido.ghtml>. Acesso em: 8 mar. 2025.

86BRAGA, Nathália. “Erro em reconhecimento facial constrange psicóloga em conferência”. **Defensoria Pública do Estado do Rio de Janeiro**, Rio de Janeiro, 12 jul. 2024. Disponível em: <https://defensoria.rj.def.br/noticia/detalhes/29955-Erro-em-reconhecimento-facial-constrange-psicologa-em-conferencia>. Acesso em: 8 mar. 2025.

87REIS, Carolina; ALMEIDA, Eduarda; DA SILVA, Felipe; DOURADO, Fernando. “Relatório sobre o uso de tecnologias de reconhecimento facial e câmeras de vigilância pela administração pública no Brasil”. Brasília: **Laboratório de Políticas Públicas e Internet**, 2021.

A correlação entre os vieses algorítmicos e os vieses cognitivos revela que os sistemas de Inteligência Artificial (IA) tendem a reproduzir, em suas decisões, padrões de pensamento enviesados semelhantes aos dos seres humanos porque os algoritmos são criados e treinados por pessoas, que podem, de forma consciente ou não, transferir seus próprios vieses para as máquinas. Ao serem alimentados por dados produzidos e manipulados por seres humanos, inevitavelmente carregam as influências dessas distorções, reproduzindo padrões de pensamento enviesados que podem prejudicar a imparcialidade e a justiça nas decisões automatizadas.

Desse modo, vieses de dados e heurística da disponibilidade se baseiam em informações mais acessíveis ou frequentes, podendo perpetuar preconceitos; vieses de seleção de amostras e heurística da ancoragem envolvem decisões baseadas em referências inadequadas; vieses de atribuição e erro fundamental de atribuição atribuem causas incorretas a efeitos; vieses de modelagem e heurística da correlação ilusória envolvem relações percebidas de forma equivocada; vieses de otimização e viés egocêntrico priorizam objetivos enviesados, como lucro em detrimento da equidade; vieses de retroalimentação e viés de confirmação reforçam crenças pré-existentes; vieses contextuais e viés do ator-observador envolvem interpretações enviesadas com base no contexto; vieses de saída e viés retrospectivo geram conclusões enviesadas após o fato; viés de reforço e heurística do afeto são guiados por recompensas ou emoções enviesadas; e, por fim, o viés de *feedback* humano e o viés da perspectiva de câmera são influenciados por ângulos ou perspectivas enviesadas.

A responsabilidade jurídica pelos vieses algorítmicos geralmente recai sobre desenvolvedores, organizações e plataformas, que devem adotar práticas éticas, como diversidade de dados, validação de modelos e transparência. No Brasil, essa responsabilidade está amparada em diversos marcos legais e regulatórios. O Código de Defesa do Consumidor (CDC)<sup>89</sup>, por exemplo, responsabiliza fornecedores de produtos e serviços por danos decorrentes de defeitos, o que pode abranger vieses algorítmicos que causem discriminação ou prejuízos aos usuários. A Lei Geral de Proteção de Dados (LGPD)<sup>90</sup>, por sua vez, estabelece obrigações para o tratamento de dados pessoais, exigindo transparência, segurança e não discriminação, sob risco de multas e sanções em caso de descumprimento.

---

88STF Portal de Notícias. “STF suspende compartilhamento de dados de usuários de telefônicas com IBGE”. Publicado em 07/05/2020. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=442902&ori=1>. Acesso em: 8 mar. 2025.

89BRASIL. Congresso Nacional. **Lei nº 8.078**, de 11 de setembro de 1990. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l8078compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm). Acesso em: 8 mar. 2025.

90BRASIL. Congresso Nacional. **Lei nº 13.709** de 14 de agosto de 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em: 8 mar. 2025.

Outro marco importante é o Marco Civil da Internet<sup>91</sup>, que assegura a neutralidade da rede e responsabiliza plataformas por danos resultantes de conteúdo gerado por terceiros. Essa responsabilidade pode se estender a sistemas algorítmicos que perpetuem vieses, especialmente quando contribuem para práticas discriminatórias ou prejudiciais.

---

91BRASIL. Congresso Nacional. **Lei nº 12.965**, de 22 de abril de 2014. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm). Acesso em: 8 mar. 2025.

### CAPÍTULO 3 – GOVERNANÇA ALGORÍTMICA

A governança algorítmica surge como uma prática indispensável à medida que as tecnologias digitais avançam, e seu estudo é fundamental para compreendermos as implicações jurídicas e éticas que envolvem o uso de algoritmos. Este capítulo tem o objetivo de explorar as interações entre as normas que regem os algoritmos e os direitos fundamentais, ao mesmo tempo em que busca entender as estratégias de governança e os desafios éticos que surgem com o desenvolvimento da Inteligência Artificial (IA).

Ao longo do texto, vamos examinar como a gestão de riscos é aplicada na implementação de algoritmos e refletir sobre a necessidade de um marco regulatório que seja eficaz tanto em nível global quanto nacional. Por fim, discute-se as etapas de avaliação de impacto algorítmico e as formas de mitigar os riscos da IA, destacando a importância de adotar uma abordagem equilibrada que favoreça o desenvolvimento sustentável e ético dessa tecnologia.

Algoritmos que possuem autoridade não apenas “fazem coisas” para o *software*, mas também podem fazer com que atores humanos respondam de acordo, tanto de maneira indireta (por meio da construção do conhecimento dos algoritmos de relevância pública) quanto diretamente (por meio de algoritmos de heteromação)<sup>92</sup>.

Algoritmos de heteromação são sistemas computacionais que operam sob um princípio de controle externo, automatizando decisões ou ações em que o sujeito humano não exerce autonomia direta. Diferentemente da automação tradicional, que frequentemente visa aliviar o trabalho humano em tarefas repetitivas ou mecânicas, a heteromação introduz um elemento de dependência, no qual os algoritmos não apenas substituem a intervenção humana, mas também a condicionam ou subordinam. Esses algoritmos podem exercer uma forma de controle e dominação sobre as pessoas, embora isso nem sempre seja explícito ou facilmente perceptível.

Em plataformas digitais, por exemplo, os algoritmos parecem ser os responsáveis pelas decisões, mas dependem de dados e ações humanas para funcionar. Por outro lado, os algoritmos também podem direcionar as escolhas e ações das pessoas de forma sutil.

Nas redes sociais, onde esses sistemas desempenham um papel central na curadoria de conteúdos, algoritmos de heteromação selecionam informações com base em padrões

---

92BRAZ, Matheus V. “Heteromação e microtrabalho no Brasil”. *Sociologias*, Porto Alegre, ano 23, n. 57, maio 2021, p. 134-172. Disponível em: <http://doi.org/10.1590/15174522-111017>. Acesso em: 8 mar. 2025.

predefinidos, reforçando frequentemente comportamentos e preferências já existentes, influenciando o que vemos, compartilhamos ou compramos e moldando nossos comportamentos de forma quase imperceptível.

Essa dinâmica restringe nossa autonomia e liberdade de escolha, já que nossas decisões passam a ser guiadas por dados preditivos, regras opacas e por interesses econômicos ou políticos das empresas que controlam essas plataformas. Como consequência, contribuem para a formação de bolhas de informação e limitam a exposição a perspectivas diversas, reduzindo a pluralidade de ideias e opiniões acessíveis aos usuários.

Em sistemas de recrutamento, os algoritmos realizam a triagem de candidatos com base em critérios previamente estabelecidos, muitas vezes padronizados, o que tende a suprimir elementos importantes para decisões humanas e contextualizadas. Outro aspecto importante é a exploração laboral. Plataformas como aplicativos de entrega ou transporte, os algoritmos gerenciam e monitoram o trabalho humano, impondo ritmos, metas e condições que podem ser exaustivas ou injustas. Os trabalhadores muitas vezes têm pouca autonomia sobre suas tarefas, mas carregam a responsabilidade pelos resultados.

A heteromação também pode criar uma dependência das pessoas em relação aos sistemas automatizados. Como os algoritmos tomam decisões ou simplificam processos, as pessoas podem se sentir menos capazes de agir sem a mediação da tecnologia. Dessa forma, os algoritmos de heteromação, embora pareçam neutros ou benéficos, podem reforçar estruturas de poder que limitam a autonomia e a liberdade das pessoas.

No artigo “Inteligência Ética”, DA SILVA & HIRATA<sup>93</sup> exploram as duas abordagens complementares presentes nas atividades de IA: a simbólica, baseada em sistemas lógicos, e a adaptativa, baseada em aprendizado de máquina.

Eles discutem o conceito de um sistema ético, além das normas e valores que embasam uma vida significativa, com raízes na Grécia Antiga, e delineiam as três diferentes classes desses sistemas: baseados em virtudes, em deveres e nas consequências das ações.

No desenvolvimento de sistemas inteligentes, a ética baseada nas consequências das ações é considerada a mais simples das três categorias. Ao criar um sistema para resolver problemas, de acordo com os autores, são determinadas as ações que ele pode realizar, levando em conta os possíveis impactos negativos destas ações e permitindo a implementação de medidas corretivas para mitigá-los ou minimizar suas consequências.

---

93DA SILVA, Flávio S. Corrêa, HIRATA, Nina S. T. “Inteligência Ética. Ética e Regulação na Inteligência Artificial”. *Revista da Sociedade Brasileira de Computação*, julho/2022, nº 47.

A ética centrada em deveres também recebe atenção significativa, fundamentando-se na definição clara de normas e valores, que são formalizados por meio de sistemas deontológicos formais. As normas deontológicas são aplicadas principalmente como restrições e condições que moldam e orientam o comportamento dos sistemas. A integração de mecanismos deontológicos foca, com frequência, na prevenção de violações, visando a redução de danos.

Por outro lado, observa-se uma convergência entre abordagens simbólicas e adaptativas em IA, bem como entre abordagens gerativas e discriminativas no aprendizado de máquina, avançando para maior transparência e explicabilidade.

Afora as abordagens mencionadas por DA SILVA & HIRATA, outras perspectivas éticas vêm sendo exploradas para orientar o desenvolvimento e a aplicação ética da Inteligência Artificial (IA); a ética do cuidado destaca a importância das relações interpessoais e sociais na tomada de decisões éticas em IA. Inspirada na obra de GILLIGAN<sup>94</sup>, reconhece a necessidade de considerar não apenas as consequências das ações, mas também o impacto que essas ações têm nas pessoas e nas comunidades envolvidas. Ao priorizar valores como empatia, responsabilidade e sensibilidade às necessidades dos outros, a ética do cuidado busca garantir que os sistemas de IA sejam desenvolvidos e utilizados de maneira a promover o bem-estar e a justiça social.

A ética do design<sup>95</sup> destaca a importância de incorporar princípios éticos desde as fases iniciais do desenvolvimento de sistemas inteligentes, o que inclui a consideração de questões como transparência, privacidade, equidade e responsabilidade social ao projetar algoritmos, interfaces e arquiteturas de IA. Outras perspectivas éticas apresentadas para a IA incluem a ética da justiça, baseada nas obras de John Rawls e Amartya Sen, que busca garantir que os benefícios e riscos da IA sejam distribuídos de maneira justa e equitativa; a ética da autonomia<sup>96</sup>, que defende o respeito à autonomia e liberdade dos indivíduos diante das decisões tomadas por sistemas de IA; e a ética da responsabilidade<sup>97</sup>, que destaca a importância dos desenvolvedores e operadores desses sistemas assumirem a responsabilidade pelas consequências de suas criações.

---

94GILLIGAN, Carol. **In a Different Voice: Psychological Theory and Women's Development**. Boston: Harvard University Press, 1982.

95FRIEDMAN, B.; NISSENBAUM, H. (1996). "Bias in computer systems". **ACM Transactions on Information Systems** (TOIS), 14(3), pp. 330-347.

96FLORIDI, Luciano. **The fourth revolution: how the infosphere is reshaping human reality**. Oxford University Press UK, 2014.

97VAN DEN HOVEN, Jeroen. "Computer Ethics And Moral Methodology". **Metaphilosophy**. Vol. 28, N. 3, July, 1997. Cambridge: Blackwell Publishers, pp. 234-248.

Recentemente, a ética baseada em virtudes tem ganhado mais atenção no desenvolvimento de sistemas inteligentes. Essa abordagem representa um desafio maior para a implementação, pois exige que os sistemas busquem benefícios internos e evitem malefícios, observando mecanismos externos de controle comportamental. Isso requer programar ações que sigam tanto a lógica simbólica quanto inferências baseadas no aprendizado adaptativo, levando em consideração não apenas a satisfação imediata, mas também os benefícios para a comunidade e a autorrealização. Para tanto, é essencial formular requisitos técnicos específicos e assegurar a mensuração do alinhamento dos sistemas a esses requisitos.

Considerada a mais emblemática de sua produção intelectual, em *After Virtue*<sup>98</sup>, Alasdair MacIntyre defende a revitalização da ética das virtudes inspirada na tradição aristotélica, mas ajustada às dinâmicas do mundo contemporâneo. Ele concebe as virtudes como qualidades indispensáveis que capacitam os indivíduos a alcançar os chamados “bens internos” das práticas, promovendo, assim, o pleno desenvolvimento humano. Para o autor, entender as virtudes requer, inicialmente, compreender o conceito de “prática”, a qual define como uma atividade cooperativa humana, estruturada de maneira coerente e complexa, que contém bens internos específicos. Exemplos incluem atividades como a agricultura, a ciência, as artes e os esportes.

O ponto central é que esses bens internos só podem ser plenamente percebidos por meio da excelência dentro de cada prática, em contraste com os bens externos, como riqueza ou prestígio. A vida humana, segundo o pensador, possui uma natureza essencialmente narrativa. As pessoas se percebem como personagens centrais de uma história que integra passado, presente e uma projeção de futuro. Desse modo, as virtudes são indispensáveis para alinhar a narrativa individual à tradição moral que a embasa. Essa perspectiva também reflete o vínculo entre narrativa e identidade, uma vez que as histórias que construímos sobre nós mesmos, bem como o contexto cultural que as molda, definem quem somos.

MacINTYRE<sup>99</sup> conecta a prática das virtudes às tradições culturais e históricas. Para ele, as virtudes não surgem isoladas, mas sempre se encontram inseridas em uma tradição que confere sentido às ações humanas e orienta as escolhas morais.

Inspirando-se em Aristóteles, ele defende que a vida humana se direciona a um *tèlos*, um propósito último que chama de “bem humano”. Esse bem, na sua perspectiva, não é algo

---

98MacINTYRE, A. *After Virtue* – A Study in Moral Theory. Third Edition. Indiana: University of Notre Dame Press, 2007.

99MacINTYRE, *Op. cit.*

subjetivo, mas pode ser reconhecido objetivamente por meio da razão prática e da convivência em comunidade. Portanto, as virtudes se apresentam como disposições indispensáveis para alcançar esse fim.

A ética das virtudes de Alasdair MacIntyre oferece uma perspectiva interessante para o desenvolvimento de sistemas inteligentes, especialmente ao destacar a relevância de valores humanos e práticas sociais. Segundo o autor, a ética está vinculada à busca pela excelência em práticas que promovem o bem comum. Aplicar essa visão ao campo tecnológico significa projetar sistemas que não apenas maximizem eficiência ou resultados, mas que também respeitem e fomentem valores fundamentais como justiça, solidariedade e honestidade.

Nesse sentido, sistemas inteligentes devem ser projetados para promover o bem-estar coletivo, alinhando-se aos “bens internos” das práticas humanas. Em vez de priorizar exclusivamente a redução de custos ou o aumento da produtividade, um sistema de saúde inteligente poderia ser estruturado para oferecer atendimento humanizado, reconhecendo as necessidades específicas de cada paciente. De forma semelhante, no sistema de justiça, a aplicação de tecnologias deve garantir decisões justas, claras e fundamentadas na equidade, promovendo a confiança pública e respeitando os direitos individuais.

Na administração pública, as tecnologias podem ser orientadas para melhorar a eficiência e a transparência, ao mesmo tempo que sustentam os valores éticos de responsabilidade, justiça e respeito ao interesse público.

Ademais, a ética das virtudes desempenha um papel fundamental ao direcionar o desenvolvimento desses sistemas para fomentar características como responsabilidade, prudência e empatia, tanto em seus usuários quanto em seus desenvolvedores, incentivando interações éticas, conscientes e alinhadas aos princípios do bem comum.

Outro aspecto da filosofia de MacINTYRE é a consideração do contexto social e histórico. Sistemas inteligentes devem ser sensíveis às tradições e valores das comunidades em que serão implementados, evitando soluções que desconsiderem especificidades locais ou imponham padrões éticos alheios à realidade cultural. Assim, o desenvolvimento de sistemas inteligentes, inspirado pela ética das virtudes, demanda uma abordagem holística, que integre avanços tecnológicos com reflexões éticas e sociais, priorizando a construção de tecnologias comprometidas com a dignidade humana e a promoção do bem comum.

### 3.1 NORMATIVIDADE ALGORÍTMICA CONCORRENTE

Projetados para processar informações, maximizar resultados e atender a objetivos previamente definidos – sejam eles de natureza econômica, social ou operacional –, esses sistemas criam um padrão regulatório que redefine as dinâmicas normativas contemporâneas, ao mesmo tempo em que geram temores sobre a capacidade de controle e regulação jurídica dessas novas formas de normatividade.

A pesquisadora norte-americana Shoshana Zuboff<sup>100</sup> alerta para os rumos questionáveis do uso de algoritmos, especialmente no que diz respeito à automatização da coleta de dados e à curadoria de conteúdos. Ela examina a prática da extração em massa de dados pessoais de usuários, com o objetivo de comercializar a segmentação de públicos e prever comportamentos de consumidores em potencial, o que resulta no aumento da renda publicitária.

As empresas de tecnologia, conhecidas como *big techs*, investem anualmente milhões de dólares na análise de comportamento a partir dos dados coletados. Utilizando sofisticadas ferramentas de análise de dados e Inteligência Artificial, essas empresas monitoram e reinterpretam padrões de comportamento dos usuários em suas plataformas, coletando dados variados, como histórico de navegação, interações em redes sociais, preferências de compra e até mesmo informações sobre o uso de dispositivos. Este processo de coleta e análise permite criar perfis detalhados dos usuários, entendendo seus hábitos, preferências e necessidades de forma cada vez mais precisa.

O principal motivo pelo qual as *big techs* realizam esses investimentos é, segundo os objetivos declarados, a personalização dos serviços oferecidos. Ao compreender melhor os usuários, essas empresas podem adaptar suas plataformas para fornecer conteúdo mais relevante e atraente, melhorando a experiência do usuário. Além disso, a personalização dos anúncios publicitários se torna mais eficiente, permitindo que as empresas direcionem suas campanhas de marketing para públicos específicos, aumentando as chances de conversão.

Essa abordagem direcionada não apenas melhoraria a satisfação do usuário, mas também maximizaria os lucros das empresas através de uma publicidade mais eficaz. Outro fator, alegadamente importante, seria a inovação contínua. A análise de grandes volumes de dados fornece informações que podem levar ao desenvolvimento de novos produtos e serviços.

---

100ZUBOFF, Shoshana. **A Era do capitalismo de Vigilância**. Tradução de George Schlesinger. Rio de Janeiro: Intrínseca, 2021[Edição digital].

As empresas podem, assim, identificar tendências emergentes, prever mudanças no comportamento do consumidor e se antecipar às necessidades do mercado. Isso confere uma vantagem competitiva, permitindo que as *big techs* permaneçam na vanguarda da tecnologia e mantenham sua posição de liderança no mercado. Para a pesquisadora, essa prática se baseia em uma arquitetura de extração automatizada que funciona como um grande espelho unidirecional, desconsiderando a consciência e o consentimento dos envolvidos.<sup>101</sup>

Esse imperativo de extração revela relações sociais assimétricas de poder e está inserido em um contexto maior, que ela denomina “capitalismo de vigilância”. Em sua concepção, essa estrutura é fundamentada em relações de subordinação e hierarquia. O algoritmo, por sua vez, coleta informações de usuários que não têm conhecimento das diretrizes da extração nem das formas pelas quais seus dados serão utilizados.<sup>102</sup> Conforme a autora, grandes conglomerados de tecnologia e plataformas de serviços online, disfarçam o imperativo de extração sob a justificativa de oferecer uma experiência personalizada: algoritmos e máquinas que coletam dados comportamentais com o intuito de “entender” os gostos individuais, criar perfis e sugerir conteúdos adequados a esses perfis, na alegação de prever preferências. Nesse processo, a experiência humana se torna uma mercadoria, que é mercantilizada e comercializada.

Parafrazeando TELES<sup>103</sup>, pode-se dizer que a convivência diária entre pessoas e tecnologias inaugura modos de interação que sobrepõem estratégias funcionais às subjetividades e às próprias estruturas sociais, revelando outros modos possíveis de exercício político. Essa simbiose crescente entre máquinas e humanos tende a apagar os contornos antes nítidos entre o discursivo e o funcional, convidando à reflexão sobre os regimes que moldam as subjetividades na governança da vida. A sociedade capitalista nos submete a um sistema duplo de controle. De um lado, promove a “humanização” dos indivíduos, atribuindo-lhes papéis previamente definidos. Por outro, adota um controle em um plano não “humanizado”, baseado em uma dinâmica que transcende as dicotomias tradicionais, como sujeito e objeto, natureza e cultura, público e privado, humano e não humano. Esse espaço híbrido parece configurar-se nas margens do político, em territórios periféricos.<sup>104</sup>

---

101 ZUBOFF, *Op. cit.*, pp. 162-163.

102 *Idem, ibidem*, p. 169-171.

103 TELES, Edson. “Ação Política Híbrida e a Dissolução da Cidadania”. **Revista de Filosofia Moderna e Contemporânea**, Brasília, v.8, n.3, dez. 2020, p. 81-103. Disponível em: <https://doi.org/10.26512/rfmc.v8i3.34494>. Acesso em: 8 mar. 2025

104 *Idem, ibidem*, p. 82.

Os algoritmos utilizados atualmente, especialmente por grandes empresas de tecnologia, foram inicialmente criados para ambientes corporativos e desenvolvidos com foco no mercado financeiro e de capitais, tendo sido projetados para otimizar processos, maximizar lucros e aumentar a eficiência operacional em contextos empresariais específicos. O objetivo principal era atender às necessidades de mercados altamente competitivos, como o financeiro, onde a precisão e a rapidez são fundamentais para o sucesso.

Entretanto, ao serem implementados em outras áreas, como a administração pública e as redes sociais, esses algoritmos nem sempre se adequam de maneira ideal. A lógica e os critérios de decisão incorporados nesses sistemas frequentemente desconsideram os valores e princípios fundamentais que sustentam um ambiente democrático, como a equidade e a participação cidadã. O ambiente empresarial, por sua própria natureza, privilegia a competição em detrimento da colaboração, um traço que se contrapõe às demandas de cooperação e inclusão inerentes à democracia.

AXELROD<sup>105</sup> demonstrou que a diversidade cultural surge inicialmente da escolha individual dos traços pelos agentes, gerando uma pluralidade de expressões culturais.

Contudo, essa variedade é subsequentemente afetada pela interação social, que, embora frequentemente vista como um agente de homogeneização, tende a reduzir as diferenças.

O modelo revela uma contradição: enquanto as escolhas individuais fomentam a diversidade cultural, a convivência social exerce uma tendência à uniformização das manifestações culturais. Essa dinâmica reflete um processo mais amplo de controle social, no qual os elementos culturais podem ser, paradoxalmente, tanto preservados quanto transformados, dependendo das estruturas de poder que mediam essa interação.

A relação entre liberdade de expressão e controle social se reflete em um fenômeno contemporâneo: o impacto das tecnologias digitais na forma como as informações são mediadas.

Para BALKIN<sup>106</sup>, o problema não reside nos algoritmos em si, mas nas entidades humanas e corporativas que os controlam. Hoje, as tecnologias digitais, por meio da mediação algorítmica, desempenham o papel de exercer controle social, consolidando relações de poder e influenciando as práticas de comunicação, muitas vezes de forma invisível, mas decisiva.

105AXELROD, Robert. **The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration**. Princeton: Princeton University Press, 1997.

106BALKIN, Jack M. "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation" (September 9, 2017). **UC Davis Law Review**, (2018 Forthcoming), Yale Law School, Public Law Research Paper Nº. 615, Ng. Disponível em: <https://openyls.law.yale.edu/handle/20.500.13051/4699>. Acesso em: 8 mar. 2025.

Nesse aspecto, o papel do Estado é determinante, especialmente considerando que a maior parte da comunicação digital ocorre em plataformas privadas. Em vez de censurar diretamente os indivíduos, os governos passaram a pressionar as plataformas para moderarem conteúdos considerados inadequados, o que altera a maneira tradicional de se garantir a liberdade de expressão.

Esse fenômeno dá origem à chamada “censura colateral”, quando intermediários, como provedores de internet ou redes sociais, são pressionados a censurar o discurso de terceiros para evitar sanções, o que resulta das regras de responsabilidade atribuídas a esses intermediários, que, para evitar punições, adotam práticas excessivas de controle, como bloqueio e filtragem de conteúdos.

O objetivo dos governos ao impor essa responsabilização é incentivar uma auto regulação que se traduza em restrição de conteúdos, criando um cenário em que a liberdade de expressão passa a ser governada não por decisões judiciais, mas por programadores e algoritmos. O autor identifica ainda um segundo problema nessa delegação: a restrição prévia administrativa. Em regimes de restrição prévia, a fala de um indivíduo é barrada até que um algoritmo ou funcionário corporativo autorize sua publicação, frequentemente sem aviso ou possibilidade de defesa.

Esse processo inverte o ônus da ação, colocando o indivíduo na posição de precisar de permissão prévia para se expressar, o que viola o princípio da liberdade de expressão sem censura antecipada. Além disso, quando a fala é removida ou filtrada por sistemas automatizados, o indivíduo é privado das garantias processuais estabelecidas para o devido processo legal. A regulamentação resultante, marcada pela censura colateral e por um viés autoritário, é problemática não apenas pela restrição da liberdade de expressão, mas também pela falta de transparência e pela ausência de responsabilização sobre quem define o que pode ou não ser dito. Conseqüentemente, a liberdade de expressão na era digital torna-se restringida por mecanismos invisíveis e inquestionáveis.

“Algocracia” é a combinação das palavras “algoritmo” e “democracia”, usado para descrever um sistema de governança ou de tomada de decisões em que os algoritmos – em vez de seres humanos ou instituições tradicionais – desempenham papel central na organização, controle e regulação de processos sociais e econômicos.

Em uma algocracia, decisões importantes, como a distribuição de recursos, o julgamento de comportamentos ou até mesmo a administração de políticas públicas, são

fortemente influenciadas ou totalmente determinadas por algoritmos, que operam de forma autônoma e, muitas vezes, de maneira opaca.

Supomos que o termo tenha sido cunhado por Aneesh Aneesh no artigo *Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies*<sup>107</sup>:

Em um sentido ideal típico, a nova forma de gestão – ou o que chamo de algocracia, isto é, o governo dos algoritmos – se diferencia de seu predecessor industrial principalmente em dois aspectos. Primeiro, a dominação é cada vez menos distribuída através de elaboradas hierarquias de trabalhadores; ao contrário, é cada vez mais realizada por meio de sistemas de informação e software que estruturam as formas possíveis de comportamento no trabalho.

Segundo, a governança algocrática parece transformar parcialmente as primeiras relações sujeito-objeto, onde um superior, como sujeito observador, deve supervisionar o trabalho de um subordinado. Essa mudança é marcada por uma relação de autoridade viabilizada por sistemas de informação e redes, onde todos são subordinados como nós nessas redes. Meu argumento relaciona a contínua desintegração da gestão vertical à arquitetura emergente dos sistemas de informação.<sup>108</sup> (Tradução nossa)

O conceito está associado a um novo paradigma de poder, no qual a autoridade política e social se desloca da figura humana para a tecnologia. Essa transição pode ser vista tanto como uma evolução tecnológica quanto como um desafio à democracia tradicional, uma vez que os algoritmos podem não ser transparentes, podem reforçar desigualdades ou podem operar de modo que não reflita os princípios democráticos fundamentais, como a equidade e a justiça.

O autor distingue entre três modos de governança organizacional: burocrática, pan-óptica e algocrática, cada uma definida por seus mecanismos de controle: escritório, vigilância e código, respectivamente. A transição para a algocracia representa uma mudança nas relações de autoridade, onde todos os indivíduos se tornam “nós” subordinados dentro de redes de informação.

Embora seja apresentada como uma forma eficiente e moderna de organizar o trabalho e os fluxos de informações, o autor alerta para os riscos dessa automatização da gestão, sugerindo que a governança algocrática pode comprometer a autonomia humana e aumentar a centralização do poder nas mãos dos algoritmos, ao eliminar intermediários e descentralizar o controle, podendo resultar em uma forma de dominação que, em vez de promover maior

107 ANEESH, A. “*Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies*”. **7th International Summer Academy on Technology Studies**, Deutschlandsberg, Austria. Disponível em: [https://www.researchgate.net/profile/A-Aneesh/publication/254843955\\_Technologically\\_Coded\\_Authority\\_The\\_Post-Industrial\\_Decline\\_in\\_Bureaucratic\\_Hierarchies/links/5bf45834a6fdcc3a8de46665/Technologically-Coded-Authority-The-Post-Industrial-Decline-in-Burea](https://www.researchgate.net/profile/A-Aneesh/publication/254843955_Technologically_Coded_Authority_The_Post-Industrial_Decline_in_Bureaucratic_Hierarchies/links/5bf45834a6fdcc3a8de46665/Technologically-Coded-Authority-The-Post-Industrial-Decline-in-Burea). Acesso em: 8 mar. 2025.

108 *Idem, ibidem*, pp 1-2.

liberdade e autonomia, acaba por reduzir as capacidades de intervenção humana nas decisões organizacionais.

A governança algorítmica suscita questões essenciais sobre o futuro do trabalho, a distribuição do capital e a relação entre indivíduos e tecnologias que estruturam as esferas econômica e social, impondo desafios éticos, jurídicos, políticos e sociológicos, tais como a concentração do poder, o avanço da vigilância massiva, a restrição da liberdade e da privacidade, além da exclusão de instâncias decisórias humanas fundamentais.

Um dos pilares da teoria de KELSEN<sup>109</sup> é o conceito de validade, que se refere à conformidade com os procedimentos formais de criação normativa estipulados pelo ordenamento jurídico, caracterizado, portanto, como autorreferente, procedimental e fechado a influências externas, o que assegura sua coerência e completude. Sua abordagem representa uma concepção de ordem jurídica como um conjunto de valores que separa o Direito da Moral, propondo que juízos de valor devem ser apartados de juízos de fato para encontrar respaldo na realidade empírica. Dessa forma, a realidade e o Direito são distinguidos em dois prismas: o “ser” e o “dever-ser”.

O reconhecimento da validade de uma norma é determinado apenas pela sua inclusão dentro de um sistema jurídico específico, independentemente do seu conteúdo. Assim, qualquer tipo de conteúdo pode ser considerado Direito. Não existe conduta humana que, em razão de sua essência ou características intrínsecas, esteja impossibilitada de integrar o conteúdo de uma norma jurídica. A validade de uma norma jurídica não se anula pelo simples fato de seu conteúdo ser contrário ao de uma norma pertencente a uma ordem jurídica diferente daquela cuja norma fundamental serve como base para sua validade.<sup>110</sup>

No sistema jurídico, o código lícito/ilícito orienta a produção de sentido. Analogamente, argumentamos que os algoritmos operam sob o código binário eficiente/ineficiente, o que os torna mediadores invisíveis de uma emergente normatividade concorrente e/ou paralela.

A normatividade algorítmica emergente, que categoriza comportamentos e antecipa decisões humanas, pode ser entendida como um sistema social em si, que compete com o Direito ao estabelecer padrões normativos concorrentes ou paralelos, colocando em questão a legitimidade das normas jurídicas tradicionais.

---

109KELSEN, Hans. **Teoria pura do direito**. 7a ed. Tradução de João Baptista Machado. São Paulo: Martins Fontes, 2006.

110KELSEN, *Op. cit.*, p. 221.

Para uma compreensão mais elaborada, trataremos de alguns dos conceitos abordados por LUHMANN<sup>111</sup> em sua teoria dos sistemas sociais.

A tese em análise fundamenta-se na ideia de que a sociedade moderna se estrutura a partir da diferenciação funcional, um modelo que organiza a sociedade em subsistemas autônomos, cada qual responsável por uma função específica. Esse paradigma representa uma ruptura com formas anteriores de organização social, como a segmentação e a estratificação, predominantes em sociedades agrárias e pré-modernas.

Na organização segmentária, típica de sociedades primitivas, as funções sociais eram distribuídas entre tribos ou comunidades relativamente independentes. Posteriormente, com o avanço para um modelo estratificado, a sociedade passou a ser dividida em hierarquias, nas quais o *status* social era determinado principalmente por nascimento ou herança. Com a complexificação da sociedade moderna, um modelo mais dinâmico e adaptável se fez necessário: a diferenciação funcional, que delega funções sociais a sistemas especializados.

A sociedade moderna é composta por diversos sistemas autônomos, cada um com funções específicas. Esses sistemas operam de maneira autorreferente, ou seja, seguem suas próprias regras e lógicas internas para lidar com a complexidade e alcançar seus objetivos sociais. Exemplos disso incluem a economia, fundamentada no dinheiro; o direito, estruturado em normas jurídicas; e a política, organizada em torno do poder.

A diferenciação funcional é considerada um avanço em relação aos modelos anteriores por sua flexibilidade e capacidade de adaptação às rápidas mudanças ambientais. Apesar da autonomia operacional, os sistemas não atuam de forma isolada. Ao contrário, são interdependentes, e essa interconexão é gerida por meio de “acoplamentos estruturais”, mecanismos que conectam os sistemas sem comprometer sua autonomia funcional.

Os acoplamentos estruturais desempenham papel central na interação entre sistemas autônomos, permitindo que eles se influenciem mutuamente sem perder sua independência. Exemplos incluem a interferência do sistema político na economia por meio de políticas fiscais ou a regulamentação jurídica das interações entre política e economia. Esses acoplamentos são essenciais para manter a coesão social em uma sociedade funcionalmente diferenciada, garantindo que, mesmo sem um mecanismo centralizado de coordenação, os sistemas operem de maneira integrada.

---

<sup>111</sup>LUHMANN, Niklas. **La sociedad de la sociedad**. México: Universidad Iberoamericana, 2006.

Um aspecto nodal da diferenciação funcional é a autonomia operacional dos sistemas. Na sociedade moderna, cada sistema social utiliza regras e normas próprias para desempenhar suas funções de forma independente. A política utiliza o poder como meio de comunicação, enquanto a economia opera com base no dinheiro.

A autorreferencialidade torna possível que os sistemas se reproduzam e mantenham sua coerência sem dependência direta de outros sistemas e sua autonomia operacional é o que permite à sociedade moderna lidar com a crescente complexidade. Cada sistema é especializado em sua função, o que aumenta a eficiência geral da sociedade.

Quando, dessa forma, se origina um sistema social, falaremos de diferenciação em referência ao que, como consequência dessa diferenciação, surge como ambiente. Uma diferenciação desse tipo (e este é o caso do sistema social) pode ocorrer em um âmbito indefinido de possibilidades dotadas de sentido – um âmbito que só pode ser especificado por meio da diferenciação; ou seja, pode ocorrer em um mundo que não possui outros limites. Contudo, a diferenciação também pode acontecer no interior de sistemas já formados. Desejamos designar como diferenciação dos sistemas apenas esse caso ou – se quisermos destacar a distinção mencionada – como diferenciação interna do respectivo sistema.<sup>112</sup> (Tradução nossa).

Do conceito de diferenciação funcional surge o conceito de “contingência”, o qual caracteriza-se por um âmbito indefinido de possibilidades dotadas de sentido – um âmbito que só pode ser especificado por meio da diferenciação.

No âmbito da diferenciação sistêmica, qualquer transformação se caracteriza como sendo, ao mesmo tempo, dupla ou mesmo múltipla em sua natureza. Isso ocorre porque a modificação de um sistema parcial implica, simultaneamente, mudanças no ambiente que envolve os outros sistemas parciais. Assim, cada evento ocorre de maneira repetida, com variações que dependem do sistema de referência adotado. Nas comunicações da sociedade funcionalmente diferenciada, é fundamental transmitir constantemente pontos de vista que envolvam tanto a agregação quanto a delimitação. No entanto, esse processo já não ocorre por meio de sinais perceptíveis, ou, quando ocorre, é de maneira bastante restrita.

Em questões tecnológicas, quando a ausência de conhecimento cientificamente seguro representa um risco para o investimento de capital, uma decisão adequada só é possível se houver uma compreensão precisa dessa diferença. Não basta apenas considerar a alteridade do outro; é necessário cuidar da própria diferença. A distinção em si deve orientar a operação, ou seja, essa operação deve ser definida por essa distinção específica e não por outra qualquer.

---

112LUHMANN, *Op. cit.*, p. 473.

A comunicação de uma distinção revela justamente a conexão entre os elementos que foram diferenciados. A unidade da operação e a diferença no esquema de observação devem se manifestar de forma integrada. Só assim é possível reproduzir o processo de diferenciação.

Em função disso, as formas de diferenciação social se distinguem conforme as diferenças que são impostas às observações, desde que essas se mantenham como operações que possuam a capacidade de estabelecer vínculos.<sup>113</sup>

Os algoritmos que integram sistemas de Inteligência Artificial desafiam a lógica da diferenciação funcional. Eles ampliam a contingência e introduzem uma normatividade concorrente, criando não apenas riscos concretos, mas também temores associados à imprevisibilidade e ao potencial comprometimento de direitos fundamentais.

A contingência, no sentido da teoria de Niklas Luhmann, está presente na própria criação dos algoritmos. As escolhas dos desenvolvedores, os dados utilizados para o treinamento e os parâmetros ajustados durante o processo de desenvolvimento introduzem elementos de incerteza e imprevisibilidade. No entanto, uma vez que um algoritmo é implementado, ele passa a operar de forma relativamente autônoma, reproduzindo as normas que foram incorporadas a ele.

LUHMANN<sup>114</sup> argumenta que, na sociedade, a comunicação adapta seus meios aos desafios enfrentados. Os meios de difusão, ao repetir informações e transformá-las em redundância, ampliam o alcance do público. Essa redundância, por sua vez, fortalece os laços sociais, mas não contribui para a geração de novos conhecimentos. Portanto, tanto a diferenciação funcional quanto os algoritmos de Inteligência Artificial, introduzem incertezas e imprevisibilidades, que afetam a operação dos sistemas sociais.

A contingência presente no desenvolvimento e na implementação dos algoritmos reflete a complexidade das escolhas dos desenvolvedores e dos dados utilizados, e a comunicação, ao reproduzir normas, precisa se atualizar para manter a coesão social. Dessa forma, a interação entre esses elementos, tanto nos sistemas tecnológicos quanto nos sociais, exige uma integração cuidadosa das diferenças para garantir a continuidade e a adaptação dos processos.

Na contemporaneidade, a constante demanda por novidades é atendida por esses meios, que se renovam continuamente devido à obsolescência das informações antigas. Com os meios de difusão, a redundância social acelera o tempo, mas aumenta a incerteza sobre a

---

113LUHMANN, *Op. cit.*, pp. 480-481.

114*Idem, ibidem*, p. 155.

aceitação ou rejeição das informações compartilhadas como base para ações futuras. A vasta quantidade de participantes impossibilita avaliar os reais efeitos da comunicação. Os meios de massa fomentam controvérsias e conflitos, mas não esclarecem se as mensagens foram acolhidas, rejeitadas ou simplesmente esquecidas, criando uma simulação de entendimento sem garantias de integração social. Nesse contexto, a evolução social oscila entre a estagnação e a descoberta de soluções para desafios emergentes.

A invenção da escrita, por exemplo, intensificou o rigor religioso, sendo utilizada como ferramenta de homogeneização motivacional, embora tenha exacerbado o ideal de unanimidade. Em resposta, a sociedade desenvolveu meios de comunicação simbolicamente generalizados, como os que regem a economia e o poder político. Tais meios ajustam as condições de comunicação para ampliar a aceitação, mesmo em situações adversas, como na troca de bens mediante pagamento ou na obediência a ordens estatais sob ameaça legitimada da força. Com a institucionalização desses meios, o espaço de aceitação das comunicações se expande, reduzindo rejeições e esquecimentos. Progressivamente, tornam-se pilares da autodescrição cultural, retratando a sociedade como fundamentada em consensos garantidos por códigos e programas. Contudo, ignoram as comunicações rejeitadas ou esquecidas, que permanecem à margem sob o conceito de “ignorância pluralística”.<sup>115</sup>

Em síntese, os meios de difusão aceleram o tempo e obscurecem o impacto real das informações compartilhadas, enquanto os meios de comunicação simbolicamente generalizados – como os que regulam a economia e o poder político –, ajustam as condições para ampliar a aceitação social. Esses mecanismos sustentam a narrativa de uma sociedade baseada em normas e consensos aparentes, mas deixam na obscuridade as comunicações rejeitadas ou esquecidas.

A “ignorância pluralística” descreve, portanto, a situação em que um grupo ou a sociedade como um todo compartilha uma lacuna de conhecimento sobre determinado tema, mas seus membros acreditam que essa ignorância é individual, enquanto os demais possuem maior entendimento. Assim, embora se perceba um consenso aparente – sustentado por normas sociais ou meios de comunicação –, muitas informações são coletivamente ignoradas ou esquecidas, ocultando uma falta real de compreensão. Esse fenômeno gera uma falsa percepção de concordância, mascarando lacunas significativas de conhecimento.

Diante dessas considerações, podemos dizer que a Inteligência Artificial, como um

---

115 LUHMANN, *Op. cit.*, pp. 155-156.

sistema tecnológico, introduz uma nova forma de comunicação e de normatividade que não se alinham perfeitamente com as estruturas de poder ou com as lógicas sociais já estabelecidas, gerando novas dinâmicas de poder e controle com algoritmos que funcionam de forma dispersa e que podem escapar às normas e aos códigos sociais existentes, criando uma contingência complexa e imprevisível, amplificando os riscos de abuso, exclusão e violação de direitos, enquanto, ao mesmo tempo, contribuem para uma sensação de “ignorância pluralística”, onde as pessoas acreditam que possuem algum grau de controle ou compreensão sobre essas tecnologias, quando, na realidade, há uma grande incerteza e um distanciamento do entendimento real de como elas funcionam e afetam a sociedade.

Afinal, o que realmente significa democracia? Embora o termo tenha se feito presente em diversas manifestações globais desde 2011, ele pode ser interpretado como excessivamente abstrato. De maneira geral, a democracia refere-se ao fortalecimento da participação cidadã e à ampliação dos instrumentos de democracia direta. Por trás dessa concepção, há uma crítica à representação política tradicional, frequentemente considerada incapaz de refletir de forma genuína a soberania popular.

A ideia de soberania popular está, invariavelmente, ligada a um “povo”. Contudo, é aqui que surgem as dificuldades. O “povo” não é algo fixo ou imutável que se expressa de forma direta. Ele só ganha forma por meio de discursos, os quais são, por natureza, parciais e suscetíveis a críticas. A razão disso reside no fato de que todo discurso é uma representação. A confusão surge justamente devido à multiplicidade de significados da palavra “representação”.

A representação política é realizada por representantes eleitos, que se apresentam como porta-vozes e legisladores do “povo”. No entanto, a palavra “representação” possui outros sentidos, como no contexto da arte. A diferença principal entre ambos é que, enquanto a representação política não possui um original preexistente, a representação artística, ao ser realizada, cria o objeto que pretende representar.

Os desafios da representação política se refletem em sua capacidade de abarcar a pluralidade social. Muitas vezes, os representantes acabam mais alinhados aos interesses de grupos específicos ou das elites, marginalizando as vozes dos menos favorecidos. Essa situação pode gerar desconfiança no sistema democrático e um sentimento de alienação entre os cidadãos.

Em uma sociedade saudável, as instituições e práticas sociais devem assegurar que os cidadãos possam viver sem o medo constante de serem dominados, direta ou indiretamente. A liberdade é o princípio fundante de qualquer regime democrático. Sem ela, o processo de autogoverno – essencial à definição moderna de democracia – se desmancha, cedendo lugar a formas de controle, manipulação e opressão.

A dominação, no contexto filosófico e político, refere-se a uma relação de poder em que uma parte exerce controle ou influência sobre outra, de modo que a parte dominada tem sua liberdade e autonomia comprometidas. Esse controle pode ocorrer de várias formas, como através da coerção, manipulação, dependência econômica, ou até mesmo pela imposição de normas e valores que limitam a capacidade de ação ou decisão do dominado.

Entre os pensadores contemporâneos, Philip Pettit é amplamente reconhecido por revitalizar e modernizar a teoria republicana, especialmente através de sua obra *Republicanism: A Theory of Freedom and Government*<sup>116</sup>, em que propõe uma versão neo republicana da liberdade, que ele chama de “liberdade como não-dominação”, distinguindo-a da liberdade liberal tradicional, que é frequentemente entendida como “liberdade como não-interferência”.

A dominação não se restringe apenas a situações de força física ou coerção explícita; ela também pode ocorrer de maneiras mais sutis, como em estruturas sociais ou econômicas que perpetuam desigualdades e restringem as oportunidades de certos grupos ou indivíduos.

Em geral, a dominação é vista como uma violação da liberdade porque impede que as pessoas vivam de acordo com suas próprias escolhas e interesses, sujeitando-as à vontade ou aos interesses de outros. Todavia, a dominação também pode ser entendida como uma violação ao princípio da igualdade.

A dominação abrange desigualdades estruturais, como as relacionadas a classe social, raça e gênero, que limitam as oportunidades de determinados grupos. PETTIT defende que a liberdade está intrinsecamente ligada às condições sociais e econômicas, pois práticas discriminatórias e desigualdades materiais não só violam a liberdade, mas também reforçam formas de dominação, prejudicando a igualdade.<sup>117</sup>

Uma sociedade justa deve garantir que todos os seus membros sejam igualmente protegidos contra formas de dominação – sejam estas explícitas, como coerção física, ou sutis, como a manipulação. Nessa senda, a igualdade não se limita a uma distribuição formal de

---

<sup>116</sup>PETTIT, Philip. **Republicanism**: A Theory of Freedom and Government. Oxford: Clarendon Press, 1997.

<sup>117</sup>PETTIT, *Op. cit.*, pp. 110-119.

direitos, mas requer uma igualdade material no exercício da liberdade, livre de relações desiguais de poder.

Assim como a ignorância pluralística distorce a percepção de um consenso social, a dominação também distorce a percepção da liberdade. Quando as condições de comunicação e os meios de difusão de informações não são adequadamente geridos, há uma tendência a promover uma aceitação superficial e inconsciente de normas e valores, que podem esconder uma real falta de liberdade e compreensão. Em outras palavras, a aceleração do tempo e a forma como as informações são compartilhadas e aceitas na sociedade podem ser comparáveis aos mecanismos de dominação, onde o verdadeiro entendimento e a autonomia ficam comprometidos.

Os algoritmos, ao funcionarem como mecanismos de controle (seja por meio de comunicação ou por dominação social), afetam a liberdade e autonomia dos indivíduos, limitando sua capacidade de agir conforme suas próprias escolhas, além de criar uma aparência de harmonia ou consenso que, na realidade, esconde desigualdades e distorções.

Podemos compreender o sistema democrático como uma estrutura autopoietica que gerencia tensões e paradoxos internos de maneira contínua. Esse sistema não tende somente à sobrevivência; busca também sua reprodução, adaptando-se às demandas de seus ambientes interno e externo. Dessa forma, a interação entre os elementos construtivos e legitimadores da democracia e as forças disruptivas ou de negação, evidencia a complexidade de sua operação.

Essa interação pode ser interpretada como uma manifestação dos códigos binários que estruturam o sistema, como “inclusão/exclusão” e “governo/oposição”. O ponto nodal, contudo, reside na forma como o sistema processa essas dualidades, convertendo conflitos em energia produtiva para sua evolução.

Os algoritmos de sistemas inteligentes intensificam e reconfiguram essas dinâmicas ao intermediar grande parte da comunicação no sistema democrático contemporâneo. Atuando como mediadores, tanto ampliam a circulação de ideias quanto exacerbam a polarização entre elementos construtivos e forças disruptivas, apresentando o potencial de reforçar bolhas informacionais, alimentar discursos radicais ou manipular percepções públicas, amplificando as forças que desafiam a estabilidade do sistema democrático. A tensão gerada por essas forças cria “ruídos” que o sistema precisa traduzir em comunicação válida e aprendizado.

Uma oposição política radical ou a disseminação de desinformação por meio de algoritmos pode desafiar a democracia, mas também impulsioná-la a se adaptar, fortalecendo

instituições, redefinindo normas e promovendo ajustes tecnológicos e sociais. O ambiente externo desempenha um papel igualmente decisivo. Pressões sociais, econômicas e tecnológicas, frequentemente intensificadas por algoritmos, afetam diretamente o equilíbrio do sistema. A desigualdade no acesso à informação ou a manipulação de dados por sistemas inteligentes pode intensificar a polarização e expandir o alcance das forças antidemocráticas.

Em contrapartida, a criação de mecanismos regulatórios que promovam uma interação ética com os algoritmos podem reequilibrar essas dinâmicas, gerando ciclos de renovação democrática.

O paradoxo da democracia, amplificado pela atuação dos algoritmos, reside em sua necessidade de incluir o dissenso e lidar com forças que podem ameaçar sua integridade. Essa característica paradoxal é, ao mesmo tempo, sua maior fraqueza e sua maior força. A sobrevivência do sistema democrático depende de sua capacidade de equilibrar tensões entre integração e exclusão, consenso e dissenso, continuidade e ruptura. Assim, a resiliência democrática no futuro dependerá de sua capacidade de integrar a tecnologia como uma ferramenta voltada à inclusão, à transparência e à inovação social.

Diante dos argumentos apresentados, defendemos o princípio da não-dominação algorítmica como um novo paradigma de proteção aos direitos humanos, centrada na dignidade humana, na equidade e na isonomia.

### 3.2 DIREITOS FUNDAMENTAIS E O RISCO TECNOLÓGICO

Na esfera jurídica, o termo “risco” denota perigo, a possibilidade de dano, a previsibilidade de perda ou responsabilidade pelo dano, abrangendo eventos incertos e futuros que, embora imprevistos, são temidos por sua capacidade de acarretar prejuízos ou danos.

Entendemos que o risco está ligado aos marcadores sociais da diferença, especialmente quando combinados em uma mesma experiência, o que está atualmente relacionado aos estudos de interseccionalidade. Portanto, é considerado que a vulnerabilidade está presente em situações de risco, este último caracterizado pela direta implicação nas condições precárias de vida dos seres humanos.

Um exemplo recente no Brasil merece atenção. Com um déficit de 10 mil funcionários, em 2022, a direção do Instituto Nacional do Seguro Social (INSS) decidiu fazer análises dos pedidos de concessão de benefícios por meio de Inteligência Artificial como alternativa para o

atendimento da fila para perícias de cerca de 1,8 milhão de segurados.

Todavia, o sistema de Inteligência Artificial utilizado negou mais de 300 mil benefícios, sendo a maioria assistenciais<sup>118</sup>. Observou-se que a interação humana seria indispensável à correta análise dos documentos necessários à concessão de benefícios previdenciários, isto porque o modelo de IA usada pelo instituto não era capaz de identificar que um pedido irregular poderia ser suprido pela apresentação de um documento faltante e não que o segurado não tivesse direito a receber o benefício.

O uso da Inteligência Artificial na Administração Pública levanta algumas preocupações quanto à moralidade e eticidade dessa tecnologia nas práticas de governo, já que, como no exemplo do INSS, a ausência de um operador e supervisor humano operou em uma escala de erro que produziu a negação de um direito fundamental à assistência social a milhares de beneficiários.

Impende assinalar as distintas características dos direitos fundamentais: 1) não se sujeitam à perda pelo decurso do tempo (imprescritíveis); 2) não admitem transferência (inalienáveis); 3) são, geralmente, irrenunciáveis; 4) são invioláveis, estando além do alcance de leis infraconstitucionais ou atos de autoridades públicas; 5) são universais, abrangendo todos os indivíduos; 6) são marcados pela efetividade, impondo ao Poder Público a adoção de mecanismos para assegurar sua realização; 7) são interdependentes, uma vez que as diversas previsões constitucionais, embora autônomas, se entrelaçam para alcançar suas finalidades; e 8) são complementares, exigindo uma análise conjunta e abrangente para atingir os objetivos delineados pelo legislador constituinte, não sendo passíveis de interpretações isoladas.

Os direitos fundamentais, como manifestações normativas constitucionais, têm sua base fundante no princípio da dignidade da pessoa humana, sendo assegurados pelo Estado. Assim, compete ao Estado definir medidas que evitem restrições ilegítimas quando os indivíduos exercem esses direitos.

Incorporando princípios de cunho mais social em suas declarações de direitos, a Constituição Federal de 1988 estabeleceu a dignidade da pessoa humana como um dos fundamentos da República Federativa do Brasil (art. 1º, inciso III), reservando um título próprio para abordar os direitos e garantias fundamentais.

---

118ROCHA, Rosely; MUNIZ, Marise. “INSS usa robôs para analisar pedidos de benefícios e milhares são negados”. **Central Única dos Trabalhadores-CUT**. Disponível em: <https://www.cut.org.br/noticias/inss-usa-robos-para-analisar-pedidos-de-beneficios-e-mais-de-300-mil-sao-negados-1cd3>. Acesso em: 8 mar. 2025.

Em sintonia com o padrão de outras Constituições internacionais, a Carta Magna Brasileira não descarta a possibilidade da existência de direitos não explicitamente previstos. O artigo 5º, §2º, reconhece a eventual emergência ou identificação de outros direitos fundamentais, decorrentes do sistema e dos princípios adotados pela própria Constituição Federal.

DIMOULIS & MARTINS<sup>119</sup> definem os direitos fundamentais como “direitos público subjetivos de pessoas (físicas ou jurídicas), contidos em dispositivos constitucionais e, portanto, que encerram caráter normativo supremo dentro do Estado, tendo como finalidade limitar o exercício do poder estatal em face da liberdade individual.” Parafrazeando os autores, para compreender a função dos direitos fundamentais, é essencial conceber a relação entre o Estado e cada indivíduo como a interação entre duas esferas.

Os direitos fundamentais asseguram a autonomia da esfera individual e, simultaneamente, delineiam situações que exigem certo tipo de interação, classificando-se em três categorias ou espécies, de acordo com a natureza do relacionamento entre o Estado e o indivíduo. Tal abordagem permite uma distinção conceitual entre os direitos negativos (de resistência), os direitos sociais e os direitos políticos, conforme delineado pela doutrina contemporânea. Nesse passo, a primeira categoria se constitui dos direitos negativos ou de resistência à intervenção estatal, os quais permitem aos indivíduos resistir a uma possível atuação do Estado, numa hipótese de que a esfera de atuação do Estado não deve interferir na esfera do indivíduo. A essência do direito reside na imposição direta ao Estado de abster-se de interferir. Trata-se de um direito de natureza negativa, pois impõe ao Estado o dever de não agir, ou seja, uma obrigação de omitir-se. Essa obrigação traduz-se em um imperativo de abstenção (*Unterlassungsgebot*), consistente na vedação de qualquer intervenção na esfera de liberdade assegurada pela Constituição.<sup>120</sup>

A segunda categoria formulada pelos autores, referente aos direitos de *status* positivo, também denominados direitos “sociais” ou de prestação, abrange os direitos que facultam ao indivíduo requerer uma determinada atuação por parte do Estado com o propósito de aprimorar as condições de vida, assegurando os fundamentos materiais essenciais para o exercício da liberdade, incluindo as liberdades de *status* negativo. Dessa forma, incumbe ao Estado agir conforme preconizado pela Constituição (isto é, intervir na esfera do indivíduo).

---

119DIMOULIS, Dimitri; MARTINS, Leonardo. **Teoria geral dos direitos fundamentais**. 5. ed. rev., atual. e ampl. São Paulo: Atlas, 2014, p. 116.

120Idem, *ibidem*, pp. 117-120.

De maneira correlata, o indivíduo possui o direito (positivo) de receber algo, seja material ou imaterial. A terminologia “direitos sociais” é justificada, segundo os autores, pelo seu intento de promover a melhoria de vida de amplos estratos da população, por meio de políticas públicas e ações concretas de política social. Adiante, a terceira categoria, dos direitos políticos ou de *status* ativo, oferece a possibilidade de participar na determinação da política estatal de forma ativa (o indivíduo pode interferir no Estado). Esses direitos são classificados como ativos, pois facultam ao indivíduo uma forma de “inserção” no domínio das decisões políticas conduzidas pelas autoridades estatais, permitindo que o “eu” adentre o espaço estatal. Entre os exemplos mais representativos estão o direito de eleger os representantes políticos, por meio do sufrágio, e o de contribuir diretamente para a formação da vontade política, seja por meio de referendos ou da participação em partidos políticos.<sup>121</sup>

BOBBIO<sup>122</sup> ao defender a historicidade dos direitos, destaca que são “[...] nascidos em certas circunstâncias, caracterizadas por lutas em defesa de novas liberdades contra velhos poderes, e nascidos de modo gradual, não todos de uma vez e nem de uma vez por todas”.

Sobre a dignidade da pessoa humana, SARLET<sup>123</sup> aponta que é relevante destacar – retomando a ideia central que permeava o pensamento clássico –, que a dignidade se configura como uma característica intrínseca à pessoa humana, sendo, portanto, inalienável e irrenunciável. Tal atributo qualifica o ser humano em sua essência e não pode ser separado de sua condição. Assim, não se pode admitir a hipótese de que alguém formule uma pretensão para que lhe seja conferida dignidade, uma vez que esta não se trata de algo a ser outorgado. Nesse sentido, a dignidade, enquanto qualidade inerente à própria natureza humana, deve ser reconhecida, respeitada, promovida e protegida, não podendo ser criada, concedida ou extinta, ainda que possa sofrer violações, pois é parte constituinte e inseparável de cada ser humano.

Mulheres, em todas as suas diversidades de raça/etnia, idade, orientação sexual ou classe social, formam um grupo social que historicamente tem sido subjugado ao poder masculino de maneira persistente e duradoura, devido ao seu sexo. Esta subordinação se manifesta não apenas através de leis discriminatórias, mas também por meio de práticas sociais que buscam manter uma diferença cultural e material entre homens e mulheres no que diz respeito ao poder social.

---

121DIMOULIS & MARTINS. *Op. cit.*, p. 126.

122BOBBIO, Norberto. **A Era dos Direitos**. 7a reimpressão. Tradução de Carlos Nelson Coutinho. Apresentação de Celso Lafer. Rio de Janeiro: Elsevier, 2004, p. 25.

123SARLET, Ingo Wolf. **Dignidade da Pessoa Humana e Direitos Fundamentais na Constituição Federal de 1988**. 4.ed. rev. atual. Porto Alegre: Livraria do Advogado, 2006, p. 41.

A performatividade do gênero feminino, ao ser vista como inerente à identidade da mulher, reforça as distinções entre homens e mulheres com base em estereótipos de gênero, como delineado na alínea “a” do artigo 5º do Decreto nº 4.377/2002<sup>124</sup> (Convenção CEDAW).

Essa disposição visa combater o estigma cultural que impede o pleno respeito às mulheres e reduzir as desvantagens materiais que enfrentam em razão de sua condição biológica. Para assegurar a máxima eficácia dessa norma, o cerne essencial do direito fundamental contido na alínea “a” do artigo 5º, do Decreto nº 4.377/2002, deve ser interpretado como um direito subjetivo negativo. Isso significa que as mulheres têm o direito de exigir que todos (sejam os poderes públicos ou os particulares) se abstenham de qualquer ação que resulte em distinção entre homens e mulheres baseada em funções estereotipadas de gênero.

Observa-se aqui a aplicação do princípio da vedação ao retrocesso, o qual determina que as normas de proteção, sejam domésticas ou internacionais, devem sempre garantir mais direitos às pessoas, não permitindo retrocessos na busca pela máxima efetividade dos direitos humanos.

Consideremos a interseção entre a teoria dos sistemas sociais autopoieticos<sup>125</sup> e o processo de tratamento de dados para elaboração de sistemas de Inteligência Artificial, para especular sobre como as estruturas sociais e normativas consagram a persistência da assimetria sexual e da desigualdade de gênero.

Inspirado pelas contribuições dos biólogos Humberto Maturana e Francisco Varela, Niklas Luhmann (1927-1998) introduziu a substituição da ideia de sistemas abertos pela concepção de sistemas autopoieticos, que se caracterizam por serem autorreferentes e operacionalmente fechados.

Um sistema autopoietico é um sistema que se auto-reproduz e se auto-organiza continuamente. O termo “autopoietico” deriva do grego, onde “auto” significa “próprio” e “poiese” significa “criação” ou “produção”, caracterizado por sua capacidade de produzir e manter a si mesmo, sem depender de influências externas para sua operação contínua.

A autorreferencialidade é um conceito central nessa teoria, que aborda a capacidade dos sistemas de se definirem a partir de suas próprias estruturas.

---

124BRASIL. Congresso Nacional. **Decreto nº 4.377**, de 13 de setembro de 2002. Promulga a Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher, de 1979. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/2002/d4377.htm](http://www.planalto.gov.br/ccivil_03/decreto/2002/d4377.htm). Acesso em: 8 mar. 2025.

125LUHMANN, Niklas. **Sistemas sociais**: esboço de uma teoria geral. Tradução de Antônio C. Luz Costa, Roberto Dutra Torres, Marco Antônio dos Santos Casanova. Petrópolis: Vozes, 2016.

Ou seja, os sistemas autopoieticos determinam suas operações e códigos internamente, sem depender de influências externas.

Outro conceito essencial dessa teoria é o fechamento operacional, que sugere que os sistemas autopoieticos operam de maneira autossuficiente, isolados de seu ambiente no que se refere às operações que realizam. Em outras palavras, os sistemas sociais traçam limites em relação ao seu entorno e interagem com ele com base em suas próprias regras internas. Esse mecanismo de controle sobre suas fronteiras operacionais garante a integridade e a autonomia do sistema.

Ao aplicar a teoria luhmanniana dos sistemas sociais ao patriarcado, pode-se compreendê-lo como um subsistema dentro do sistema social mais amplo, que interage com outros subsistemas como a economia, a política e a religião, trocando informações e símbolos, que perpetuam e reforçam as normas, valores e hierarquias patriarcais, consolidando sua estrutura e influência. A interconexão entre esses subsistemas revela como os sistemas sociais mantêm suas estruturas internas, ao mesmo tempo em que se adaptam e dialogam com o ambiente externo.

Para Niklas Luhmann, o ambiente compreende tudo que está externo aos sistemas sociais e não participa diretamente de suas operações autopoieticas. Nesse sentido, o patriarcado – enquanto conjunto de normas, valores e práticas – não pode ser classificado como “ambiente”, pois está intrinsecamente sobreposto nas operações de outros sistemas, como o jurídico, familiar e político. Por outro lado, segundo sua teoria, estruturas são padrões de expectativas e normas que estabilizam comunicações dentro dos sistemas.

O patriarcado atua precisamente como uma estrutura que molda interações em múltiplos sistemas sociais, definindo hierarquias e papéis. No entanto, sua natureza vai além de uma simples “influência”. Operando por meio de um código binário central (masculino/feminino), que se desdobra em subcódigos como capaz/incapaz, forte/frágil, potente/impotente, o patriarcado, como um sistema autopoietico, reproduz-se por meio de suas próprias comunicações (discursos, normas, instituições), adaptando-se a mudanças sociais sem perder sua função central: manter a dominação masculina.

Tais distinções servem para naturalizar a superioridade masculina (ex.: associar o masculino a força e capacidade), justificar a dominação masculina em esferas como família, trabalho e política, e excluir ou subordinar grupos fora do padrão masculino hegemônico.

O patriarcado não é um “efeito colateral” da sociedade – é um sistema autônomo, que se retroalimenta.

Quando criamos uma lei contra a violência doméstica, por exemplo, o patriarcado não desaparece: ele se adapta, encontrando novas formas de controle. A força do patriarcado está em sua capacidade de criar códigos invisíveis, dividindo o mundo em masculino e feminino (homem = forte, mulher = frágil), justificando a desigualdade (‘é natural’, ‘sempre foi assim’) e se disfarçando de modernidade, em empresas que usam o ‘feminismo’ para vender produtos, mas não contratam mulheres para liderança; homens ‘apoiam’ a igualdade, mas não dividem tarefas domésticas. Essa é a armadilha: o patriarcado não é estático. Ele se renova para parecer menos opressor, sem abandonar seu núcleo: a dominação masculina.

Apesar de autônomo, o patriarcado interage com outros sistemas sociais via acoplamentos estruturais. Na família, define papéis de gênero nas relações domésticas, no direito, influencia leis que reforçam hierarquias de gênero e na economia organiza a divisão sexual do trabalho e desigualdades salariais. Cada sistema processa essas influências segundo suas próprias lógicas, sem determinar as operações do patriarcado.

Os padrões normativos do subsistema patriarcal estão profundamente conectados a disfunções sociais, como a objetificação da mulher e a restrição ou relativização de sua autonomia. Estes fatores resultam em um tecido social distorcido e na deterioração da comunicação, evidenciando como os mecanismos sistêmicos moldam a esfera social e apontando para o fato de que as disfunções sociais decorrem da submissão dos campos comunicativos às exigências de estruturas formalmente organizadas, resultando na disseminação de ideologias que naturalizam a inferiorização das mulheres.

Sugere-se, portanto, que a diferenciação funcional nas sociedades modernas inclui subsistemas baseados em hierarquias de poder, além dos subsistemas tradicionais, como o direito, a economia e a política. Essa perspectiva reconhece o patriarcado como uma entidade autônoma com capacidade de adaptação e resiliência, similar a outros subsistemas. Sua persistência não é apenas um reflexo de estruturas sociais, mas resultado de sua capacidade de se autorreproduzir. Ele não é um epifenômeno de outros sistemas, mas um ator autônomo que molda e é moldado por eles.

Sua transversalidade é uma clara demonstração de sua capacidade de se infiltrar em múltiplas esferas sociais.

Entender o patriarcado como subsistema altera radicalmente as estratégias para combatê-lo. Enquanto uma estrutura pode ser desmontada por meio de reformas pontuais em sistemas específicos, um sistema autopoietico exige intervenções que ataquem sua lógica operacional interna, rompendo com os códigos do sistema através de uma reengenharia da comunicação.

Os algoritmos, assim como os sistemas sociais, exibem características autorreferenciais, criando padrões normativos internos que, sem a agência humana constante, não dependem diretamente de validações externas.

A percepção de danos futuros potencializados pelos algoritmos, projeta o risco como um elemento estruturante nas respostas jurídicas que visam proteger os direitos fundamentais e os valores coletivos em uma sociedade profundamente interconectada por sistemas algorítmicos. A sociedade contemporânea, marcada por um excesso de possibilidades auto produzidas, intensifica a complexidade social e transforma o risco em um elemento comunicativo onipresente, intrinsecamente ligado aos processos decisórios e à comunicação social que cerca sua gestão. Diante disso, o medo se manifesta como uma reação subjetiva ao risco, intensificando a percepção de vulnerabilidade em face da imprevisibilidade e da opacidade dos sistemas algorítmicos.

Os algoritmos de IA operam como sistemas fechados em termos de sua lógica interna, mas estão em constante interação com um ambiente externo mais amplo e complexo. Essa interação é essencial para o funcionamento dos algoritmos, mas também revela suas limitações, já que eles não conseguem abarcar toda a riqueza e variabilidade do ambiente. Tal dinâmica ajuda a explicar por que os algoritmos podem reproduzir vieses, falhar em contextos imprevistos ou gerar resultados que não correspondem plenamente às expectativas humanas.

Assim, a autorreferencialidade nos sistemas de Inteligência Artificial pode criar um ciclo de retroalimentação, onde os vieses inicialmente presentes nos dados de treinamento são amplificados e perpetuados pelo próprio sistema, reforçando as desigualdades e injustiças presentes na sociedade em relação às mulheres.

As estruturas sociais são padrões de expectativas que orientam as ações. No contexto da IA, os algoritmos, ao serem projetados e treinados com dados, incorporam expectativas e padrões de comportamento que podem se tornar normativos. Ou seja, eles estabelecem o que é considerado “normal” ou “esperado” dentro de um determinado sistema, influenciando decisivamente as interações humanas.

Entre as principais questões que surgem está a capacidade do Direito de perceber o risco algorítmico e de construir elementos internos para sua gestão.

Afora isso, surge a possibilidade de modificar as estruturas jurídicas, conforme as necessidades do sistema, para lidar com os desafios impostos pelas novas tecnologias.

Evidencia-se, portanto, a necessidade do Direito identificar, gerir e comunicar o risco algorítmico de maneira eficaz, enquanto enfrenta os medos que permeiam os operadores jurídicos tanto quanto a sociedade. A comunicação jurídica, em relação ao risco algorítmico, deve se expandir para incluir outras comunicações jurídicas, permitindo um processo recursivo e reflexivo dentro do próprio sistema jurídico, o qual não deve apenas reagir aos riscos algorítmicos de forma pontual, mas integrar um processo contínuo de aprendizado e adaptação, permitindo que as práticas jurídicas se ajustem de maneira mais orgânica e crítica às complexidades do mundo digital. Isso pode envolver a revisão de normas, a criação de novos mecanismos de controle e a promoção de um diálogo mais amplo entre diferentes áreas do direito e da sociedade.

### 3.3 INTELIGÊNCIA ARTIFICIAL NO BRASIL: ESTRATÉGIAS DE GOVERNANÇA, ÉTICA E DESENVOLVIMENTO SUSTENTÁVEL

No âmbito da Estratégia Brasileira para a Transformação Digital (E-Digital)<sup>126</sup>, aprovada pelo Decreto nº 9.319/2018<sup>127</sup>, a Inteligência Artificial foi destacada como de importância prioritária devido aos seus impactos no país.

O Ministério da Ciência, Tecnologia, Inovações e Comunicações, por meio da Portaria MCTIC nº 1.122/2020<sup>128</sup>, definiu a área de Inteligência Artificial como uma prioridade para projetos de pesquisa e desenvolvimento tecnológico de 2020 a 2023, resultando na formulação da Estratégia Brasileira de Inteligência Artificial (EBIA)<sup>129</sup>.

---

126BRASIL. **Estratégia Brasileira para a Transformação Digital (E-Digital)**. Disponível em: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>. Acesso em: 8 mar. 2025.

127BRASIL. Congresso Nacional. **Decreto nº 9.319**, de 21 de março de 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/decreto/D9319.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/D9319.htm). Acesso em: 17 fev. 2024.

128BRASIL. **Portaria MCTIC nº 1.122**, de 19 de março de 2020. Disponível em: [https://antigo.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria\\_MCTIC\\_n\\_1122\\_de\\_19032020.html](https://antigo.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria_MCTIC_n_1122_de_19032020.html). Acesso em: 8 mar. 2025.

129BRASIL. **Estratégia Brasileira de Inteligência Artificial – EBIA**. Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação. Julho de 2021. Disponível em: [https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento\\_referencia\\_4-979\\_2021.pdf](https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf). Acesso em: 8 mar. 2025.

Essa Estratégia tem como objetivo orientar as iniciativas do Estado Brasileiro no avanço da pesquisa, inovação e desenvolvimento de soluções em inteligência artificial, promovendo seu uso consciente, ético e para o benefício de um futuro melhor.

Para esse fim, a EBIA estabelece nove eixos temáticos como base do documento, assim apresentados:

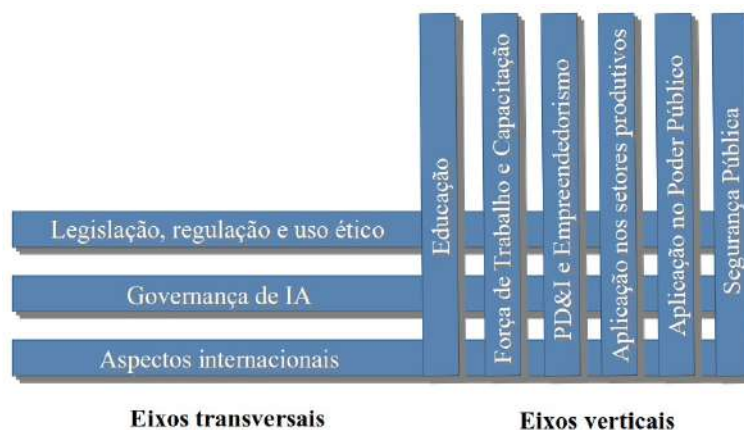


Figura 1: *Op. cit.*, p. 6

A Estratégia Brasileira de Inteligência Artificial oferece uma análise da situação atual da IA no Brasil e no mundo, destaca os desafios a serem enfrentados, propõe uma visão de futuro e apresenta um conjunto de ações estratégicas para alcançar essa visão, tendo como ponto de partida “a definição de objetivos estratégicos que levam em consideração todo o ecossistema tecnológico, e que poderão posteriormente ser desdobrados em ações específicas.”

Estabelece, para tanto, os seguintes objetivos:

- Contribuir para a elaboração de princípios éticos para o desenvolvimento e uso de IA responsáveis;
- Promover investimentos sustentados em pesquisa e desenvolvimento em IA;
- Remover barreiras à inovação em IA;
- Capacitar e formar profissionais para o ecossistema da IA;
- Estimular a inovação e o desenvolvimento da IA brasileira em ambiente internacional;
- Promover ambiente de cooperação entre os entes públicos e privados, a indústria e os centros de pesquisas para o desenvolvimento da Inteligência Artificial.<sup>130</sup>

Uma das avaliações de impacto mais relevantes no cenário nacional, estipulada pela Lei Geral de Proteção de Dados Pessoais, é conhecida como Relatório de Impacto sobre a

<sup>130</sup>BRASIL. EBIA. *Op. cit.*, p. 7.

Proteção de Dados Pessoais (RIPD)<sup>131</sup>. A análise abrange os processos de manipulação de dados pessoais que possam representar ameaças às liberdades individuais e aos direitos fundamentais. O propósito é examinar as salvaguardas e os mecanismos de mitigação de risco, conforme definido no artigo 5º, inciso XVII, da LGPD.

Por fim, recentemente o Brasil também adotou a Avaliação de Impacto Regulatório (AIR), que foi instituída pelo art. 5º, da Lei de Liberdade Econômica (Lei nº 13.874/2019<sup>132</sup>), pelo art. 6º, da Lei Geral das Agências Reguladoras Federais (Lei nº 13.848/2019<sup>133</sup>), e regulamentada por meio do Decreto nº 10.411/2020<sup>134</sup>.

A AIR é um instrumento de gestão e implementação de políticas públicas<sup>135</sup> que tem por objetivo realizar uma avaliação prévia à edição de atos normativos para verificar a razoabilidade do seu impacto, bem como subsidiar a tomada de decisão (art. 2º, I, do Decreto nº 10.411/2020).

VIEIRA & BARRETO (2019, pp. 64-68) apresentam importantes contribuições no contexto desta pesquisa, destacando as iniciativas brasileiras de boa governança.

Em 2007, foi criado o Programa de Fortalecimento da Capacidade Institucional para Gestão em Regulação (PRO-REG), por meio do Decreto nº 6.062/2007<sup>136</sup>, alterado pelo Decreto nº 8.760/2016<sup>137</sup>.

O objetivo principal do PRO-REG é aprimorar a qualidade da regulação no governo federal, fortalecendo o sistema regulatório e facilitando o exercício pleno das funções pelos diversos atores envolvidos, ao mesmo tempo em que busca aprimorar a coordenação entre as instituições participantes, os mecanismos de prestação de contas e a participação da sociedade civil no processo de monitoramento.

---

131 Importante frisar que o RIPD ainda configura como objeto de futura regulamentação pela ANPD, tendo havido tomada de subsídios, mas sem previsão de publicação da minuta da resolução para consulta pública.

132 BRASIL. Congresso Nacional. **Lei nº 13.874**, de 20 de setembro de 2019. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/lei/113874.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/113874.htm). Acesso em: 8 mar. 2025.

133 BRASIL. Congresso Nacional. **Lei nº 13.848**, de 25 de junho de 2019. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/lei/113848.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/113848.htm). Acesso em: 8 mar. 2025.

134 BRASIL. Congresso Nacional. **Decreto nº 10.411**, de 30 de junho de 2020. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2020/decreto/d10411.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/d10411.htm). Acesso em: 8 mar. 2025.

135 BRASIL. Ministério da Economia. “Normativos da União deverão ter análise de impacto regulatório para serem editados”. Disponível em: <https://www.gov.br/economia/pt-br/assuntos/noticias/2020/julho/normativos-da-uniao-deverao-ter-analise-de-impacto-regulatorio-para-serem-editados>. Acesso em: 8 mar. 2025.

136 BRASIL. Congresso Nacional. **Decreto nº 6.062**, de 16 de março de 2007. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2007/decreto/d6062.htm](https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6062.htm). Acesso em: 8 mar. 2025.

137 BRASIL. Congresso Nacional. **Decreto nº 8.760**, de maio de 2016. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2016/decreto/d8760.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8760.htm). Acesso em: 8 mar. 2025.

Inicialmente, o programa concentrou-se no diagnóstico do ambiente regulatório brasileiro e na capacitação para melhorar a qualidade regulatória. Com esse objetivo, o PRO-REG desenvolveu diversas ações, incluindo:

- (a) Disseminação de boas práticas regulatórias no âmbito da administração pública federal;
- (b) Fomento e difusão de iniciativas regulatórias para aprimorar a abordagem estratégica nas decisões sobre políticas públicas e regulatórias;
- (c) Consolidação e expansão do uso da Análise de Impacto Regulatório (AIR);
- (d) Implementação de ações voltadas à gestão do estoque regulatório, à promoção da transparência, controle social e responsabilização no processo regulatório;
- (e) Ampliação do diálogo sobre qualidade regulatória com diversos atores do sistema regulatório;
- (f) Realização de programas de treinamento e capacitação sobre qualidade regulatória para os principais envolvidos no processo.

Em 2011, foi sancionada a Lei de Acesso à Informação (Lei nº 12.527/2011<sup>138</sup>), que regulamenta o direito fundamental de acesso às informações públicas no Brasil, incentivando a criação de uma cultura de transparência. Entre suas diretrizes, destacam-se:

- (a) A observância da publicidade como princípio geral e do sigilo como exceção;
- (b) A divulgação de informações de interesse público, independentemente de solicitações;
- (c) O uso de meios de comunicação viabilizados pela tecnologia da informação;
- (d) O fomento ao desenvolvimento da cultura de transparência na administração pública;
- (e) O desenvolvimento do controle social sobre a administração pública.

No modelo regulatório, quatro projetos de lei tramitaram no Brasil: um de iniciativa da Câmara dos Deputados – PL 21/2020<sup>139</sup> que, conforme seu histórico, foi apresentado em fevereiro de 2020 propondo na ementa: “estabelece princípios, direitos e deveres para o uso de IA no Brasil”; e dois de iniciativa do Senado Federal, o PL 5051/2019<sup>140</sup> (“estabelece os princípios para o uso da IA no Brasil”) e o PL 5691/2019<sup>141</sup> (“institui a Política Nacional de Inteligência Artificial”).

---

138BRASIL. Congresso Nacional. **Lei nº 12.527**, de 18 de novembro de 2011. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Acesso em: 8 mar. 2025.

139BRASIL. Câmara dos Deputados. **Projeto de Lei 21/2020**. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 8 mar. 2025.

140BRASIL. Congresso Nacional. **Projeto de Lei 5051/2019**. Disponível em: <https://www.congressonacional.leg.br/materias/materias-bicameras/-/ver/pl-5051-2019>. Acesso em: 8 mar. 2025.

141BRASIL. Senado Federal, **Projeto de Lei 5691/2019**. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 8 mar. 2025.

Em relação aos projetos que tramitaram no Senado Federal, o PL 5051/2019 buscava estabelecer os princípios para o uso da IA no Brasil, tendo em vista o reconhecimento do bem-estar humano.

No Art. 2º do texto inicial: (I) o respeito à dignidade humana, à liberdade, à democracia e à igualdade, (II) o respeito aos direitos humanos, à pluralidade e à diversidade; (III) a garantia da proteção da privacidade e dos dados pessoais; (IV) a transparência, a confiabilidade e a possibilidade de auditoria dos sistemas; (V) a supervisão humana. E a disciplina do uso tendo por objetivo a promoção e a harmonização da valorização do trabalho humano e o desenvolvimento econômico (Art. 3º). Também definia a responsabilidade civil por danos decorrentes do uso de IA na figura do seu supervisor (Art. 4º, § 2º), tendo em vista que a lei determina a tomada de decisões por uma IA como auxiliar da decisão humana (Art. 4º), ou seja, é necessária e fundamental a supervisão humana desses instrumentos, reconhecendo suas implicações. Importante nesse projeto, também, é a formação de diretrizes para a atuação da Administração Pública em (I) promover educação sobre a IA; (II) políticas de proteção e qualificação do trabalho humano; (III) a garantia da adoção gradual da IA; (IV) a ação proativa na regulação das aplicações de IA.

O PL 5691/2019, em seu texto inicial, da mesma maneira que o outro PL da Casa, não desenvolve sobre os conceitos como o PL da Câmara, restringindo-se aos princípios e diretrizes da pretendida Política Nacional de Inteligência Artificial, a qual busca implementar uma política para o desenvolvimento tecnológico aliado à transparência democrática das decisões baseadas em IA, até mesmo as autônomas, e integra um processo de transição para os mecanismos de automação, mitigando prejuízos ao emprego (art. 3º, XI), que se relaciona diretamente com a responsabilidade humana e das organizações que desenvolvem e operam IA, obrigando-as a estabelecer relações abertas, inteligíveis e debatidas sobre o seu uso, sustentando formas de decisões rastreáveis e princípios de governança ligados à proteção dos riscos ligados à IA e a proteção de dados pessoais – tendo em vista o cumprimento da LGPD<sup>142</sup>.

Destacamos a importância do tema com a criação da Comissão de Juristas pelo Ato nº 4, de 2022, do Presidente do Senado<sup>143</sup>, a qual teve como objetivo auxiliar na elaboração de um

142BRASIL. Congresso Nacional. **Lei nº 13.709**, de 14 de agosto de 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/ato2015-2018/2018/lei/113709.htm](https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/113709.htm). Acesso em: 8 mar. 2025.

143BRASIL. Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo Sobre Inteligência Artificial no Brasil. **Relatório Final**. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/2-relatorio-final-versao-completa-cjsubia.pdf>. Acesso em: 8 mar. 2025.

substitutivo para guiar a análise dos Projetos de Lei nº 5.051/2019, 21/2020 e 872/2021, os quais buscam estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da Inteligência Artificial no país. Em síntese, a comissão descreve procedimentos metodológicos e destaca o caráter público da avaliação de impacto algorítmico, podendo até mesmo criar uma base de dados correspondente.

Em maio de 2023, a Comissão apresentou Relatório Final de conclusão dos trabalhos, o qual incluiu um anteprojeto de lei que foi convertido no PL 2.338/2023<sup>144</sup>, o qual estabelece normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável da IA, apresentado pelo Senador Rodrigo Pacheco (PSD-MG), Presidente do Senado Federal.

O projeto de lei, em seu art. 2º, dispõe sobre o desenvolvimento, a implementação e o uso de sistemas de Inteligência Artificial no Brasil, tendo como fundamentos:

- I. a centralidade da pessoa humana;
- II. o respeito aos direitos humanos e aos valores democráticos;
- III. o livre desenvolvimento da personalidade;
- IV. a proteção ao meio ambiente e o desenvolvimento sustentável;
- V. a igualdade, a não discriminação, a pluralidade e o respeito aos direitos trabalhistas;
- VI. o desenvolvimento tecnológico e a inovação;
- VII. a livre iniciativa, a livre concorrência e a defesa do consumidor;
- VIII. a privacidade, a proteção de dados e a autodeterminação informativa;
- IX. a promoção da pesquisa e do desenvolvimento com a finalidade de estimular a inovação nos setores produtivos e no poder público; e
- X. o acesso à informação e à educação, e a conscientização sobre os sistemas de Inteligência Artificial e suas aplicações.

Na Exposição de Motivos, o grupo de notáveis juristas aponta suas preocupações em abrir-se “para o diálogo com a sociedade a partir de uma perspectiva multissetorial e multidisciplinar, bem como local, regional e global, com o objetivo de reunir visões plurais.”

Nessa quadra, esse novo marco legal tem um duplo objetivo. De um lado, estabelecer direitos para proteção do elo mais vulnerável em questão, a pessoa natural que já é diariamente impactada por sistemas de inteligência artificial, desde a recomendação de conteúdo e direcionamento de publicidade na Internet até a sua análise de elegibilidade para tomada de crédito e para determinadas políticas públicas. De outro lado, ao dispor de ferramentas de governança e de um arranjo institucional de fiscalização e supervisão, criar condições de previsibilidade acerca da sua interpretação e, em última análise,

---

144BRASIL. Senado Federal. **Projeto de Lei nº 2.338/2023**. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 8 mar. 2025.

segurança jurídica para inovação e o desenvolvimento econômico-tecnológico.

A Comissão de Juristas tem como intenção harmonizar uma estratégia que leve em consideração os riscos, ao mesmo tempo em que adota um enfoque regulatório fundamentado em princípios de direitos. Isso inclui a previsão de mecanismos de governança que garantam transparência e reconheçam os esforços das entidades econômicas que demonstram uma gestão eficaz dos riscos associados à criação e implementação de sistemas de Inteligência Artificial.

Nesse movimento pendular e dinâmico de categorização de riscos e intensificação de direitos e deveres, vale ressaltar o reconhecimento e internalização de que a realidade brasileira está permeada por desigualdades e assimetrias estruturais – o racismo sendo uma central entre elas e essencial neste contexto. A esse respeito, além de adotar definições sobre discriminação direta e indireta – incorporando, assim, definições da Convenção Interamericana contra o Racismo, promulgada com status de grupos (hiper)vulneráveis tanto para a qualificação do que venha ser um sistema de alto risco como para o reforço de determinados direitos.

A Avaliação de Impacto Algorítmico (AIA) envolve, fundamentalmente, o compromisso das agências públicas em avaliar os efeitos das tecnologias algorítmicas segundo critérios de equidade, justiça e mitigação de vieses, especialmente quanto aos impactos sobre grupos sociais minoritários. Esse processo constitui um instrumento essencial para a promoção de uma governança tecnológica ética e inclusiva.

Com vistas a atender esse compromisso, a necessidade da avaliação de impacto algorítmico é mencionada no artigo 22, que a torna obrigatória para os agentes de Inteligência Artificial quando o sistema é considerado de alto risco durante a avaliação preliminar.

O parágrafo primeiro do artigo 24 detalha os elementos essenciais a serem considerados nessa avaliação, desde os riscos associados ao sistema até as medidas de transparência para o público. Outrossim, o parágrafo segundo ressalta a importância do princípio da precaução ao utilizar sistemas que possam causar impactos irreversíveis, exigindo que a avaliação leve em conta evidências incipientes, incompletas ou especulativas.

Por fim, os parágrafos terceiro e quarto conferem à autoridade competente a prerrogativa de estabelecer outros critérios e elementos para a avaliação de impacto, incluindo a participação dos diferentes segmentos sociais afetados e a regulamentação da periodicidade de atualização, levando em consideração o ciclo de vida dos sistemas de alto risco e os campos de aplicação.

Dessa forma, foram estabelecidas duas etapas de avaliação: a preliminar, a ser conduzida pelo fornecedor da tecnologia, e a de impacto algorítmico, mais abrangente e obrigatória para sistemas classificados como de alto risco.

O texto fornecido pela Comissão de Juristas estabelece, no artigo 24, que a metodologia da avaliação de impacto incluirá, no mínimo, as seguintes etapas: I – preparação; II – compreensão do risco; III – redução dos riscos identificados; IV – monitoramento.

No contexto da proposta de lei brasileira mais desenvolvida até o momento, representada pelo PL 2.338/2023, adotou-se uma abordagem semelhante à União Europeia em relação à classificação de riscos, com diretrizes específicas para aplicações de alto risco e risco excessivo. Contudo, em comparação com a abordagem europeia, a escolha do termo “risco excessivo” em vez de “risco inaceitável” sugere uma atenuação da proteção contra os danos algorítmicos, como evidenciado nas nuances do projeto de lei. Isso inclui a falta de uma lista mais abrangente de aplicações de risco excessivo, como sistemas de reconhecimento de emoções ou traços de personalidade, policiamento preditivo ou armas autônomas letais, além de disposições menos rigorosas para o uso de vigilância biométrica à distância.

A obrigação de conduzir avaliações de impacto algorítmico sobre os direitos humanos é abordada na Seção III do projeto de lei, que estipula que essa avaliação é obrigatória para os agentes de Inteligência Artificial quando o sistema for considerado de alto risco após uma avaliação preliminar. Os aspectos a serem analisados e registrados incluem os riscos conhecidos e previsíveis associados ao sistema de Inteligência Artificial, o número de pessoas potencialmente afetadas, a gravidade das consequências e os esforços necessários para mitigação, bem como informações sobre a lógica de funcionamento do sistema e seu histórico de testes e avaliação de medidas de mitigação de impactos nos direitos.

No ordenamento jurídico brasileiro, a adoção de relatórios de impacto não é uma prática estranha. Na esfera ambiental, a Política Nacional de Meio Ambiente já prevê a avaliação de riscos ambientais como um de seus instrumentos (de acordo com o artigo 9º, da Lei nº 6.938, de 31 de agosto de 1981<sup>145</sup>), com a intenção de prevenir potenciais danos ao meio ambiente decorrentes da atividade empresária.

Mais recentemente, a Lei Geral de Proteção de Dados também introduziu o relatório de impacto à proteção de dados pessoais, destinado a descrever os “processos de tratamento de dados pessoais que possam acarretar riscos às liberdades individuais e aos

---

145BRASIL. Congresso Nacional. **Decreto nº 9.319**, de 21 de março de 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/decreto/D9319.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/D9319.htm). Acesso em: 8 mar. 2025.

direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco” (artigo 5º, XVII, da Lei nº13.709/2018<sup>146</sup>).

Quanto ao conteúdo da Avaliação de Impacto Algorítmico (AIA), conforme esclarecido por um relatório do Instituto de Pesquisa Data & Society<sup>147</sup> (2021), esta deve abordar três questões fundamentais: “o que um sistema faz; quem tem autoridade para intervir no que o sistema faz; e quem deve tomar decisões sobre o que o sistema pode fazer”. Assim, o objetivo da AIA é fornecer transparência sobre o funcionamento e os propósitos do sistema algorítmico, registrar os impactos e medidas de prevenção de danos, e identificar os responsáveis, possibilitando posterior fiscalização pelas autoridades competentes e, idealmente, pela própria sociedade.

### 3.4 PRÁTICAS DE GOVERNANÇA E GESTÃO DE RISCOS NA IMPLEMENTAÇÃO DE ALGORITMOS

A gestão de algoritmos pode abranger uma variedade de perspectivas, desde abordagens estritamente legais e regulatórias até uma abordagem puramente técnica. Em geral, a gestão prioriza a responsabilização, transparência e garantias técnicas, com a escolha da abordagem dependendo de fatores como a natureza do algoritmo, seu contexto e uma análise de riscos. Essa gestão se refere às práticas e políticas que visam controlar e regular os algoritmos, assim como seus impactos sobre indivíduos e a sociedade. Ela se baseia em ferramentas que operam tanto no algoritmo em si quanto em seu ambiente, incluindo conjuntos de dados e elementos de design.

A governança algorítmica deve ser uma parte essencial da proteção contra vieses, regulando o uso de algoritmos em diferentes contextos, desde plataformas digitais até instituições públicas. Diferente da ideia tradicional de governança, que envolve apenas normas e regulações, ela precisa se basear na interação entre diversos atores e no uso de processos computacionais para coordenar decisões. Seu objetivo será garantir que os algoritmos operem de maneira transparente, justa e responsável, minimizando impactos negativos e promovendo o bem-estar coletivo.

---

146BRASIL. Congresso Nacional. **Lei nº 13.709**, de 14 de agosto de 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em: 8 mar. 2025.

147MOSS, Emanuel, WATKINS, Elizabeth, SINGH, Ranjit, ELISH, Madeleine Clare, METCALF, Jacob. **Assembling Accountability**: Algorithmic Impact Assessment for the Public Interest (June 29, 2021). Disponível em: <http://dx.doi.org/10.2139/ssrn.3877437>. Acesso em: 8 mar. 2025.

Um dos elementos obrigatórios a serem estabelecidos nessa governança será a de compreender como os algoritmos funcionam e influenciam a sociedade e garantir sua transparência. Outro aspecto fundamental será a responsabilização: quem responderá por erros ou injustiças cometidas por sistemas automatizados? Ademais, a regulação deverá ser adaptada a diferentes setores, como saúde, segurança e comunicação, garantindo que o uso da inteligência artificial esteja alinhado com os interesses coletivos.

A inclusão da perspectiva de gênero será um aspecto essencial na governança algorítmica. Por essa razão, será fundamental implementar estratégias para identificar e corrigir essas distorções, promovendo um ambiente mais equitativo. Nas plataformas digitais, isso significará impedir que os algoritmos reforcem estereótipos de gênero em anúncios, recomendações de conteúdo ou filtros automáticos. No setor público, será essencial que sistemas usados em políticas sociais garantam um tratamento justo para mulheres e outros grupos vulneráveis.

A gestão de risco algorítmico será uma parte vital desse processo, tratando de implementar um conjunto de práticas com vistas a prever e mitigar os impactos negativos do uso de algoritmos. Isso envolve identificar riscos, como vieses discriminatórios e decisões equivocadas, medir os efeitos dessas falhas por meio de indicadores de precisão e justiça social, e aplicar soluções como auditorias, testes de robustez e diversificação dos dados de treinamento. O monitoramento contínuo também é indispensável para garantir que os padrões éticos e regulatórios sejam cumpridos ao longo do tempo.

Ao incorporar a perspectiva de gênero, a gestão de risco algorítmico se tornará uma importante ferramenta para combater desigualdades. Isso significa, por exemplo, evitar que algoritmos favoreçam perfis específicos em processos de seleção para empregos ou concessão de crédito. Também implica considerar as necessidades de mulheres em situação de vulnerabilidade, como mães solo, garantindo que tenham acesso a programas de assistência social.

Uma solução importante será o desenvolvimento de sistemas de aprendizado contínuo, que serão capazes de incorporar dados mais representativos e reduzir distorções ao longo do tempo. Também será necessário definir limites claros para o uso de determinadas variáveis, como estado civil e endereço, prevenindo decisões discriminatórias.

Tais iniciativas devem caminhar em direção a construir uma governança algorítmica mais justa e inclusiva, garantindo que a tecnologia beneficie a sociedade como um todo,

sem reproduzir desigualdades.

Embora governança e gestão de risco sejam conceitos distintos, são complementares e se fortalecem mutuamente. A governança estabelece diretrizes normativas e éticas, enquanto a gestão de risco as traduz em ações práticas e contínuas. Assim, governança e gestão de risco não apenas minimizam os impactos negativos, mas também maximizam os benefícios dos algoritmos.

Nos debates legislativos em diversos países, os relatórios de avaliação de impacto algorítmico (AIA) são percebidos pela maioria da comunidade técnica e pela sociedade civil como uma ferramenta essencial para promover transparência e controle social equilibrados, sem inibir a inovação e o potencial econômico.

### 3.4.1 Princípios Fundamentais da Proteção no Cenário Jurídico Global

Com o avanço da Inteligência Artificial (IA) e o surgimento de desafios éticos, a Comissão Europeia apresentou, em 2019, as Diretrizes Éticas para a Inteligência Artificial Confiável<sup>148</sup>, relacionados aos seus benefícios e potenciais riscos, como parte da Estratégia Europeia com o objetivo de garantir o uso responsável da tecnologia, promovendo a proteção de grupos vulneráveis e a cooperação internacional.

A iniciativa parte da preocupação de que essa tecnologia, ao mesmo tempo em que traz benefícios para os indivíduos e para a sociedade, também apresenta erros, riscos e impactos negativos que podem ser difíceis mensurar e identificar, constituindo um conjunto de princípios e orientações desenvolvidos pela Comissão Europeia para promover o uso ético e responsável da Inteligência Artificial (IA) na Comunidade. Desde então têm tido destaque no enfoque da UE à IA, tendo como objetivos a responsabilidade e a cooperação internacional.

O Guia é direcionado a todos que tenham interesse ou estejam envolvidos de alguma forma no design, desenvolvimento, implementação ou utilização de sistemas de Inteligência Artificial. Além do mais, também é relevante para aqueles que podem ser afetados por esses sistemas. Suas recomendações e requisitos possuem ampla abrangência, abarcando empresas, coordenadores de projetos, pesquisadores, serviços públicos, entidades governamentais,

---

148UNIÃO EUROPEIA. European Commission. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. Communication: **Building Trust in Human Centric Artificial Intelligence**. Brussels. 8.4.2019. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>. Acesso em: 8 mar. 2025.

agências, instituições, organizações da sociedade civil, indivíduos, profissionais e consumidores.

No espírito das preocupações voltadas à equidade, não discriminação e solidariedade, o Guia prioriza a proteção de dois grupos sociais específicos. De um lado, estão os considerados vulneráveis – crianças, pessoas com deficiência, minorias étnicas e outros grupos que, ao longo da história, enfrentam desvantagens estruturais ou o risco de exclusão social. De outro, encontram-se os indivíduos submetidos a relações assimétricas, como empregados e consumidores.

O propósito central do Guia é prevenir injustiças oriundas de vieses que possam prejudicar esses grupos. Para tanto, requer que os dados utilizados no treinamento de sistemas de Inteligência Artificial (IA) reflitam ampla representatividade, contemplando a diversidade das comunidades e populações.

A compreensão do Guia passa necessariamente pelos quatro princípios éticos que o sustentam: (1) respeito à autonomia humana, (2) prevenção de danos, (3) justiça e (4) explicabilidade. O respeito à autonomia humana reafirma o valor da dignidade intrínseca de cada pessoa, reconhecendo-a como sujeito moral e não como objeto de análise ou manipulação. Esse princípio rejeita o uso da IA para coerção, condicionamento ou submissão, assegurando que a tecnologia proteja o ser humano em sua integridade física, mental, cultural e decisória. Também é essencial para garantir acesso igualitário aos direitos, benefícios e oportunidades proporcionados pelos sistemas de IA.

O Guia destaca a necessidade de controle sobre riscos à saúde mental, vigilância indevida e manipulações injustificadas, propondo que os sistemas de IA sejam desenvolvidos para complementar e potencializar as capacidades humanas, respeitando os princípios do design centrado no ser humano. Essa abordagem exige supervisão contínua durante todo o ciclo de vida dos sistemas, garantindo que decisões significativas sejam mantidas sob o controle dos seres humanos.

O princípio da prevenção de danos está intimamente ligado à robustez técnica e segurança. No que diz respeito à justiça, esse princípio se desdobra em duas dimensões: substancial e procedimental. Na dimensão substancial, busca-se uma distribuição justa de custos e benefícios, protegendo contra preconceitos, discriminações e estigmatizações, além de garantir a proporcionalidade e o equilíbrio entre os meios e fins.

Na dimensão procedimental, assegura-se o direito de contestar decisões tomadas por sistemas de IA, com transparência e identificação clara da autoridade responsável.

Esse aspecto está relacionado ao princípio da explicabilidade, que visa garantir a confiança nos sistemas, permitindo que suas decisões sejam compreensíveis e passíveis de questionamento. Nos casos em que o sistema opera como uma “caixa-preta”, é imprescindível a adoção de mecanismos adicionais, como rastreabilidade e auditabilidade, ajustados conforme o contexto e a gravidade das possíveis consequências.

Além dos quatro princípios éticos, o Guia baseia-se em sete requisitos fundamentais, aplicáveis durante todo o ciclo de vida dos sistemas de IA: (I) agência e supervisão humana, (II) robustez técnica e segurança, (III) privacidade e governança de dados, (IV) transparência, (V) diversidade, não discriminação e justiça, (VI) bem-estar ambiental e social e (VII) prestação de contas. Esses requisitos refletem a ideia de responsabilidade compartilhada entre os desenvolvedores – responsáveis pela concepção e implementação dos sistemas – e os implantadores, que os utilizam em seus negócios e operações. A agência e a supervisão humana, em particular, exigem que os sistemas de IA promovam uma sociedade democrática e respeitem os direitos fundamentais. Desde a fase de desenvolvimento, é essencial realizar avaliações contínuas de riscos e instituir canais para acolher *feedback* externo, especialmente em casos em que a IA possa afetar direitos fundamentais. Ademais, os usuários devem estar aptos a tomar decisões autônomas e informadas sobre o uso da tecnologia, compreendendo seu funcionamento e podendo contestá-la, quando necessário.

Especial atenção deve ser dada a sistemas que influenciem comportamentos humanos por meio de técnicas difíceis de identificar, particularmente em processos subconscientes. É indispensável garantir o direito de não ser submetido a decisões exclusivamente automatizadas que possam causar impactos jurídicos ou danosos, reafirmando a centralidade do ser humano frente à tecnologia.

No que diz respeito à supervisão, o Guia destaca a importância da implementação de mecanismos de governança, alguns dos quais são explicitamente mencionados. Estes incluem:

- *Human-in-the-loop*<sup>149</sup> (HITL), que se refere à capacidade de intervenção humana em cada ciclo de decisão do sistema. Entende-se que isso pode não ser sempre possível ou desejável.
- *Human-on-the-loop*<sup>150</sup> (HOTL), que diz respeito à capacidade de intervenção humana durante o ciclo de design do sistema e no monitoramento de sua operação.

---

149 *Human-in-the-loop* (HITL): “Humano na malha” ou “Humano no ciclo”, em tradução livre.

150 *Human-on-the-loop* (HOTL): “Humano no ciclo, porém não na malha”, em tradução livre.

- *Human-in-command*<sup>151</sup> (HIC), que aborda a capacidade de supervisão abrangente da atividade do sistema, incluindo seus impactos econômicos, sociais, jurídicos e éticos. Isso inclui a habilidade de decidir quando e como utilizar o sistema em determinadas situações, podendo até mesmo envolver a decisão de não utilizá-lo em determinado contexto. Também se destaca a capacidade de estabelecer níveis de discricção humana durante o uso ou de contornar as decisões do sistema.

Em relação à explicabilidade, sua dimensão técnica requer que as decisões dos sistemas de Inteligência Artificial sejam compreensíveis e rastreáveis por seres humanos. As informações precisam ser transmitidas levando em consideração o conhecimento da pessoa afetada. Mesmo que ocorram *trade-offs*<sup>152</sup> entre explicabilidade e precisão do sistema, a necessidade de explicabilidade não pode ser negligenciada.

O Guia sublinha a importância de compreender a influência da IA nas decisões e escolhas estratégicas das organizações, apontando que a explicabilidade deve abranger tanto a lógica do sistema quanto o modelo de negócios que o sustenta. Também estabelece que os sistemas de IA não devem se passar por seres humanos, garantindo que os usuários saibam que estão interagindo com máquinas e possam tomar decisões informadas. A diversidade, não discriminação e justiça devem ser incorporadas para evitar vieses em todas as fases do ciclo de vida da IA, com supervisão rigorosa no desenvolvimento, coleta de dados e programação de algoritmos. Outrossim, o Guia ressalta a importância de incluir diversas perspectivas, respeitar as individualidades dos usuários e promover o bem-estar social e ambiental. Exige ainda auditorias detalhadas sobre dados, algoritmos e processos, para garantir responsabilidade sem comprometer informações sensíveis.

A responsabilidade inclui minimizar danos, relatar impactos negativos e avaliar riscos éticos, com auditorias independentes quando se tratam de direitos fundamentais. O Guia adverte que, se os compromissos éticos não forem atendidos, o uso da IA não deve ser autorizado.

Por fim, o Guia destaca a responsabilidade central do tomador de decisão ao enfrentar situações de *trade-offs*. Esse indivíduo deve tratar tais dilemas com responsabilidade, assegurando que as decisões tomadas sejam continuamente revisadas quanto à sua adequação.

---

151 *Human-in-command* (HIC): “Humano no comando”, em tradução livre.

152 A expressão *trade-offs* pode ser traduzida para o português como “compromissos” ou “trocas”. No contexto, ela se refere à situação em que você precisa fazer uma escolha entre duas ou mais opções, sabendo que ao optar por uma delas, você terá que abrir mão de algum benefício ou característica da outra opção. Resumidamente, *trade-offs* envolve a ideia de equilibrar benefícios e sacrifícios ao fazer uma escolha.

Além disso, é imperativo que ajustes necessários ao sistema sejam prontamente implementados, sempre que se fizerem exigidos pelas circunstâncias.

A *accountability*<sup>153</sup> também fala da capacidade do sistema de corrigir eventuais impactos adversos e injustos assim que são identificados.

As Diretrizes da Comissão Europeia configuram-se como um referencial indispensável para tratar das questões críticas relacionadas à crescente aplicação da Inteligência Artificial e servem como um ponto de partida fundamental para os debates, especialmente no que concerne à proteção dos direitos das pessoas afetadas por esses sistemas – com destaque para aquelas em que haja situação de vulnerabilidade –, bem como para a definição clara dos deveres e das responsabilidades atribuídos aos desenvolvedores e implementadores dessas tecnologias.

A dimensão ética da IA não é um recurso de luxo ou um complemento: ela precisa ser uma parte integral do desenvolvimento da IA. Ao buscar uma IA centrada no ser humano, baseada na confiança, protegemos o respeito por nossos valores sociais fundamentais e estabelecemos uma marca distintiva para a Europa e sua indústria como líder em IA de ponta que possa ser confiável em todo o mundo.<sup>154</sup> (Tradução nossa).

A UNESCO aprovou a Resolução 41 C/73<sup>155</sup>, de 22 de novembro de 2021, adotando o primeiro conjunto de regras globais sobre a ética na Inteligência Artificial (IA), na sua 41ª Conferência Geral, que aconteceu em Paris.

A Recomendação salienta a proteção de dados pessoais, promovendo maior transparência e controle para os titulares. Proíbe o uso de IA para pontuação social e vigilância em massa, por considerar tais práticas como violações de direitos humanos e liberdades fundamentais. Por outro lado, sugere que os Estados-membros garantam que a responsabilidade final permaneça com os seres humanos, vedando a atribuição de personalidade jurídica às tecnologias. Em termos de implementação, recomenda a criação de

---

153 *Accountability* pode ser traduzida como “responsabilidade” ou “prestação de contas”. Ela se refere à obrigação de indivíduos ou organizações serem responsáveis por suas ações, decisões e seus impactos, e de serem capazes de prestar contas apropriadas por eles. Em relação aos sistemas de Inteligência Artificial, a *accountability* implica que aqueles que desenvolvem e implementam esses sistemas são responsáveis por garantir que eles operem de maneira ética e justa, e devem ser capazes de explicar e corrigir qualquer impacto adverso que possam causar.

154 UNIÃO EUROPEIA. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. Communication: **Building Trust in Human Centric Artificial Intelligence**. Brussels. 8.4.2019, p. 9. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificialintelligence>. Acesso em: 8 mar. 2025.

155 UNESCO. Report Of The Social And Human Sciences Commission (SHS). Paris, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>. Acesso em: 8 mar. 2025.

uma “Avaliação de Impacto Ético” para prever os impactos dos sistemas de IA, além de analisar as estruturas legais existentes e considerar a criação de um “Responsável pela Ética da IA” independente.

A Recomendação também estabelece valores fundamentais de respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana, visando assegurar diversidade e inclusão, além de preservar a paz e a justiça em sociedades interconectadas. Esses valores são transformados em princípios e políticas aplicáveis para garantir sua efetiva operacionalização.

Em 13 de junho de 2024, o Parlamento Europeu e o Conselho da União Europeia aprovaram o Regulamento (UE) 2024/1689<sup>156</sup>, estabelecendo regras harmonizadas para a Inteligência Artificial (IA) na região. Conhecido como “Regulamento da Inteligência Artificial”, a lei visa promover uma IA segura, ética e centrada no ser humano, alinhando-se aos valores fundamentais da União Europeia.

O principal objetivo do regulamento é melhorar o funcionamento do mercado interno, criando um quadro jurídico uniforme para o desenvolvimento, comercialização e uso de sistemas de IA na União Europeia, incluindo garantir a proteção da saúde, segurança e direitos fundamentais dos cidadãos, bem como promovendo a inovação e a confiança nos sistemas de IA. O regulamento classifica os sistemas de IA em quatro categorias, com base no risco que representam:

1. Risco Inaceitável: Sistemas que representam uma ameaça clara e inaceitável aos direitos e segurança dos cidadãos, como sistemas de pontuação social. Esses sistemas são proibidos;
2. Alto Risco: Sistemas que podem afetar significativamente os direitos e segurança dos indivíduos, como IA em setores críticos (saúde, transporte, energia). Esses sistemas estão sujeitos a requisitos rigorosos de transparência, rastreabilidade e supervisão humana;
3. Risco Limitado: Sistemas que apresentam riscos moderados, como *chatbots*. Esses sistemas devem cumprir requisitos específicos de transparência, informando os usuários de que estão interagindo com uma IA;
4. Risco Mínimo: Sistemas que representam baixo ou nenhum risco, como filtros de *spam*. Esses sistemas estão sujeitos a poucas ou nenhuma exigência regulatória.

---

156UNIÃO EUROPEIA. **Regulamento 2024/1689** do Parlamento Europeu e do Conselho de 13 de junho de 2024 (Regulamento da Inteligência Artificial). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689>. Acesso em: 8 mar. 2025.

Para os sistemas de IA classificados como de alto risco, o regulamento impõe uma série de obrigações, incluindo:

- **Gestão de Dados e Governança:** Assegurar a qualidade e representatividade dos dados utilizados para treinar os sistemas de IA;
- **Documentação Técnica:** Manter registros detalhados sobre o desenvolvimento e funcionamento dos sistemas de IA;
- **Transparência e Explicabilidade:** Fornecer informações claras sobre o funcionamento dos sistemas e permitir que os usuários compreendam e contestem decisões automatizadas;
- **Supervisão Humana:** Garantir que os sistemas de IA possam ser supervisionados e controlados por humanos, especialmente em decisões críticas;
- **Avaliação de Conformidade:** Realizar avaliações periódicas para garantir que os sistemas de IA atendam aos requisitos estabelecidos pelo regulamento.

### **3.4.2 Princípios Fundamentais da Proteção no Cenário Jurídico Nacional**

O sistema jurídico brasileiro passou por uma abertura à institucionalização dos Direitos Humanos a partir da promulgação da Constituição de 1988. Nesse sentido, o Direito Constitucional e o Direito Internacional encontraram um ponto de interseção essencial, com o propósito de salvaguardar um valor fundamental: a dignidade da pessoa humana, a qual surge como elemento central do Estado Democrático de Direito no Brasil (art. 1º, III, CRFB/88), desempenhando papel vital no cerne do conceito de Direitos Humanos.

Considerando a natureza fundamental dos interesses envolvidos – relacionados aos direitos de intimidade, liberdade e igualdade – trazemos foco no direito à dignidade humana (art. 1º, III, CRFB/88) como o direito fundamental que pode estar em risco. Dele irradiam os direitos da personalidade e a garantia de inviolabilidade da intimidade, da vida privada, da honra e da imagem das pessoas (art. 5º, X, CRFB/88), à igualdade e à não discriminação (art. 5º, *caput*, CRFB/88), bem como à proteção de dados pessoais (art. 5º, LXXIX, CRFB/88).

A essência da busca pela dignidade da pessoa humana reside na garantia de que todos possam desfrutar de uma vida desprovida de arbitrariedade e violência, tendo como meta subjacente permitir que cada indivíduo se desenvolva plenamente e participe ativamente na vida política, social e cultural da comunidade em que se insere.

A proteção dos Direitos Humanos é uma construção sujeita a desenvolvimento contínuo conforme os limites da viabilidade política e da razoabilidade intelectual de cada tempo. A conquista do atual patamar de proteção dos Direitos Humanos exigiu uma trajetória de processos emancipatórios, muitas vezes revolucionários. Estes processos moldaram a percepção de várias sociedades, levando-as a reconhecer determinados elementos sociais como direitos inalienáveis, a serem conquistados e mantidos de forma duradoura. Ao lado dos direitos e garantias constitucionais, os direitos provenientes de tratados e convenções de Direitos Humanos devem ser invocados sempre que houver violação por parte do Estado, ou mesmo por particulares. O art. 5º, §2º, da Constituição Federal de 1988, diz que os direitos e garantias expressos ali não excluem outros decorrentes do regime e dos princípios por ela adotados, ou dos tratados internacionais em que o Brasil seja parte.

MAZZUOLI<sup>157</sup> analisa que, devido à disposição expressa do §2º, do artigo 5º, da Constituição Federal, estipulando que os direitos e garantias previstos no texto constitucional não excluem outros provenientes de tratados internacionais dos quais o Brasil seja signatário, os direitos derivados de acordos internacionais de direitos humanos detêm a qualidade de norma material constitucional. Isso significa que tais direitos não são formalmente inseridos na Constituição, mas têm caráter constitucional intrínseco. Por conseguinte, eles podem servir como padrão para a fiscalização da constitucionalidade por meio do método difuso.

Entretanto, em 2004, a Emenda Constitucional nº 45<sup>158</sup> foi aprovada, incluindo o §3º ao artigo 5º, estabelecendo a viabilidade de tratados e convenções internacionais sobre Direitos Humanos serem aprovados, em cada uma das Casas do Congresso Nacional, através de dois turnos de votação, obtendo três quintos dos votos dos membros respectivos. Nesse caso, esses acordos se equiparam a Emendas Constitucionais, o que equivale a dizer que são normas constitucionais tanto em sentido material quanto em sentido formal. Devido à inclusão desses dispositivos no texto constitucional, torna-se possível aplicar o método concentrado para o controle de constitucionalidade. Todavia, o STF julgou o HC 87.585<sup>159</sup> em dezembro de 2008. Antes desse momento, os tratados e convenções internacionais de

---

157MAZZUOLI, Valerio. **Controle jurisdicional da convencionalidade das leis**. 4 v, 2a. ed. rev., atual. e ampl. São Paulo: Revista dos Tribunais, 2011, pp. 146-147.

158BRASIL. Congresso Nacional. **Emenda Constitucional nº 45**, de 30 de dezembro de 2004. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/emendas/emc/emc45.htm](https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc45.htm). Acesso em: 8 mar. 2025.

159BRASIL. Supremo Tribunal Federal. **Habeas Corpus 87.585-8-TO**. Relator: Min. Marco Aurélio. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=597891>. Acesso em: 8 mar. 2025.

Direitos Humanos eram considerados pelo Supremo Tribunal Federal como tendo *status* de lei ordinária.

O resultado prático disso era que uma outra lei ordinária subsequente poderia revogar as disposições originadas desses instrumentos internacionais.

A partir da decisão do mencionado HC, o STF estabeleceu, majoritariamente, que os tratados e convenções internacionais de Direitos Humanos adentram ao sistema jurídico brasileiro como normas supralegais. Essa visão prevalecente foi sustentada pelo Ministro Gilmar Mendes.

Normas supralegais, nesse contexto, possuem um nível hierárquico superior ao das leis ordinárias, porém inferior ao da Constituição. Isso significa que uma lei ordinária não pode revogar disposições provenientes de instrumentos internacionais de Direitos Humanos.

Há uma exceção: os tratados e convenções internacionais de Direitos Humanos que são incorporados conforme o procedimento estipulado no §3º, do artigo 5º, isto é, com uma maioria qualificada e uma votação em dois turnos em cada Casa Legislativa. Nesses casos, os acordos internacionais de Direitos Humanos não têm *status* de normas supralegais, mas sim de Emendas Constitucionais. Em relação aos tratados e convenções internacionais que não abordam questões de Direitos Humanos, o entendimento do STF é que possuem a natureza de lei ordinária.

A Convenção Americana sobre Direitos Humanos (Pacto de São José da Costa Rica), ratificado pelo Brasil através da promulgação do Decreto nº 678/1992<sup>160</sup>, tem como propósito o fortalecimento e a promoção de direitos no continente americano e é obrigatório para aqueles Estados que o ratifiquem ou que se aderem a ele, contemplando em seu bojo um rol de direitos considerados não apenas essenciais, mas fundamentais, que constituem princípios subjacentes à dignidade humana.

O Protocolo Adicional à Convenção Americana Sobre Direitos Humanos em Matéria de Direitos Econômicos, Sociais e Culturais, (Protocolo de San Salvador), foi ratificado pelo Brasil através da promulgação do Decreto nº 3.321/1999<sup>161</sup>, incorporando, assim, um rol de direitos sociais adicionais ao ordenamento pátrio, tais como: direito ao trabalho (Art. 6), condições justas, equitativas e satisfatórias de trabalho (Art. 7), entre outros.

---

160BRASIL. Congresso Nacional. **Decreto nº 678**, de 6 de novembro de 1992. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/d0678.htm](http://www.planalto.gov.br/ccivil_03/decreto/d0678.htm). Acesso em: 8 mar. 2025.

161BRASIL. Congresso Nacional. **Decreto nº 3.321**, de 30 de dezembro de 1999. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decreto/d3321.htm](https://www.planalto.gov.br/ccivil_03/decreto/d3321.htm). Acesso em: 8 mar. 2025.

No que se refere ao princípio de igualdade perante a lei e prevenção de discriminação, a Corte Interamericana de Direitos Humanos destacou que a noção de igualdade deriva da unidade fundamental da humanidade e está inseparavelmente ligada à dignidade essencial humana. Na fase atual do desenvolvimento do Direito Internacional, o princípio fundamental de igualdade e não discriminação conquistou *status* no âmbito do *jus cogens*. Esse princípio serve como base legal tanto para a ordem interna quanto internacional e permeia todo o sistema jurídico. Os Estados têm a responsabilidade de evitar qualquer ação que possa, direta ou indiretamente, levar à criação de situações de discriminação *de jure* ou *de facto*.

A Corte já ressaltou que, enquanto a obrigação geral definida no artigo 1.1 da Convenção Americana se refere ao dever do Estado de respeitar e garantir os direitos contidos neste tratado “sem discriminação”, o artigo 24 protege o direito à “igual proteção perante a lei”.

De acordo com a jurisprudência da Corte, o artigo 24 da Convenção também incorpora uma diretriz que visa garantir a igualdade substancial. Portanto, o direito à igualdade estipulado nesta disposição possui uma dimensão formal, protegendo a igualdade diante da lei, bem como uma dimensão material ou substantiva, que exige “a implementação de medidas afirmativas em prol de grupos historicamente marginalizados ou discriminados, com base nos fatores mencionados no artigo 1.1 da Convenção Americana”<sup>162</sup>.

A Convenção Sobre a Eliminação de Todas as Formas de Discriminação Contra a Mulher (CEDAW), ratificada no Brasil pela promulgação do Decreto nº 4.377/2002, em seu art. 5º, alínea “a”, dispõe expressamente:

Os Estados-Partes tomarão todas as medidas apropriadas para:  
a) Modificar os padrões socioculturais de conduta de homens e mulheres, com vistas a alcançar a eliminação dos preconceitos e práticas consuetudinárias e de qualquer outra índole que estejam baseados na ideia da inferioridade ou superioridade de qualquer dos sexos ou em funções estereotipadas de homens e mulheres.

Dessa forma, os princípios introduzidos pelo Direito Internacional por meio da Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW), visam erradicar qualquer forma de distinção, restrição ou exclusão fundamentada no sexo, cujo efeito ou intenção comprometa ou anule o reconhecimento, o desfrute ou o

---

162CORTE INTERAMERICANA DE DIREITOS HUMANOS: **Caso Márcia Barbosa e Outros Vs. Brasil, Exceções Preliminares, Mérito, Reparações e Custas**, Sentença de 7 de setembro de 2021, par. 138-141. Disponível em: <http://cepia.org.br/wp-content/uploads/2021/12/Sentenc%CC%A7a-marcia-barbosa.pdf>. Acesso em: 8 mar. 2025.

exercício dos direitos das mulheres. Nesse ponto específico, o Comitê CEDAW destacou que a disseminação de estereótipos de gênero dentro do sistema judicial tem um impacto grave no pleno desfrute dos direitos humanos das mulheres. Esses estereótipos podem criar obstáculos ao acesso à justiça em todas as áreas e afetar especialmente as mulheres que são vítimas ou sobreviventes de violência<sup>163</sup>.

No que diz respeito aos tratados de direitos humanos, é fundamental ressaltar a importância de uma interpretação que sempre considere a norma mais benéfica ao ser humano. Isso significa que esses tratados devem ser analisados com base no princípio *pro homine*<sup>164</sup> (princípio da primazia das normas mais favoráveis), por meio do qual o intérprete (e o aplicador do direito) deve escolher a norma que melhor proteja o ser humano como sujeito de direitos em cada situação específica.

CANOTILHO<sup>165</sup> examina a estrutura constitucional, diferenciando princípios e regras e destacando o papel fundamental dos princípios estruturantes na organização e interpretação do sistema constitucional. De acordo com o jurista, os princípios fundamentais que moldam a constituição adquirem significado e aplicação através da interação dinâmica com outros princípios, que por sua vez se desdobram em subprincípios e regras para definir seu alcance e implementação no contexto jurídico e político. Nesse passo, os pilares do Estado de Direito se manifestam em princípios secundários como a constitucionalidade, a legalidade e a sujeição do legislador aos direitos fundamentais, que por sua vez são detalhados por outros preceitos constitucionais especializados. Por conseguinte, se os tratados e convenções internacionais de direitos humanos podem ser considerados equivalentes às emendas constitucionais (nos termos do art. 5º, §3º, da CRFB/88) ou, ainda, possuírem *status* supralegal (conforme o entendimento atual do STF, expresso no RE 349.703/RS<sup>166</sup>), especialmente se ratificados antes da Emenda Constitucional 45/2004, é certo que, desfrutando

---

163ORGANIZAÇÃO DAS NAÇÕES UNIDAS, Comitê CEDAW: **Recomendação Geral nº 33 sobre o acesso das mulheres à justiça**, 3 de agosto de 2015, CEDAW/C/GC/33, par. 26, p. 14. Disponível em: <https://assets-compromissoeatitude-ipg.sfo2.digitaloceanspaces.com/2016/02/Recomendacao-Geral-n33-Comite-CEDAW.pdf>. Acesso em: 8 mar. 2025.

164BRASIL. Supremo Tribunal Federal. **ADI 4.439/DF**, Relator: Ministro Roberto Barroso. Data de julg.: 27/09/2017. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=15085915>. Acesso em: 8 mar. 2025.

165CANOTILHO, J. J. Gomes. **Direito Constitucional e Teoria da Constituição**. 7a ed., 11ª reimpressão. Coimbra: Almedina, 2003, pp. 1165-1175.

166BRASIL. Supremo Tribunal Federal. **Recurso Extraordinário nº 349.703/RS**. Plenário, julgamento em 3 de dezembro de 2008; relator: Ministro Carlos Ayres Britto. DJe nº 236, 12 de dezembro de 2008. Disponível em: [https://jurisprudencia.s3.amazonaws.com/STF/IT/RE\\_349703\\_RS\\_1278971642175.pdf?AWSAccessKeyId=AKIARMMD5JEA067SMCVA&Expires=1753805389&Signature=hvpGhky2Qkv4IERixp0NTGK0Lwo%3D](https://jurisprudencia.s3.amazonaws.com/STF/IT/RE_349703_RS_1278971642175.pdf?AWSAccessKeyId=AKIARMMD5JEA067SMCVA&Expires=1753805389&Signature=hvpGhky2Qkv4IERixp0NTGK0Lwo%3D). Acesso em: 8 mar. 2025.

da condição de normas materialmente constitucionais, o Decreto nº 678/1992 e o Decreto nº 4.377/2002 devem também ser utilizados como parâmetro para o controle da produção e interpretação normativa no âmbito interno.

Como signatário da Declaração Universal sobre Bioética de Direitos Humanos<sup>167</sup> – homologada unanimemente em 19 de outubro de 2005, em Paris, por aclamação dos 191 países-membros, na 33ª Sessão da Conferência Geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) –, o Brasil assumiu o compromisso internacional de recusar “atos contrários aos Direitos Humanos, às liberdades fundamentais e à dignidade humana”, bem como:

[...] tomar todas as medidas adequadas de caráter legislativo, administrativo ou de qualquer outra natureza, de modo a implementar os princípios estabelecidos na presente Declaração e em conformidade com o direito internacional e com os direitos humanos. Tais medidas devem ser apoiadas por ações nas esferas da educação, formação e informação ao público. (Artigo 22, letra a).

### 3.4.3 Princípios e Limitações na Regulação da IA no Brasil

Em Sessão Deliberativa Ordinária realizada em 10 de dezembro de 2024<sup>168</sup>, o PL nº 2.338, de 2023, foi aprovado pelo Senado Federal e encaminhado à Câmara dos Deputados para apreciação. Os Projetos de Lei nºs 5.051 e 5.691, de 2019; 21, de 2020; 872, de 2021; 3.592, de 2023; além dos Projetos nºs 210 e 266, de 2024, foram considerados prejudicados e arquivados.

O Capítulo I – Das Disposições Preliminares, traz como principal objetivo da proposta estabelecer normas gerais para uma governança responsável dos sistemas de Inteligência Artificial, equilibrando o avanço tecnológico com a proteção dos direitos fundamentais.

Todavia, o texto aprovado pelo Senado apresenta falhas no que tange à sua conformidade com o princípio constitucional da reserva legal.

Consagrado no artigo 5º, inciso II, da Constituição Federal, este princípio determina que “ninguém será obrigado a fazer ou deixar de fazer alguma coisa senão em virtude de lei”. Esse preceito garante que apenas a lei, em sentido formal, pode criar obrigações ou restringir direitos, assegurando segurança jurídica e previsibilidade nas relações sociais.

167UNESCO. **Declaração Universal sobre Bioética de Direitos Humanos**. Paris, 2006. Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000146180\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000146180_por). Acesso em: 8 mar. 2025.

168BRASIL. Senado Federal. **PL 2338/2023**. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9881643&ts=1734719326347&disposition=inline>. Acesso em: 8 mar. 2025.

A minuta encaminhada à Câmara dos Deputados propõe-se como um marco regulatório pioneiro, contudo, apresenta lacunas que comprometem a concretização desses objetivos, especialmente quando analisado sob a perspectiva hermenêutica da Constituição como um instrumento normativo de transformação social.

Fugiria ao escopo desta pesquisa analisar minuciosamente o texto aprovado pelo Senado, mas algumas considerações se fazem necessárias.

Em uma análise preliminar dos artigos 1º a 30 da minuta de lei, observa-se que grande parte dos dispositivos apresentados se baseia em princípios amplos e indeterminados, que requerem concretização normativa e regulamentar.

Nos artigos 1º a 14, os princípios estabelecidos, como a centralidade da pessoa humana (art. 2º, I), transparência (art. 3º, VI) e não discriminação ilícita ou abusiva (art. 3º, IV), refletem direitos fundamentais de caráter genérico e demandam maior especificação normativa.

A centralidade da pessoa humana exige uma interpretação contextualizada, especialmente em situações como a proteção de grupos vulneráveis (art. 2º, XIV) ou a governança de sistemas de IA de alto risco (art. 14). A ausência de critérios técnicos claros nesses dispositivos implica a necessidade de regulamentação complementar para operacionalizar os princípios estabelecidos.

A mesma lógica se aplica ao artigo 4º, que apresenta definições como “discriminação abusiva” e “ciclo de vida da IA”. Embora esses conceitos sejam essenciais para delimitar o alcance da lei, eles carecem de critérios objetivos e padrões técnicos que viabilizem sua aplicação prática. No tocante à efetividade dos direitos fundamentais, a simples previsão de princípios sem um detalhamento técnico adequado revela-se insuficiente para assegurar a realização plena dos direitos. Torna-se necessário que a regulamentação estabeleça parâmetros claros e objetivos, permitindo a implementação concreta dos direitos previstos, tanto em termos de proteção individual quanto coletiva, evitando que a indeterminação normativa comprometa a eficácia da norma. Nos dispositivos que abordam os direitos das pessoas afetadas por sistemas de IA (artigos 5º e 6º), a imprecisão normativa dos direitos fundamentais torna-se ainda mais evidente.

O direito à explicação (art. 6º, I) é consagrado como um princípio, mas depende de regulamentação técnica que defina os parâmetros para a explicação adequada, especialmente em sistemas de alto risco.

O mesmo ocorre com a revisão humana prevista no art. 6º, III, que não esclarece os critérios para essa supervisão, deixando lacunas que podem comprometer sua efetividade. Já nos artigos 15 a 30, a necessidade de concretização normativa também é evidente.

A classificação de sistemas de IA de alto risco (art. 15) apresenta critérios amplos como “alto potencial danoso” e “grau de reversibilidade dos danos”, que carecem de especificação técnica e metodológica. A ausência de parâmetros objetivos para mensuração dificulta a aplicação uniforme do dispositivo e pode levar a interpretações divergentes.

A verdadeira realização dos direitos fundamentais na minuta de lei dependerá da efetividade de uma regulamentação complementar, que deve ser realizada de forma cuidadosa e atenta às necessidades da sociedade e às peculiaridades do campo tecnológico. A regulamentação não pode ser vista como uma simples extensão da lei, mas como uma tarefa que visa garantir a concretização dos direitos e a promoção de uma sociedade mais justa e equitativa no contexto do uso de tecnologias avançadas como a IA. O sucesso dessa tarefa exigirá uma ação coordenada entre o legislador, as autoridades reguladoras, o Judiciário e a sociedade civil, que, trabalhando juntos, poderão garantir a eficácia da lei, a proteção dos direitos fundamentais e o desenvolvimento responsável da Inteligência Artificial.

Embora o texto destaque a proteção de grupos vulneráveis, como crianças, idosos e pessoas com deficiência, faltam mecanismos concretos e imediatamente aplicáveis que garantam a efetividade dessas disposições. A normatividade dos direitos fundamentais exige mais do que previsões genéricas; são necessários instrumentos jurídicos vinculantes que assegurem a aplicação prática dessas proteções, evitando que se limitem a promessas abstratas.

O projeto carece de mecanismos redistributivos ou medidas concretas para reduzir desigualdades estruturais, frustrando a concepção constitucional de justiça distributiva e inclusão. Metas como equidade e justiça social permanecem declarativas, sem força normativa para impactar a realidade. Para superar essas fragilidades, algumas alterações são indispensáveis. Em primeiro lugar, é essencial estabelecer critérios objetivos para os conceitos de “alto risco” e “risco sistêmico”, atualmente vagos, o que compromete a segurança jurídica e a previsibilidade. Critérios claros são fundamentais para que a norma atenda ao princípio da legalidade e proporcione segurança aos envolvidos.

O §2º prevê a criação de regimes regulatórios simplificados, os quais serão definidos pelo Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA).

De acordo com o art. 4º, IX, o SIA se configura como um mecanismo regulatório coordenado pela autoridade competente, com a função de promover a cooperação entre autoridades setoriais e entidades reguladoras. A proposta remete a uma lógica de governança distribuída e colaborativa, embora sua eficácia dependa da autonomia e da clareza das funções atribuídas a cada órgão. Este ponto é especialmente relevante em um contexto de rápida evolução tecnológica, já que a criação do SIA parece depender de uma infraestrutura ainda não consolidada, podendo retardar a eficácia da lei.

O art. 25 determina que desenvolvedores ou aplicadores de sistemas de IA de alto risco realizem uma avaliação de impacto algorítmico antes de sua introdução no mercado. Essa exigência representa um modelo de precaução alinhado ao direito internacional, notadamente ao Regulamento da Inteligência Artificial da UE, visando à proteção de direitos fundamentais. Porém, a definição ampla de “alto risco” e a falta de clareza sobre “impactos adversos” demandam regulamentação mais precisa para evitar lacunas no cumprimento dessa obrigação.

A metodologia da avaliação de impacto algorítmico deve considerar rigorosamente os “riscos e benefícios aos direitos fundamentais”, mesmo diante da complexidade e rápida evolução dos algoritmos. Além disso, a flexibilização da avaliação, conforme prevista pela autoridade setorial, precisa ser criteriosa para evitar que sistemas de alto risco sejam introduzidos sem uma análise rigorosa.

O projeto de lei precisa superar fragilidades e estruturar-se de maneira mais concreta e vinculante, de modo a efetivar os valores constitucionais em um campo tão sensível quanto a Inteligência Artificial. Apesar de avanços pontuais, o texto atual corre o risco de não atender às demandas por transformação social e proteção de direitos fundamentais que norteiam a Constituição Brasileira.

Destarte as considerações anteriores, gostaríamos de dar especial atenção às disposições contidas no art. 3º, IV, que aborda a “discriminação ilícita ou abusiva”, cujos conceitos estão definidos no art. 4º, XII.

Quando uma nova legislação, como a referente à IA, não fornece todos os detalhes necessários para a sua aplicação prática, o Judiciário pode, nesse caso, recorrer às leis antidiscriminatórias como parâmetros para preencher as lacunas da norma e assegurar que os direitos fundamentais, especialmente aqueles relacionados à igualdade e não discriminação, sejam efetivamente protegidos.

As normas que tratam da discriminação racial, da discriminação contra a mulher, ou da proteção dos direitos das crianças e dos idosos, já contêm definições claras de discriminação e mecanismos de reparação. Elas estabelecem critérios objetivos para a aplicação desses direitos, como a necessidade de igualdade de tratamento e o combate a práticas discriminatórias em diferentes contextos. Entretanto, o texto apresenta um novo conceito: o da discriminação abusiva.

A expressão “discriminação abusiva”, conforme utilizada na minuta da lei sobre Inteligência Artificial, parece referir-se a uma forma de discriminação que, além de ser indevida ou injusta, ultrapassa os limites do aceitável em determinado contexto.

Na seara jurídica, discriminação designa o tratamento desigual de indivíduos ou grupos com base em características como raça, sexo, orientação sexual, origem étnica, religião, idade, entre outras, sem justificativa objetiva e razoável para tal diferenciação. O tratamento desigual pode manifestar-se de forma direta, quando ocorre uma distinção explícita, ou indireta, quando um critério aparentemente neutro impacta desproporcionalmente certos grupos.

A discriminação negativa, frequentemente associada ao termo “discriminação”, ocorre quando indivíduos ou grupos são tratados de maneira desigual ou prejudicial. Esse tipo de discriminação é injustificável e ilegal em muitos sistemas jurídicos, incluindo o brasileiro. Exemplos de discriminação negativa incluem racismo, sexismo, homofobia ou qualquer outro tratamento desigual não justificado por razões objetivas, que cause danos ou exclusão ao grupo discriminado. Por outro lado, a discriminação positiva (ou ação afirmativa) visa promover a igualdade substancial, reconhecendo que, em algumas situações, o tratamento igualitário pode perpetuar desigualdades históricas. Por conseguinte, a discriminação positiva busca corrigir desigualdades, oferecendo um tratamento desigual com o intuito de alcançar a equidade e proporcionar oportunidades iguais para grupos historicamente marginalizados ou desfavorecidos.

Um exemplo clássico de discriminação positiva é a adoção de cotas raciais ou cotas para pessoas com deficiência em universidades ou concursos públicos. Estas políticas têm por objetivo garantir que indivíduos de grupos minoritários ou vulneráveis, que sofreram séculos de exclusão social, tenham acesso a oportunidades que, de outra forma, seriam inacessíveis devido a condições desiguais de partida.

A expressão “discriminação abusiva”, conforme apresentada na minuta da lei, sem especificar se se refere à discriminação negativa ou positiva, sugere um conceito que abrange tanto a discriminação negativa excessiva quanto a limitação da discriminação positiva.

A discriminação abusiva, assim, parece apontar para um limite da discriminação negativa, ou seja, o tratamento desigual deve ser necessário, proporcional e justificado, tanto em políticas de ação afirmativa (discriminação positiva) quanto em casos de discriminação negativa. O abuso se daria quando, embora a discriminação inicialmente pudesse ser justificada (para corrigir desigualdades), ultrapassasse os limites da razoabilidade, tornando-se excessiva, desnecessária ou desproporcional. Isso significaria que qualquer política de proteção a minorias seria suscetível de ser rotulada como abusiva em nome da igualdade formal. Porém, a justiça social não se resume à simples ideia de tratar todos de maneira idêntica. O tratamento desigual pode ser necessário justamente para corrigir desigualdades históricas.

Considerar a proteção de minorias historicamente subalternizadas ou marginalizadas como “desproporcional” ou “desnecessária”, permitiria, por exemplo, acusações de “racismo reverso” ou a ideia de homens sendo vítimas de misandria. Essas expressões são frequentemente usadas de forma distorcida para sugerir que as ações afirmativas ou políticas de igualdade seriam uma forma de discriminação contra grupos historicamente dominantes, como pessoas brancas ou homens. No entanto, racismo reverso e misandria são conceitos contestáveis, pois se baseiam em uma interpretação equivocada da natureza das desigualdades estruturais.

O conceito de discriminação positiva – como as cotas raciais ou a afirmação de políticas para mulheres e grupos marginalizados – visa corrigir desigualdades históricas que resultaram em exclusão sistêmica e injustiça estrutural. Estas políticas são implementadas com o objetivo de promover uma igualdade real, não de criar uma inversão das opressões. O objetivo é criar um campo de oportunidades mais igualitário.

Observa-se que as Constituições dos Estados Democráticos contemporâneos incluíram um leque de direitos fundamentais, abrangendo catálogos de direitos individuais, coletivos, sociais, econômicos e culturais, além de outros direitos expressos ou implícitos.

Esse conjunto, conhecido como “bloco de constitucionalidade”, garante, por um lado, a imutabilidade ou abolição desses direitos por meio de reforma constitucional e, por outro, sua abertura mediante cláusula específica de recepção dos direitos humanos presentes em

instrumentos e também internacionais.

No caso da Constituição Brasileira, esse mecanismo é observado no artigo 5º, §2º. Contudo, frente a possibilidade de expansão dos direitos fundamentais positivados na ordem constitucional vigente por meio da cláusula expansiva, é imperativo ponderar cuidadosamente o sentido e alcance que a Constituição atribui a ela, evitando a proliferação de direitos que não preservem a fundamentalidade característica da dignidade da pessoa humana e prevenindo, assim, a banalização desses direitos.

Se mal interpretada, a expressão “discriminação abusiva” pode abrir espaço para distorções ideológicas que buscam argumentar que políticas de ação afirmativa resultam em injustiças contra grupos majoritários e socialmente privilegiados. Isso, de fato, criaria um falso equilíbrio que ignoraria o contexto das desigualdades estruturais.

#### 3.4.4 Governança Pública, Ética e Inteligência Artificial

Diante das ações de decisão e predição realizadas, ao considerar uma possível aplicação da Inteligência Artificial na Administração Pública é útil dividir a análise em duas dimensões: interna e externa. A dimensão interna abrange as atividades de gestão organizacional e processual dentro da entidade pública, oferecendo facilidades como comunicação e compartilhamento de informações entre agentes e órgãos públicos. Por outro lado, a dimensão externa diz respeito ao uso da tecnologia para interação com os cidadãos, visando facilitar atividades e solucionar problemas de forma online. Embora essa contribuição possa resolver problemas administrativos e reduzir custos, a incorporação da IA na Administração Pública levanta preocupações sobre possíveis retrocessos.

Reconhecida na ADPF 347/DF<sup>169</sup>, a teoria do estado de coisas inconstitucional, no dizer de CAMPOS<sup>170</sup>, significa a ideia de a “omissão inconstitucional decorrer da falha de coordenação entre o Legislativo e o Executivo, a implicar deficiências na consecução de políticas públicas.” Nesse sentido, a ausência de coesão no plano constitucional precipita sua disfuncionalidade, evidenciada pela falha estrutural, a qual será detalhada adiante, conforme o autor:

169BRASIL. Supremo Tribunal Federal. **Arguição de Descumprimento de Preceito Fundamental (ADPF 347)**. Tribunal Pleno. Relator Ministro Marco Aurélio. Data do Julg.: 04/10/2023. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=4783560>. Acesso em: 8 mar. 2025.

170CAMPOS, Alexandre de Azevedo. “O Estado de Coisas Inconstitucional e o litígio estrutural”. **CONJUR**. Disponível em: <http://www.conjur.com.br/2015-set-01/carlos-campos-estado-coisas-inconstitucional-litigio-estrutural>. Acesso em: 8 mar. 2025.

A falha estrutural configura-se, normalmente, como ausência ou deficiência de políticas públicas. Graves e sistemáticas violações de direitos são originadas e agravadas por falhas estruturais nos procedimentos de desenho, implementação, avaliação e financiamento de políticas públicas. Essas falhas têm raízes em prolongadas omissões dos agentes e autoridades públicos, em limitações das políticas públicas correspondentes, na falta de medidas administrativas, legislativas ou orçamentárias voltadas na superação dos problemas de direitos.<sup>171</sup>

A doutrina do Estado de Coisas Inconstitucional, erigida pela jurisprudência da Corte Constitucional da Colômbia, surge como resposta aos casos persistentes de fracasso das políticas públicas e às suas conseqüentes transgressões aos direitos fundamentais. Nessa esteira, destaca-se a discrepância entre as estruturas de poder delineadas na parte orgânica da Constituição, caracterizada pela recorrente omissão nas políticas públicas ao longo dos anos, e os preceitos consagrados na parte dogmática do mesmo texto constitucional, os quais visam fomentar a inclusão social, a igualdade material e a proteção da dignidade humana.

De tal modo, evidencia-se um claro entrave no processo político e institucional, impedindo a resolução dos litígios estruturais que afetam um grande número de pessoas vítimas de violações aos direitos humanos, devido à falta de vontade política e à ineficácia da estrutura administrativa.

Segundo CAMPOS<sup>172</sup>, o conceito de estado de coisas inconstitucional poderia ser discernido pelo Tribunal Constitucional como um embate estrutural quando se vislumbram os seguintes requisitos: a) a violação em larga escala e difundida dos direitos fundamentais de um expressivo contingente de indivíduos; b) a prolongada negligência das autoridades em cumprir suas responsabilidades na salvaguarda e promoção desses direitos; c) a resolução das violações aos direitos demanda a implementação de medidas complexas por uma multiplicidade de órgãos, implicando alterações estruturais que podem requerer a alocação de recursos públicos, revisão das políticas públicas vigentes ou a formulação de novas políticas, entre outras ações; e d) a possibilidade de sobrecarga do sistema judiciário, caso todos aqueles que tiverem seus direitos violados buscarem individualmente reparação junto ao Poder Judiciário.

Em síntese, são três os pressupostos do Estado de Coisas Inconstitucional:  
a) a constatação de um quadro não simplesmente de proteção deficiente, e sim de violação massiva, generalizada e sistemática de direitos fundamentais, que afeta a um número amplo de pessoas;

---

171CAMPOS, *Op. cit.*

172*Idem, ibidem.*

- b) a falta de coordenação entre medidas legislativas, administrativas, orçamentárias e até judiciais, verdadeira “falha estatal estrutural”, que gera tanto a violação sistemática dos direitos, quanto a perpetuação e agravamento da situação;
- c) a superação dessas violações de direitos exige a expedição de remédios e ordens dirigidas não apenas a um órgão, e sim a uma pluralidade destes – são necessárias mudanças estruturais, novas políticas públicas ou o ajuste das existentes, alocação de recursos etc.

Assim, é factível identificar a viabilidade de aplicar o mencionado conceito de estado de coisas inconstitucional ao modelo jurídico brasileiro, especialmente no que diz respeito à automatização do processo decisório estatal por Inteligência Artificial (IA) sem a devida e rigorosamente necessária supervisão humana, haja vista que, ao introduzir um sistema de IA na Administração Pública, é indispensável que haja uma regulamentação específica que o respalde. De acordo com os princípios constitucionais que orientam as atividades administrativas, é necessário que exista uma norma explícita que autorize essa implementação de maneira apropriada, do contrário, tal ação seria considerada ilegal. A utilização da IA poderia ser vista como irregular no momento em que começasse a determinar quais medidas são mais relevantes para a Administração Pública, ou quais processos devem ser considerados pelo funcionário público, assumindo, assim, funções tanto de gestor quanto de magistrado.

Segundo VIEIRA & BARRETO<sup>173</sup>, a governança, a gestão de riscos e a integridade (GRC) constituem um conjunto de práticas interligadas com o objetivo de agregar valor às instituições públicas e privadas. Essas iniciativas operam de maneira integrada para assegurar o cumprimento das metas, lidar adequadamente com as incertezas e promover a conduta ética. Esse modelo organiza e desenvolve as capacidades em todos os níveis da organização, buscando um desempenho sustentável, pautado pela integridade, levando em conta as incertezas, os processos internos e o atendimento às exigências externas.

Nas rotinas laborais e na gestão pública, o avanço tecnológico transformou radicalmente a maneira como interagimos com o trabalho e os serviços do Estado. Tais mudanças, embora promovam ganhos de eficiência, também levantam questões sobre privacidade, autonomia e a dignidade humana. No ambiente de trabalho, ferramentas tecnológicas de monitoramento passaram a capturar, registrar e analisar cada detalhe da atividade dos empregados.

---

<sup>173</sup>VIEIRA, James Batista; BARRETO, Rodrigo T. de Souza. **Governança, gestão de riscos e integridade**. Brasília: ENAP, 2019, p. 11.

Desde os resultados de testes de qualidade vinculados a técnicos específicos até o uso de dispositivos que registram todas as teclas digitadas, o que antes era mediado pela supervisão direta tornou-se um processo automatizado e contínuo. A vigilância permanente coloca o trabalhador sob um olhar que não se desvia, criando um contexto de controle absoluto, onde a privacidade e a liberdade individual são praticamente anuladas.

Na esfera da Administração Pública, a automação, impulsionada por sistemas de Inteligência Artificial, trouxe novas oportunidades, mas também desafios inéditos. Se, por um lado, a introdução desses sistemas nos serviços públicos têm o potencial de acelerar processos e reduzir falhas humanas, por outro, a ausência de transparência em decisões automatizadas e o risco de desconexão com as necessidades reais da população geram inquietações legítimas. Mais grave ainda é a possibilidade de violação de direitos fundamentais, como o direito à privacidade, o direito à escolha e ao consentimento informado. Quando algoritmos substituem deliberações humanas, a relação entre o cidadão e o Estado pode se tornar desumanizada e distante.

Ao se delegar à máquina o que é essencialmente humano – a reflexão crítica e o julgamento moral – cria-se um cenário onde a responsabilidade se dilui, dispersando-se entre operadores, programadores e o próprio algoritmo. Essa responsabilidade difusa denuncia um vácuo do sujeito decisor, desmaterializando o ator jurídico e fragilizando o suporte da responsabilização. Tais situações, apesar de ocorrerem em contextos distintos, convergem em um ponto essencial: ambas ilustram como a tecnologia pode transformar relações humanas fundamentais e, ao mesmo tempo, fragilizar direitos constitucionais se não for cuidadosamente regulamentada. No trabalho, o empregado perde espaço para exercer sua autonomia, enquanto na Administração Pública o cidadão pode se ver excluído de processos decisórios importantes.

No artigo intitulado “*The Paradox of Regulatory Discretion*”<sup>174</sup>, a autora investiga a relação entre a discricionariedade regulatória e a eficácia das políticas públicas, focando na aplicação de sistemas de Inteligência Artificial na Administração. A pesquisa compara três autoridades regulatórias britânicas nos setores de comunicações, gás e eletricidade, e serviços de água, onde avalia como a discricionariedade afeta a implementação de políticas e o alcance de objetivos regulatórios. Os resultados mostram que, embora a discricionariedade permita ajustes mais flexíveis às condições específicas, ela pode gerar inconsistências e ineficiências,

---

174YEUNG, K. (2018), “Algorithmic regulation: A critical interrogation”. *Regulation & Governance*, 12: 505-523. Disponível em: <https://doi.org/10.1111/rego.12158>. Acesso em: 8 mar. 2025.

em especial quando carece de mecanismos adequados de prestação de contas e transparência.

A Constituição de um Estado é a pedra angular que estrutura sua organização política e jurídica, ao mesmo tempo em que consagra os direitos fundamentais como elementos centrais de um ordenamento voltado ao bem comum. Esses direitos, universais ou incorporados pelo Direito Positivo, constituem a espinha dorsal dos princípios da liberdade e da igualdade, fundamentos indispensáveis para a promoção da dignidade humana e a realização da justiça social. Encarregado de preservar os valores éticos e jurídicos, o Estado deve implementar sistemas de controle que promovam a integridade das ações públicas, respeitando os fundamentos da Constituição – pautado pela ética em suas ações –, por meio de medidas de autocontrole ou controle externo.

A conexão entre esses valores constitucionais e o Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal<sup>175</sup> demonstra a relevância da ética na Administração Pública e constitui um importante desdobramento dos princípios constitucionais, abordando questões éticas e morais, ressaltando a importância da ética na Administração Pública – onde o equilíbrio entre legalidade e propósito é essencial para fortalecer a moralidade dos atos administrativos –, e garantindo a conformidade com os valores éticos pelos agentes públicos.

A Administração Pública, compreendida como o conjunto de entidades governamentais encarregadas das atividades administrativas, está sujeita a princípios fundamentais. Seguindo rigorosamente o que estabelece a CRFB/88, a Administração Pública é regida por princípios constitucionais, conforme estabelecido no artigo 37, os quais incluem os princípios da legalidade, impessoalidade, moralidade, publicidade e eficiência.

A eficiência na Administração Pública está associada à qualidade dos serviços prestados, acompanhada do desempenho da gestão governamental, como previsto na EC nº 19/1998<sup>176</sup>. Isso representa um dos modos pelos quais a sociedade pode exigir da Administração Pública a oferta de serviços de qualidade.

Uma vez que a gestão pública está sujeita a um regime jurídico que a restringe às ações permitidas por lei, surge o princípio da legalidade, respaldado no art. 37, *caput*, da CRFB/88. Ou seja, as atividades do setor público devem ser realizadas de acordo com normas legais previamente estabelecidas.

---

175BRASIL. Congresso Nacional. **Decreto nº 1.171**, de 22 de junho de 1994. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decreto/d1171.htm](https://www.planalto.gov.br/ccivil_03/decreto/d1171.htm). Acesso em: 8 mar. 2025.

176BRASIL. Congresso Nacional. **Emenda Constitucional nº 19**, de 14 de junho de 1998. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/Emendas/Emc/emc19.htm](https://www.planalto.gov.br/ccivil_03/constituicao/Emendas/Emc/emc19.htm). Acesso em: 8 mar. 2025.

Portanto, qualquer violação desse princípio constitui um ato ilegal, uma vez que não é aceitável uma ação que não tenha sido explicitamente autorizada pela lei.

Além dos princípios constantes da CRFB/88, a Lei nº 9.784/1999<sup>177</sup>, o art. 2º, dita que a Administração Pública obedecerá, dentre outros, aos princípios da legalidade, finalidade, motivação, razoabilidade, proporcionalidade, moralidade, ampla defesa, contraditório, segurança jurídica, interesse público e eficiência.

Inerente ao princípio da legalidade, o princípio da finalidade se refere à compreensão de que todas as ações da Administração Pública devem estar alinhadas com os objetivos previstos pela lei. O propósito da finalidade é garantir que os agentes públicos ajam de acordo com os objetivos legais estabelecidos, utilizando condutas que sejam adequadas, certas e lícitas para alcançar tais objetivos em todos os atos administrativos. Isso quer dizer que as atividades administrativas devem ser direcionadas para a realização dos fins estabelecidos pela legislação, evitando desvios de conduta ou o uso indevido do poder público. Assim, o princípio da finalidade atua como um mecanismo que orienta a atuação da Administração Pública, garantindo que suas ações estejam em consonância com os interesses públicos e os propósitos estabelecidos pela lei.

Portanto, a moralidade na Administração Pública não reside em uma mera formalidade, mas se constitui como alicerce essencial na construção de uma sociedade justa e equânime. Ao equipararmos a supervisão ética ao controle constitucional, defendemos que a ética não pode ser relegada a uma questão periférica, mas deve ser central na análise da legitimidade e legalidade dos atos administrativos. Dessa forma, a incorporação dos princípios éticos na rotina da gestão estatal figura como imprescindível não apenas para assegurar a conformidade jurídica, mas também para fomentar a confiança da comunidade nas estruturas governamentais.

Embora a Inteligência Artificial possa ser útil nos processos decisórios, ela não é onisciente. Não tem a capacidade de conhecer, aprender e formular respostas de forma intuitiva como os seres humanos, mas ajuda no processamento das informações que recebe e, assim, aprimora o sistema de suporte administrativo ao apresentar as melhores opções disponíveis.

Todavia, ao considerarmos algumas técnicas de *machine learning*, podemos observar que os direitos da personalidade ficam suscetíveis de sofrer violações. Isto porque uma base

---

177BRASIL. Congresso Nacional. Lei nº 9.784, de 29 de janeiro de 1999. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l9784.htm](https://www.planalto.gov.br/ccivil_03/leis/l9784.htm). Acesso em: 8 mar. 2025.

de dados será utilizada pela IA para processar o máximo de informações possível e convertê-las em dados para atingir o objetivo desejado. Em tais situações, onde há acesso extenso aos dados de todos ou da maioria dos cidadãos, há um risco maior de violação da privacidade, resultando em uma possível quebra do sigilo dos dados. A LGPD traz claramente quais são os fundamentos relacionados à proteção de dados pessoais, que servem para embasar toda e qualquer ação que envolva seu tratamento.

A violação do sigilo refere-se ao acesso a informações pessoais dos indivíduos, as quais deveriam ser obtidas em obediência aos artigos 6º a 9º, combinados com os artigos 23 a 26, da Lei Geral de Proteção de Dados<sup>178</sup>, ou por meio de uma autorização judicial.

Dados anonimizados são essenciais para o avanço da Inteligência Artificial, da Internet das Coisas, do aprendizado de máquina, das cidades inteligentes e da análise de comportamentos, entre outros. Sempre que possível, uma organização, seja pública ou privada, deve proceder à anonimização dos dados pessoais, pois isso aprimora a segurança da informação na organização e aumenta a confiança de seus públicos em seus serviços. Se um dado for anonimizado, a Lei Geral de Proteção de Dados (LGPD) não se aplicará a ele. Um dado só é considerado efetivamente anonimizado se não for possível, por meios técnicos ou outros, reconstruir o caminho para identificar o titular do dado. Se de alguma forma a identificação ocorrer, ele não é, de fato, um dado anonimizado, mas sim apenas pseudo anonimizado, e estará sujeito à LGPD.

Embora a Administração Pública deva adotar o princípio da minimização dos dados, ou seja, compartilhar apenas os dados estritamente necessários para a finalidade específica, os titulares dos dados devem ser informados sobre a finalidade do compartilhamento e como seus dados serão utilizados.

De uma abordagem analítica, podemos deduzir uma consideração sobre o compartilhamento de dados pela Administração Pública, notadamente quando se omite a anonimização de dados pessoais sensíveis dos cidadãos brasileiros: a violação do direito à direito à privacidade.

A vigilância constante, ou mesmo a mera percepção dela, leva à autocensura e à alteração de comportamento das pessoas. Agravando esta situação, a vigilância possibilitada pela Inteligência Artificial aumenta significativamente a sensação de ser monitorado sem interrupção.

---

178BRASIL. Congresso Nacional. **Lei nº 13.709**, de 14 de agosto de 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2015-2018/2018/Lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm). Acesso em: 8 mar. 2025.

O'NEIL<sup>179</sup> faz uma crítica contundente ao uso de modelos de decisão automatizados (ADM), aos quais ela chama de “algoritmos de destruição em massa”, mostrando o paradoxo de sua capacidade de gerar receita para mercados digitais, como os de mecanismos de busca, agregadores de *leads*<sup>180</sup> e marqueteiros, enquanto, ao mesmo tempo se configuram como um peso econômico e social, devido aos custos ocultos que impõem. A crítica se estende à insuficiência desses modelos em incorporar valores humanos essenciais, como justiça, equidade e o bem comum. Estes valores, por sua natureza, são resistentemente intangíveis e não podem ser quantificados de forma exata.

A autora aponta que, embora economistas tentem calcular custos relacionados a questões como poluição, contaminação por esgoto ou a extinção de espécies, os números jamais serão capazes de representar de maneira precisa tais danos, assim como a justiça ou a equidade, conceitos que não podem ser reduzidos a simples modelos matemáticos. Ela questiona a eficácia dos modelos algorítmicos ao sugerir que muitos deles, ao simplificarem excessivamente realidades complexas, acabam por falhar em capturar a verdadeira essência dos fenômenos que pretendem analisar. Sua sugestão é de que a única forma de corrigir tais falhas é abandonar modelos inadequados, como o de valor agregado, e, em vez de buscar ferramentas que tentem medir a eficácia de um servidor público, dedicar-se à criação de ciclos positivos de *feedback*, que envolvam servidores e usuários dos serviços no processo de avaliação. Esse ciclo virtuoso permitiria aprimorar a administração pública de maneira colaborativa e não punitiva.

Tais melhorias, no entanto, dependem da incorporação de valores humanos no processo de criação e análise dos modelos, como um esforço consciente de imbuir os sistemas com ética e justiça, mesmo que isso signifique sacrificar certa eficiência. Além disso, o texto amplia a crítica ao mostrar como os ADM, quando aplicados ao sistema judicial, podem replicar preconceitos e discriminações históricas. No caso dos modelos de reincidência criminal, os algoritmos podem reforçar desigualdades ao basearem suas decisões em dados como o histórico de crédito e as conexões sociais de um indivíduo. Em situações como essas, a autora propõe que os algoritmos sejam “emburrecidos”, ou seja, que se reduza a precisão de suas conclusões para evitar que discriminação e desigualdade sejam reforçadas.

---

179O'NEIL, Cathy. **Algoritmos de Destruição em Massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução de Rafael Abraham. Santo André: Rua do Sabão, 2020.

180*Leads* são potenciais clientes ou contatos que demonstraram algum interesse em um produto ou serviço, geralmente ao fornecerem informações como nome, e-mail ou outros dados de contato. Essas informações são coletadas por meio de diferentes canais, como sites, redes sociais ou campanhas publicitárias, e são consideradas uma oportunidade para as empresas se relacionarem com essas pessoas e convertê-las em clientes.

A ideia de “emburrecer” os algoritmos surge como uma resposta necessária para garantir que todos sejam tratados de maneira igual, como preconiza o princípio da igualdade no direito. O texto conclui com um olhar sobre os movimentos atuais para a auditoria de algoritmos, destacando exemplos de iniciativas acadêmicas que buscam detectar e corrigir vieses nos sistemas automatizados.

Projetos<sup>181</sup> como o da Universidade de Princeton, que utiliza robôs virtuais para simular diferentes perfis de pessoas e observar o tratamento que elas recebem, mostram o potencial de auditoria e transparência para identificar problemas e trabalhar para corrigi-los. Esses esforços, segundo a autora, representam passos importantes na busca por uma governança ética da tecnologia, garantindo que os algoritmos sirvam ao bem comum, sem comprometer valores fundamentais como a justiça e a equidade. Entre as conclusões, a autora alerta para o perigo de uma utopia tecnológica que se baseia em uma fé cega na capacidade dos algoritmos de resolverem problemas humanos e sociais.

A introdução e integração da Inteligência Artificial (IA) na Administração Pública apresentam um potencial considerável para otimizar os serviços prestados ao cidadão, proporcionando maior eficiência, agilidade e redução de custos. No entanto, esse avanço tecnológico traz consigo uma série de questões complexas que devem ser cuidadosamente avaliadas, pois a adoção da IA não pode se dar de maneira acrítica, sem que se considerem os impactos sobre os direitos fundamentais e os valores éticos que fundamentam a própria ordem jurídica do Estado. O controle exercido por agentes públicos é fundamental para evitar que decisões automatizadas, tomadas por sistemas de IA, resultem em impactos negativos ou injustos, em especial no que se refere à discriminação, à exclusão social ou à violação de direitos fundamentais.

A IA, por mais sofisticada que seja, não possui a capacidade de avaliar as complexidades das situações humanas de forma ética e contextualizada como um ser humano poderia fazer. O princípio da centralidade do sujeito humano afirma sua posição insubstituível como único ente capaz de exercer juízo crítico, responsabilidade moral e reflexão ética em processos decisórios. Diferentemente dos sistemas automatizados, que operam com base em padrões e estatísticas, o ser humano possui consciência, historicidade e intencionalidade.

---

181O WebTAP (*Web-Based Thematic Analysis Platform*) é uma plataforma desenvolvida para auxiliar pesquisadores na realização de análise temática, uma técnica qualitativa utilizada para identificar, analisar e interpretar padrões (ou “temas”) dentro de conjuntos de dados qualitativos, como entrevistas, textos ou transcrições.

Ele interpreta contextos, pondera consequências e toma decisões que vão além da lógica probabilística dos algoritmos. Esse conceito se desdobra em três dimensões fundamentais: a dimensão epistemológica destaca que o ser humano é o único agente dotado de compreensão simbólica e crítica. Enquanto as máquinas processam dados, o sujeito humano interpreta a realidade, atribui significados e reelabora informações a partir de sua experiência subjetiva.

A dimensão ético-moral evidencia que a tomada de decisões envolve valores, dilemas e implicações morais. Embora a máquina possa simular respostas éticas, ela não possui consciência para avaliar o impacto de suas decisões sobre a dignidade e a justiça. Apenas o ser humano pode assumir responsabilidade moral por suas escolhas. Por fim, a dimensão jurídico-normativa destaca que, no direito, a responsabilização é inerente à condição de sujeito.

Delegar decisões normativas à Inteligência Artificial cria um vácuo de *accountability*, pois sistemas são ferramentas, não agentes autônomos. A centralidade do sujeito humano resgata o princípio fundamental de que decisões jurídicas e políticas devem estar sempre ancoradas em alguém que possa responder por seus atos.

Diante dos riscos algorítmicos, a centralidade do sujeito humano surge como um contraponto à falsa noção de autonomia dos sistemas de IA. Reafirmá-la significa impedir a transferência de responsabilidade para entidades sem subjetividade, garantindo que a tecnologia permaneça um meio, e não um fim em si mesma.

Ao contrário do que muitos discursos tecnológicos sugerem, a IA não substitui a autonomia humana, apenas a simula.

Sempre que delegamos uma decisão a um sistema automatizado sem um olhar crítico, corremos o risco de transformar o sujeito – aquele que deveria estar no centro da reflexão e da responsabilidade – em um mero operador de um mecanismo opaco e impessoal. Isso se torna especialmente problemático quando a IA começa a influenciar decisões que afetam vidas humanas, como na aplicação do direito, na gestão de recursos e na formulação de políticas públicas.

Manter o ser humano no centro das decisões significa reafirmar que a tecnologia deve permanecer um instrumento, e não uma instância autônoma de julgamento. A responsabilidade não pode ser diluída em códigos e algoritmos; deve estar sempre vinculada a um sujeito capaz de responder por suas escolhas. Portanto, a supervisão humana não deve ser

relegada a um plano secundário, mas deve ser entendida como uma garantia de que as decisões tomadas pelo sistema estejam sempre em conformidade com os princípios fundamentais da Constituição. Ademais, é indispensável que, na aplicação da Inteligência Artificial na Administração Pública, se adote uma técnica interpretativa robusta, como a ponderação de princípios, a qual permite equilibrar os diferentes valores e direitos que, em determinadas situações, podem estar em colisão. Nesse sentido, a técnica da ponderação entre princípios deve ser utilizada de modo a garantir que os direitos fundamentais sejam devidamente respeitados, sem que a aplicação da IA resulte em excessos ou restrições desproporcionais. A liberdade, a igualdade e a dignidade da pessoa humana não podem ser relegadas a segundo plano em nome da eficiência ou da inovação tecnológica.

### 3.5. DA AVALIAÇÃO DE IMPACTO À MITIGAÇÃO DE RISCOS

O Canadá, visando a regulação de impacto algorítmico, elaborou o *Algorithmic Impact Assessment tool*<sup>182</sup>, uma ferramenta online, com perguntas que determinam o nível de impacto de um sistema de decisão automatizado, com fatores baseados em design, algoritmo, tipo de decisão, impacto e dados.

Outra ferramenta online que propõe criar fichas técnicas para documentar o contexto, as limitações e os potenciais vieses dos dados utilizados no ajuste de algoritmos, é o *Datasheets for Datasets*<sup>183</sup>.

No Reino Unido, o Ada Lovelace Institute lançou o *Examining Black Box*<sup>184</sup>, um relatório destinado a esclarecer os termos em auditorias de algoritmos e avaliações de impacto algorítmico, bem como o estado atual da pesquisa e prática. Concentrando-se na auditoria de algoritmos e na avaliação de impacto algorítmico, para cada um identificaram duas abordagens-chave. A primeira é dividida em duas ferramentas, a auditoria de viés e a inspeção regulatória. A segunda, composta por duas ferramentas, a avaliação de impacto

182CANADÁ. **Algorithmic Impact Assessment tool.** Disponível em: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>. Acesso em: 8 mar. 2025.

183GEBRU, Timnit; MORGENSTERN, Jamie; VECCHIONE, Briana; VAUGHAN, Jennifer Wortman; WALLACH, Hanna; DAUMÉ III, Hal; CRAWFORD, Kate. “Datasheets for Datasets”. **ArXiv:1803.09010**. Dec. 2021. Disponível em: <https://arxiv.org/pdf/1803.09010>. Acesso em: 8 mar. 2025.

184REINO UNIDO. Ada Lovelace Institute; DataKind UK. **Examining Tools for assessing algorithmic systems the Black Box: Tools for assessing algorithmic systems.** Disponível em: <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-%0AExamining-the-Black-Box-Report-2020.pdf>. Acesso em: 8 mar. 2025.

de riscos (*ex ante*) e a avaliação de impacto algorítmico (*ex post*).

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE)<sup>185</sup> compartilhou um *framework* para ajudar na classificação de sistemas de AI. Tal classificação pode ser útil na avaliação de riscos destes sistemas.

Para COZMAN & KAUFMAN<sup>186</sup>, atualmente, não há um padrão consistente de referência que possibilite a comparação da eficácia das técnicas de interpretabilidade e mitigação de viés. De acordo com os autores, um exemplo relevante de técnica de interpretabilidade, amplamente reconhecida, é o SHAP (*SHapley Additive exPlanations*).

Eles analisam a contribuição das variáveis individualmente, ou seja, avalia o efeito das interações entre os atributos separadamente, além de fornecer uma visão global do modelo.

Vale notar que o SHAP não é uma técnica singular, conforme apontam os autores, mas sim um conjunto de técnicas, cada uma delas com diferentes níveis de adequação para diferentes modelos de Inteligência Artificial. Utilizando fundamentos da teoria dos jogos cooperativos, o SHAP calcula o impacto de cada atributo na predição feita pelo modelo. A conclusão dos autores é de que a categorização dos sistemas de Inteligência Artificial por níveis de risco, conforme sugerido pela proposta de regulamentação da Comissão Europeia, representa um caminho adequado. O foco das autoridades reguladoras deve recair sobre os usos da IA que apresentam os maiores riscos para indivíduos e para a sociedade, levando em consideração os *trade-offs* entre os riscos e os benefícios que devem ser analisados tanto pelos desenvolvedores quanto pelos usuários e reguladores.

Um dos maiores equívocos ao utilizar sistemas de IA, dizem os autores, reside na chamada “promessa de objetividade”. Isso se refere à suposição de que os algoritmos, por serem executados por máquinas, garantem uma objetividade e/ou neutralidade livres de erros humanos. O mais apropriado seria tratar os sistemas de IA como aliados dos especialistas humanos, em vez de considerá-los autônomos. A ética aplicada à IA deve priorizar a mitigação de riscos, pois a abordagem adotada não pode ser universal. Tão pouco é viável controlar todos os avanços e usos dessa tecnologia. Assim, torna-se essencial estabelecer prioridades e concentrar esforços nos riscos mais significativos, conforme apontado por COZMAN & KAUFMAN.<sup>187</sup>

185OECD. “OECD Framework for the Classification of AI systems”, OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, 2022. Disponível em: <https://doi.org/10.1787/cb6d9eca-en>. Acesso em: 8 mar. 2025.

186COZMAN, Fabio Gagliardi, KAUFMAN, Dora. “Viés no aprendizado de máquina em sistemas de Inteligência Artificial: a diversidade de origens e os caminhos de mitigação”. *Revista USP*. São Paulo, n. 135, p. 195-210, outubro/novembro/dezembro 2022.

187COZMAN & KAUFMAN. *Op. cit.*, p. 209.

Nesse contexto, a União Europeia desponta como referência global. Por meio do Regulamento Geral de Proteção de Dados (GDPR) e da Estratégia de Inteligência Artificial da UE, tem liderado esforços para garantir que o uso de IA seja ético, transparente e responsável. Os princípios de governança da IA delineados pela OCDE e pelo Grupo de Especialistas de Alto Nível sobre IA da Comissão Europeia, por sua vez, exercem influência direta no modelo de avaliação de impacto algoritmo proposto adiante, propiciando uma gestão mais segura e equilibrada dessa tecnologia.

#### Objetivos do Modelo de Avaliação de Impacto Algorítmico:

- Garantir a conformidade com a Constituição Brasileira, a LGPD e os regulamentos internacionais;
- Proteger os direitos fundamentais, como a dignidade humana, a privacidade e a igualdade, nos processos algorítmicos;
- Promover a transparência e explicabilidade nos sistemas algorítmicos, garantindo a fiscalização das decisões automatizadas;
- Mitigar riscos e efeitos negativos da IA, como discriminação, exclusão social e danos à privacidade;
- Fomentar a responsabilidade social no uso de sistemas algorítmicos, promovendo um ambiente ético e justo.

#### Etapas do Modelo de Avaliação de Impacto Algorítmico (AIA):

##### 1. Identificação do Sistema Algorítmico

- Descrição do algoritmo, sua função e os objetivos;
- Identificação dos dados utilizados, considerando a fonte, a qualidade e a adequação dos dados em relação aos direitos dos indivíduos;

##### 2. Análise do Contexto

- Avaliação do contexto em que o algoritmo será implementado, incluindo os riscos sociais, econômicos e legais;
- Consideração do impacto do algoritmo nas relações de poder, considerando princípios de justiça social e equidade;

##### 3. Identificação de Impactos Potenciais

- Impactos Éticos e Sociais: Análise de potenciais discriminações, exclusões ou danos aos direitos fundamentais, com ênfase na dignidade humana e na igualdade;

- Impactos Econômicos: Consideração dos efeitos do algoritmo no mercado de trabalho e na competitividade;
- Impactos Ambientais: Caso aplicável, análise dos efeitos do algoritmo sobre o meio ambiente;
- Impactos Legais: Avaliação da conformidade com a Constituição Brasileira, a LGPD e outras legislações internacionais aplicáveis;

#### 4. Análise de Riscos

- Risco de Discriminação: Verificação de possíveis vieses no algoritmo que possam gerar discriminação baseada em gênero, raça, etnia ou classe social;
- Risco de Prejuízo à Privacidade: Avaliação do impacto sobre os dados pessoais, com foco na conformidade com a LGPD;
- Risco de Opacidade: Análise da explicabilidade e auditabilidade das decisões algorítmicas.
- Risco de Exclusão Social: Identificação de possíveis impactos negativos em grupos vulneráveis, como minorias ou pessoas com deficiência.
- Risco Econômico: Avaliação dos impactos da automação no emprego e na renda, com foco na justiça social e no desenvolvimento sustentável.

#### 5. Propostas de Mitigação

- Revisão de Dados e Algoritmos: Aperfeiçoamento dos dados de treinamento e do próprio algoritmo para mitigar vieses e promover a justiça;
- Garantias de Transparência: Desenvolvimento de mecanismos para garantir a explicabilidade das decisões algorítmicas;
- Monitoramento Contínuo: Estabelecimento de sistemas de monitoramento para avaliar os impactos a longo prazo e realizar ajustes conforme necessário;
- Capacitação e Formação: Capacitação de profissionais envolvidos no desenvolvimento e na implementação de algoritmos para que compreendam as implicações éticas e legais;

#### 6. Relatório de Impacto Algorítmico

- Elaboração de um relatório detalhado, que incluirá os resultados da avaliação, as recomendações para mitigação e as ações corretivas sugeridas;
- Publicação do relatório, garantindo a transparência e permitindo a consulta pública;
- Implementação das Ações de Mitigação;
- Implementação das medidas corretivas e acompanhamento de sua efetividade.

O Grupo de Especialistas de Alto Nível sobre Inteligência Artificial da OCDE, em suas diretrizes e relatórios sobre a governança da IA, classificam os riscos envolvidos na implementação de sistemas algorítmicos e propõem tratamentos para cada categoria de risco identificada.

As classificações de risco e os tratamentos recomendados são essenciais para a realização de uma Avaliação de Impacto Algorítmico (AIA), pois permitem que as organizações identifiquem, analisem e mitiguem, ou eliminem, os potenciais efeitos adversos de sistemas de IA em vários contextos.

Os riscos são classificados em diferentes níveis, conforme o impacto potencial desses sistemas, levando em conta o contexto de sua aplicação, a magnitude dos danos e a probabilidade de ocorrência, podendo ser agrupada nas seguintes categorias:

1. Riscos de Alto Impacto – Estes riscos envolvem consequências significativas para os indivíduos ou para a sociedade, podendo afetar gravemente os direitos fundamentais, como a dignidade, a privacidade, a liberdade e a não discriminação. Exemplos incluem:

- Discriminação e Viés: A IA pode reproduzir ou amplificar vieses sociais ou estruturais existentes, como preconceitos de raça, gênero, etnia ou classe social, afetando grupos vulneráveis;
- Desinformação e Manipulação: Sistemas de IA usados para criar ou disseminar desinformação, como em redes sociais, podem ter um impacto negativo nas democracias e na coesão social;
- Decisões Automatizadas sem Supervisão Humana Adequada: Decisões críticas (como em áreas de saúde, emprego e justiça) tomadas sem a devida supervisão humana podem resultar em danos irreparáveis à pessoa afetada;

2. Riscos de Impacto Médio – São riscos que, embora possam afetar gravemente alguns indivíduos ou setores, não têm um impacto tão generalizado ou irreversível quanto os riscos de alto impacto. Exemplos incluem:

- Transparência e Explicabilidade Limitada: Quando sistemas de IA são opacos e difíceis de auditar, isso pode gerar desconfiança e dificultar a responsabilização;
- Privacidade e Proteção de Dados: Embora não representem uma violação direta de direitos, falhas em proteger dados pessoais ou garantir consentimento adequado podem ter consequências significativas;

- Impactos Econômicos (Automação e Emprego): A automação de tarefas por IA pode causar mudanças significativas no mercado de trabalho, resultando em desigualdade econômica ou exclusão social;

3. Riscos de Baixo Impacto – Estes riscos têm consequências limitadas e geralmente estão associados a falhas menores ou a um controle eficaz sobre o sistema de IA. Exemplos incluem:

- Erros de Precisão ou Qualidade dos Dados: Quando os sistemas de IA não são suficientemente precisos ou bem treinados, podem cometer erros, mas esses erros tendem a ser corrigidos sem grandes repercussões;
- Falta de Adoção de Boas Práticas de Governança: Deficiências nas práticas internas de governança da IA, que não geram danos diretos ou imediatos, mas podem prejudicar a eficácia do sistema a longo prazo.

O tratamento dos riscos identificados na Avaliação de Impacto Algorítmico é uma parte determinante do processo de governança e deve ser adequado ao nível de risco classificado. Para cada tipo de risco, as diretrizes da OCDE e da Comissão Europeia sugerem abordagens específicas.

1. Tratamento de Riscos de Alto Impacto:

- Mitigação e Prevenção Rigorosa: Para riscos com alto impacto, como discriminação ou manipulação, é essencial realizar uma mitigação ativa e preventiva. Isso envolve a revisão constante dos dados utilizados para treinar os algoritmos, a implementação de mecanismos de controle de viés, e a promoção de transparência e explicabilidade dos sistemas;
- Supervisão Humana: Em sistemas de IA que possam afetar de forma significativa os direitos dos indivíduos, é crucial garantir a supervisão humana contínua, especialmente quando se trata de decisões críticas como aquelas em saúde, justiça e emprego;
- Transparência e Auditoria: Implementar mecanismos robustos de transparência, garantindo que os algoritmos possam ser auditados por órgãos independentes e que suas decisões possam ser compreendidas pelos usuários e pelas autoridades reguladoras;
- Regulamentação Estrita e Penalidades: Imposição de restrições mais rígidas e responsabilidades legais para prevenir a manipulação e o uso indevido de IA, além de sanções para as partes que não cumprirem com as normas éticas e legais;

2. Tratamento de Riscos de Impacto Médio:

- **Monitoração e Ajustes Contínuos:** Para riscos de impacto médio, como a violação de dados pessoais ou a falta de explicabilidade, deve-se realizar uma monitoração contínua do funcionamento do sistema, com ajustes regulares para corrigir problemas antes que se tornem graves;
- **Educação e Treinamento:** Capacitação das partes envolvidas, desde desenvolvedores até usuários finais, para garantir o uso adequado da IA e a conformidade com as boas práticas de proteção de dados e direitos dos indivíduos;
- **Revisão de Impactos Sociais e Econômicos:** A adoção de políticas públicas que abordem os impactos sociais e econômicos da automação, incluindo a requalificação da força de trabalho e o desenvolvimento de novas oportunidades de emprego;
- **Garantia de Conformidade com a LGPD:** Assegurar que as soluções de IA cumpram os princípios de proteção de dados pessoais estabelecidos pela Lei Geral de Proteção de Dados (LGPD), incluindo a realização de Relatórios de Impacto sobre a Proteção de Dados (RIPD).

### 3. Tratamento de Riscos de Baixo Impacto:

- **Ajustes Operacionais Menores:** Riscos de baixo impacto podem ser mitigados por ajustes simples nos processos internos de governança, como melhoria na qualidade dos dados e no treinamento dos algoritmos;
- **Adoção de Boas Práticas de Governança:** Implementação de normas e diretrizes internas que promovam a qualidade regulatória e a responsabilidade social, garantindo que os sistemas sejam eficazes e éticos, sem representar riscos significativos;
- **Monitoramento e Melhoria Contínua:** Estabelecimento de processos de monitoramento e melhoria contínua para identificar precocemente falhas que possam ser corrigidas antes que se tornem um problema maior.

A OCDE e o Grupo de Especialistas de Alto Nível sobre IA da Comissão Europeia fornecem uma sólida estrutura regulatória para que os sistemas de IA sejam utilizados de maneira responsável, transparente e ética, assegurando que os riscos sejam mitigados de acordo com sua gravidade e impacto potencial.

A conformidade com essas diretrizes não só protege os direitos fundamentais, mas também promove a confiança pública na IA, permitindo o desenvolvimento sustentável e inclusivo da tecnologia.

## CONSIDERAÇÕES FINAIS

Este trabalho procurou compreender a relação entre a cultura patriarcal e a origem da desigualdade de gênero, o que exige uma análise crítica de como, ao longo da história humana, a suposta superioridade física e mental dos homens em relação às mulheres se transformou em uma construção cultural, originando relações assimétricas que resultam em inúmeras injustiças, especialmente pela inferiorização das mulheres nessas relações.

Como exposto, os sistemas algorítmicos são, simultaneamente, produtos e produtores das dinâmicas sociais nas quais estão inseridos, refletindo em si as relações de poder e as normas culturais que os originaram. Regular esses processos, portanto, não se limita a mitigar danos, mas buscar promover a ressignificação das formas de interação e relação no espaço digital e, em última instância, na própria sociedade.

A transparência no funcionamento do algoritmo é fundamental para compreender como os resultados discriminatórios são gerados e para identificar possíveis correções técnicas aos vieses. Além disso, é necessário realizar uma análise humana dos vieses identificados, especialmente os de gênero, considerando a complexidade dos mesmos, que são expressões de tendências e distribuições estatísticas incomuns.

Detectar vieses, por si só, não é suficiente para uma avaliação completa de seus impactos éticos e jurídicos. Identificar se um viés específico constitui discriminação exige uma larga compreensão dos contextos sociais em que a IA é aplicada, o que demanda capacidade para fazer julgamentos morais.

Os algoritmos, devido à sua peculiar tecnicidade, estabelecem, sob a égide do binômio eficiência/ineficiência, um subsistema normativo. Este fenômeno exige a assimilação e a regulação dos riscos inerentes aos sistemas de Inteligência Artificial (IA), configurando uma nova fronteira para a interpretação e aplicação das normas jurídicas.

Não obstante a existência de um número considerável de normas e regulamentos voltados à governança de dados, nosso sistema jurídico carece de um conjunto normativo mais robusto, que abranja tanto a governança quanto a regulação dos algoritmos, com o intuito de mitigar e, quando possível, prevenir os riscos algorítmicos. A responsabilidade difusa e o automatismo moral são manifestações do risco algorítmico que afetam tanto a *accountability* quanto a autonomia humana em processos decisórios mediados por Inteligência Artificial.

Essa responsabilidade difusa surge quando a complexidade dos sistemas algorítmicos e a multiplicidade de agentes envolvidos em sua criação e uso tornam difícil determinar quem deve responder por suas consequências.

Programadores desenvolvem os modelos, empresas os implementam, órgãos reguladores estabelecem normas e usuários os aplicam em suas rotinas. Contudo, se nenhum desses atores assume a responsabilidade final pelos impactos da IA, forma-se um vácuo normativo e ético. A ausência de definição torna-se ainda mais problemática quando a tecnologia é utilizada para justificar decisões sem que haja um sujeito identificado que possa ser questionado. Quando um juiz, um médico ou um gestor público se apoia em um algoritmo para decidir e se exime da responsabilidade sob o argumento de que “foi o sistema que indicou esse resultado”, ocorre uma transferência indevida da autoridade humana para uma entidade que, por sua própria natureza, não pode ser responsabilizada moral ou juridicamente.

O automatismo moral, assim, manifesta-se quando os usuários deixam de questionar criticamente as recomendações dos sistemas de IA e passam a segui-las de maneira automática. Esse fenômeno decorre da crença equivocada de que a tecnologia opera com precisão e imparcialidade, levando à percepção de que um sistema algorítmico seria inerentemente mais justo ou racional do que o julgamento humano. O problema se agrava em áreas como o direito, a segurança pública e a saúde, onde as decisões impactam diretamente vidas humanas. Quando o operador ou usuário da IA abdica de sua capacidade crítica e transfere sua responsabilidade decisória para o sistema, há um esvaziamento ético: o ser humano, que deveria ocupar o centro das decisões, passa a agir como um mero executor de comandos algorítmicos, perdendo sua autonomia reflexiva e moral.

Ambos os problemas estão interligados e revelam um ciclo preocupante no contexto do risco algorítmico. A reprodução de preconceitos nos sistemas de IA reforça desigualdades estruturais sob a falsa premissa de neutralidade.

A responsabilidade difusa impede a identificação de culpados e dificulta a correção de injustiças e o automatismo moral, por sua vez, fortalece a passividade dos usuários, que, ao confiar cegamente nos sistemas, deixam de exercer seu papel crítico. Esse encadeamento leva à substituição indevida do julgamento humano pelo cálculo algorítmico, deslocando a centralidade do sujeito e criando um cenário no qual a injustiça é automatizada e ninguém se sente responsável por ela. Considerando que a IA tem o potencial de tomar decisões autônomas que afetam diretamente vidas humanas e bens, a regulação torna-se imprescindível

para estabelecer responsabilidades claras quando danos ou erros ocorrem.

A regulação não só garante que as partes prejudicadas possam buscar reparação, mas também assegura que os desenvolvedores e operadores de IA sejam responsabilizados por suas ações. Ademais, a regulação é fundamental para proteger os direitos dos cidadãos, garantindo que os produtos e serviços baseados em IA atendam a elevados padrões de qualidade e segurança, atuando também como um mecanismo para prevenir a formação de monopólios no setor.

Pela perspectiva dos sistemas de Niklas Luhmann, o sistema democrático é dotado da capacidade de autoconstrução e auto sustentação, o que lhe permite transitar entre tensões internas e externas sem perder a essência que o define: a convivência harmoniosa entre a pluralidade e o dissenso. A principal diferença reside no fato de que, hoje, os algoritmos se apresentam como agentes invisíveis dessa mediação, influenciando nossas interações de maneira sutil, porém determinante.

Acolher o dissenso como expressão legítima do pluralismo, sem que isso comprometa sua integridade ou finalidade maior, é servir à dignidade humana. Nesse sentido, a resiliência democrática está atrelada à sua capacidade de equilibrar forças opostas, conciliando continuidade e ruptura, consenso e divergência. Os algoritmos, alinhados aos valores republicanos, podem se tornar verdadeiros agentes de inclusão, transparência e inovação social, operando como instrumentos éticos de justiça e igualdade. Assim, o futuro da democracia repousa em sua habilidade de transformar tensões em avanços, afirmando-se como um sistema vivo, dinâmico e essencialmente voltado para o bem comum.

Os sistemas algorítmicos refletem dinâmicas sociais, podendo perpetuar relações de poder e normas culturais discriminatórias. Regular esses processos significa buscar ressignificação das interações no espaço digital e na sociedade. É, portanto, uma questão não apenas técnica, mas também principiológica. O Direito, amparado nos princípios constitucionais e nas técnicas de ponderação, deve promover decisões solidamente fundamentadas nos princípios éticos que sustentam a convivência democrática.

Compreender a Inteligência Artificial exige distinguir a mera simulação de processos cognitivos da efetiva presença de atributos humanos. A IA, em sua essência, é um mecanismo computacional que reproduz padrões estatísticos sem a vivência subjetiva que define o sujeito moral.

Essa distinção torna-se fundamental ao analisar os vieses algorítmicos, especialmente o viés de correspondência ou atribuição, que ocorre quando se projetam indevidamente características humanas sobre sistemas desprovidos de intencionalidade e sensibilidade.

Os julgamentos morais baseiam-se na distinção entre o certo e o errado, o justo e o injusto. Apesar de sua aparente simplicidade, resultam de processos complexos e dinâmicos influenciados por interações sociais. Trata-se de um fenômeno sofisticado que demanda habilidades humanas específicas. Portanto, a análise de questões éticas, incluindo as relacionadas à não discriminação, é melhor conduzida pelo processo cognitivo mais complexo – o humano.

Os vieses não surgem de maneira espontânea. Algoritmos de IA, quando baseados em dados, não geram preconceitos por si mesmos, mas podem reproduzi-los sem a devida intervenção humana. Esse fenômeno ocorre nos seguintes estágios: na coleta de dados, uma vez que estes refletem preconceitos preexistentes; na preparação dos dados de treinamento; no processo de seleção e processamento dos atributos fornecidos ao algoritmo; e na tomada de decisões, quando o algoritmo faz propostas e decisões ao longo do ciclo de desenvolvimento da IA.

Para a construção de sistemas verdadeiramente equitativos, é necessário que os dados de treinamento reflitam a diversidade dos grupos sociais e demográficos, o que exige uma curadoria rigorosa, garantindo que os dados não sejam apenas representativos de uma população homogênea, mas de todas as camadas sociais, de forma balanceada.

A avaliação de sistemas algorítmicos deve ser contínua e iterativa. Além de monitorar o desempenho dos modelos em tempo real, é imprescindível realizar testes periódicos para verificar se o sistema está perpetuando ou ampliando discriminações. Métodos de auditoria algorítmica podem ser empregados, com foco na identificação de vieses ocultos que possam ter sido negligenciados no processo de desenvolvimento.

Ao destacar a autorreferencialidade dos sistemas sociais, entendemos que esses sistemas possuem a capacidade de se autodefinir e auto reforçar, construindo sua própria realidade e significado a partir de suas próprias estruturas internas, donde nos permitimos especular que algoritmos e sistemas de IA também podem desenvolver suas estruturas internas de autorreprodução e auto reforço. Se um algoritmo de IA é treinado com conjuntos de dados que refletem certos preconceitos ou vieses presentes na sociedade, como as normas patriarcais, ele pode aprender e internalizar esses padrões discriminatórios.

À medida que o algoritmo continua a operar e interagir com novos dados, ele pode

perpetuar e amplificar esses vieses, reproduzindo-os em suas decisões e resultados.

Independentemente da técnica empregada, a modelagem de sistemas deve sempre ser orientada por um compromisso ético e técnico que busque evitar a perpetuação ou o agravamento das desigualdades, especialmente de gênero, assegurando a equidade nas decisões geradas. A supervisão humana é essencial para garantir que os algoritmos não apenas processem dados, mas também evitem a reprodução ou amplificação das desigualdades históricas. Uma solução “ótima” em termos puramente algorítmicos pode não ser considerada ideal se não atender aos princípios de justiça e inclusão social. O design de sistemas algorítmicos que minimizem ou eliminem os vieses de gênero requer uma integração cuidadosa entre técnicas matemáticas e computacionais, sempre acompanhadas de uma análise ética que priorize a justiça e a equidade. Isso exige a necessidade de supervisão humana constante, para garantir que os algoritmos não apenas cumpram sua função de processar dados de maneira eficiente, mas também contribuam para uma sociedade mais justa, inclusiva e livre de discriminação.

A regulamentação deve assegurar que as tecnologias adotadas não violem os direitos constitucionais, como o direito à privacidade, à liberdade, à igualdade e à dignidade da pessoa humana. Isso se reflete, principalmente, na prevenção e eliminação de vieses. O Direito, enquanto sistema especializado na estabilização de expectativas, exige inovação na criação de soluções que protejam tanto a autonomia individual quanto os direitos fundamentais diante das incertezas do mundo algorítmico.

Na legislação sobre IA, como em qualquer sistema jurídico moderno, a norma inicial, que se apresenta de maneira mais abstrata e indeterminada, é apenas o primeiro passo de um processo contínuo de concretização. Esse processo envolve não apenas a elaboração de regulamentações complementares, mas também a intervenção ativa do Judiciário e a participação da sociedade na construção de uma norma jurídica que desponta como em constante evolução. A regulamentação complementar, portanto, não deve ser encarada apenas como uma obrigação técnica e burocrática, mas como um instrumento fundamental para a realização da justiça material e a efetivação dos direitos fundamentais no contexto da IA.

A criação de normas secundárias, que possam detalhar os critérios, parâmetros e procedimentos necessários para a aplicação dos direitos estabelecidos na regulamentação de sistemas inteligentes é essencial para que as intenções legislativas se concretizem de forma plena e eficaz. Nesse sentido, o legislador deve atuar com uma visão sistêmica e dinâmica,

considerando que a realidade tecnológica e social é suscetível a mudanças rápidas e que a regulação deve ser capaz de acompanhar essas transformações sem perder sua eficácia.

Com relação ao papel do Judiciário, destacamos a importância da interpretação constitutiva das normas. A interpretação judicial, ao aplicar a norma em casos concretos, deve buscar não apenas a coerência formal, mas também a efetividade dos princípios e direitos previstos na Constituição e nas leis infraconstitucionais.

Em um contexto como o da Inteligência Artificial, onde os desafios são inéditos e muitas vezes imprevisíveis, a atuação do Judiciário será decisiva para preencher as lacunas deixadas pelos novos princípios de direitos fundamentais e para garantir a proteção desses direitos diante das inovações tecnológicas. A interpretação judicial “criativa” através da técnica de ponderação de princípios é um mecanismo que pode assegurar que os princípios, embora indeterminados, se ajustem à realidade dos casos, evitando que se tornem normas vazias ou ineficazes. Ao aplicar a ponderação de princípios, o Judiciário deve avaliar qual princípio deve prevalecer em cada caso específico, considerando a natureza do conflito, as circunstâncias e os valores constitucionais envolvidos.

A participação social também se configura como um elemento central da concretização dos direitos e da efetividade da regulamentação. A democratização da norma é uma das características essenciais de um direito que é verdadeiramente fundamental. A sociedade deve ser parte ativa na formulação das regulamentações que vão concretizar os princípios constitucionais, especialmente no que toca a áreas tão inovadoras e complexas como a Inteligência Artificial. A consulta pública, os processos de deliberação democrática e a transparência nos processos regulatórios são condições indispensáveis para garantir que os direitos fundamentais não sejam tratados de forma abstrata ou elitista, mas, sim, de maneira inclusiva e representativa da diversidade de interesses sociais.

A atuação do Estado, ao regular a IA, deve ser entendida como uma responsabilidade coletiva, que envolve não apenas os órgãos públicos, mas também a sociedade civil, os acadêmicos e os profissionais da área. O alinhamento com padrões internacionais, conforme sugerido pela articulação da regulamentação brasileira com as diretrizes da OCDE e da União Europeia, é outro ponto a ser considerado para a concretização efetiva da legislação. A globalização dos problemas relacionados à IA exige cooperação internacional, pois os sistemas de IA operam além das fronteiras nacionais e seus efeitos podem ser transnacionais.

A ordem jurídica deve estar integrada a uma rede de normas que ultrapassem o âmbito nacional, sem perder de vista as peculiaridades locais, sociais e culturais.

O Relatório de Avaliação de Impacto Algorítmico desempenha um papel decisivo na eliminação de vieses de gênero em sistemas de IA, por meio de uma abordagem regulatória contínua e supervisão constante. Ao identificar e avaliar os riscos algorítmicos, a AIA possibilita uma análise detalhada das fontes de viés nos dados utilizados no treinamento dos algoritmos, com destaque para as desigualdades históricas e culturais presentes nesses dados.

A transparência nos métodos de desenvolvimento e a possibilidade de revisão dos processos de treinamento dos algoritmos permitem a detecção desses padrões discriminatórios, oferecendo diretrizes específicas para sua correção, incluindo a prevenção da perpetuação de estereótipos de gênero. Os sistemas de IA devem ser projetados com foco na equidade e inclusão, e a regulação deve assegurar que os algoritmos operem de maneira justa, levando em consideração a diversidade humana e evitando a discriminação.

A intervenção humana é imprescindível para garantir que os resultados gerados pelos sistemas sejam analisados de forma ética e contextual, possibilitando a correção de eventuais vieses identificados.

A AIA também assegura que as tecnologias adotadas respeitem os direitos fundamentais previstos na Constituição, como a não discriminação com base no sexo, a privacidade e a dignidade humana. A regulamentação oriunda da AIA contribui para a criação de um sistema regulatório sólido, no qual os algoritmos são monitorados para evitar a perpetuação ou amplificação das desigualdades entre homens e mulheres. Dessa forma, a Inteligência Artificial pode se tornar uma ferramenta eficaz para a promoção de uma sociedade mais justa, inclusiva e livre de discriminação.

À medida que a IA continua a se desenvolver e adquirir maior autonomia, novos debates e desafios jurídicos inevitavelmente surgirão.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AMAZON WEB SERVICES (AWS). “O que é uma rede neural?”. Disponível em: <https://aws.amazon.com/pt/what-is/neural-network/>. Acesso em: 8 mar. 2025.
- ANEESH, A. “*Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies*”. **7th International Summer Academy on Technology Studies**, Deutschlandsberg, Austria. Disponível em: [https://www.researchgate.net/profile/A-Aneesh/publication/254843955\\_Technologically\\_Coded\\_Authority\\_The\\_Post-Industrial\\_Decline\\_in\\_Bureaucratic\\_Hierarchies/links/5bf45834a6fdcc3a8de46665/Technologically-Coded-Authority-The-Post-Industrial-Dcline-in-Burea](https://www.researchgate.net/profile/A-Aneesh/publication/254843955_Technologically_Coded_Authority_The_Post-Industrial_Decline_in_Bureaucratic_Hierarchies/links/5bf45834a6fdcc3a8de46665/Technologically-Coded-Authority-The-Post-Industrial-Dcline-in-Burea). Acesso em: 8 mar. 2025.
- ARENDDT, Hannah. **A condição humana**. Tradução de Roberto Raposo. 13. ed. rev. [Reimpr.]. Rio de Janeiro: Forense Universitária, 2020.
- AXELROD, Robert. **The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration**. Princeton: Princeton University Press, 1997.
- BALKIN, Jack M. “Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation” (September 9, 2017). **UC Davis Law Review**, (2018 Forthcoming), Yale Law School, Public Law Research Paper N°. 615, Ng. Disponível em: <https://openyls.law.yale.edu/handle/20.500.13051/4699>. Acesso em: 8 mar. 2025.
- BARLETTA, Fabiana, PALMEIRA, Carolina S. de Sá. “Vulnerabilidade da Mulher, Autonomia Privada e o Exercício de Direitos Reprodutivos e Sexuais”. **Vulnerabilidades e suas Dimensões Jurídicas**. (Coord.) Fabiana Barletta e Vitor Almeida. Indaiatuba: Foco, 2023.
- BEAUVOIR, Simone. **O Segundo Sexo**. Tradução de Sérgio Milliet. 2a ed. Rio de Janeiro: Nova Fronteira, 2009.
- BOBBIO, Norberto. **A Era dos Direitos**. Tradução de Carlos Nelson Coutinho. Apresentação de Celso Lafer. Rio de Janeiro: Elsevier, 2004. 7a reimpressão.
- BOURDIEU, Pierre. **O poder simbólico**. Tradução de Fernando Tomaz. Difusão Editorial: Lisboa, 1989.
- BOURDIEU, Pierre. **A dominação masculina**, 2a ed. Tradução: Maria Helena Kühner. Rio de Janeiro: Bertrand Brasil, 2002.
- BOURDIEU, Pierre. **Razões Práticas sobre a Teoria da Ação**. Tradução de Mariza Corrêa. Campinas: Papirus, 2009.

BUOLAMWINI, Joy, GEBRU, Timnit. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. **Proceedings of Machine Learning Research** 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency.

Disponível em: <https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/>. Acesso em: 8 mar. 2025.

BURLE, Caroline, CORTIZ, Diego. “Mapeamento de Princípios de Inteligência Artificial”. **CEWEB**.

Disponível em: <https://ceweb.br/publicacoes/mapeamento-de-principios-de-inteligencia-artificial/?page=1>. Acesso em: 8 mar. 2025.

BRAGA, Nathália. “Erro em reconhecimento facial constrange psicóloga em conferência”. **Defensoria Pública do Estado do Rio de Janeiro**, Rio de Janeiro, 12 jul. 2024.

Disponível em: <https://defensoria.rj.def.br/noticia/detalhes/29955-Erro-em-reconhecimento-facial-constrange-psicologa-em-conferencia>. Acesso em: 8 mar. 2025.

BRASIL. **Constituição da República Federativa do Brasil**.

Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 678**, de 6 de novembro de 1992.

Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/d0678.htm](http://www.planalto.gov.br/ccivil_03/decreto/d0678.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 1.973/1996**, de 9 de junho de 1994.

Disponível em:

[https://www.planalto.gov.br/ccivil\\_03/decreto/1996/d1973.htm#:~:text=DECRETO%20N%C2%BA%201.973%2C%20DE%201%C2%BA,9%20de%20junho%20de%201994](https://www.planalto.gov.br/ccivil_03/decreto/1996/d1973.htm#:~:text=DECRETO%20N%C2%BA%201.973%2C%20DE%201%C2%BA,9%20de%20junho%20de%201994). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 3.321**, de 30 de dezembro de 1999. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decreto/d3321.htm](https://www.planalto.gov.br/ccivil_03/decreto/d3321.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 4.377**, de 13 de setembro de 2002.

Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/2002/d4377.htm](http://www.planalto.gov.br/ccivil_03/decreto/2002/d4377.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 8.078**, de 11 de setembro de 1990.

Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/18078compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/18078compilado.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 9.319**, de 21 de março de 2018.

Disponível em:

[https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/decreto/D9319.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/D9319.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto nº 10.411**, de 30 de junho de 2020.

Disponível em:

[https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2020/decreto/d10411.htm#:~:text=DECRETO%20N%C2%BA%2010.411%20DE%2030%20DE%20JUNHO%20DE%202020&text=Regulamenta%20a%20an%C3%A1lise%20de%20impacto,que%20lhe%20confere%20o%20art.](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/d10411.htm#:~:text=DECRETO%20N%C2%BA%2010.411%20DE%2030%20DE%20JUNHO%20DE%202020&text=Regulamenta%20a%20an%C3%A1lise%20de%20impacto,que%20lhe%20confere%20o%20art.) Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 12.965**, de 22 de abril de 2014.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm).  
Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Decreto-Lei nº 4.657**, de 4 de setembro de 1942, com redação dada pela **Lei nº 12.376**, de 30 de dezembro de 2010.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decreto-lei/del4657compilado.htm](https://www.planalto.gov.br/ccivil_03/decreto-lei/del4657compilado.htm).  
Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Emenda Constitucional nº 45**, de 30 de dezembro de 2004.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/emendas/emc/emc45.htm](https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc45.htm).  
Acesso em: 8 mar. 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação. **Estratégia Brasileira para a Transformação Digital (E-Digital)**.  
Disponível em: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>. Acesso em: 8 mar. 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação. **Estratégia Brasileira de Inteligência Artificial – EBIA**. Julho de 2021.  
Disponível em: [https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento\\_referencia\\_4-979\\_2021.pdf](https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf). Acesso em: 8 mar. 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação. **Portaria MCTIC nº 1.122**, de 19 de março de 2020.  
Disponível em:  
[https://antigo.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria\\_MCTIC\\_n\\_1122\\_de\\_19032020.html](https://antigo.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria_MCTIC_n_1122_de_19032020.html). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 6.938**, de 31 de agosto de 1981.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l6938.htm](https://www.planalto.gov.br/ccivil_03/leis/l6938.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 7.116**, de 5 de janeiro de 1989.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l7716.htm](https://www.planalto.gov.br/ccivil_03/leis/l7716.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 13.709**, de 14 de agosto de 2018.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm).  
Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 13.848**, de 25 de junho de 2019.

Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/lei/113848.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/113848.htm). Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Lei nº 13.874**, de 20 de setembro de 2019.  
Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/lei/113874.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/113874.htm). Acesso em: 8 mar. 2025.

BRASIL. Câmara dos Deputados. **Projeto de Lei 21/2020**.  
Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 8 mar. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 896/2023**.  
Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/03/07/proposta-que-criminaliza-misoginia-comeca-a-tramitar-no-senado>. Acesso em: 8 mar. 2025.

BRASIL. Senado Federal. **Projeto de Lei 2338/2023**.  
Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 8 mar. 2025.

BRASIL. Senado Federal. **PL 2338/2023**. (Texto aprovado para submissão à Câmara dos Deputados).  
Disponível em:  
<https://legis.senado.leg.br/sdleg-getter/documentodm=9881643&ts=1734719326347&disposition=inline>. Acesso em: 8 mar. 2025.

BRASIL. Câmara dos Deputados. **Projeto de Lei 6.032/2005**.  
Disponível em: <https://www.camara.leg.br/propostas-legislativas/302782>. Acesso em: 8 mar. 2025.

BRASIL. Congresso Nacional. **Projeto de Lei 5.051/2019**.  
Disponível em: <https://www.congressonacional.leg.br/materias/materias-bicameras/-/ver/pl-5051-2019>. Acesso em: 8 mar. 2025.

BRASIL. Senado Federal. **Projeto de Lei 5.691/2019**.  
Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 8 mar. 2025.

BRASIL. Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo Sobre Inteligência Artificial no Brasil. **Relatório Final**.  
Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/2-relatorio-final-versao-completa-cjsubia.pdf>. Acesso em: 8 mar. 2025.

BRASIL. Ministério da Economia. “Normativos da União deverão ter análise de impacto regulatório para serem editados”.  
Disponível em: <https://www.gov.br/economia/pt-br/assuntos/noticias/2020/julho/normativos-da-uniao-deverao-ter-analise-de-impacto-regulatorio-para-serem-editados>. Acesso em: 8 mar. 2025.

BRASIL. Supremo Tribunal Federal. **Recurso Extraordinário nº 349.703/RS**. Plenário, julgamento em 3 de dezembro de 2008; relator: Ministro Carlos Ayres Britto. DJe nº 236, 12 de dezembro de 2008.

Disponível em:

[https://jurisprudencia.s3.amazonaws.com/STF/IT/RE\\_349703\\_RS\\_1278971642175.pdf?AWSAccessKeyId=AKIARMMD5JEAO67SMCVA&Expires=1753805389&Signature=hpGhky2Qkv4IERixp0NTGK0Lwo%3D](https://jurisprudencia.s3.amazonaws.com/STF/IT/RE_349703_RS_1278971642175.pdf?AWSAccessKeyId=AKIARMMD5JEAO67SMCVA&Expires=1753805389&Signature=hpGhky2Qkv4IERixp0NTGK0Lwo%3D). Acesso em: 8 mar. 2025.

BRASIL. Supremo Tribunal Federal. **Habeas Corpus 87.585-8-TO**. Relator: Min. Marco Aurélio. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=597891>. Acesso em: 8 mar. 2025.

BRASIL. Supremo Tribunal Federal. **ADI 4.439/DF**, Relator: Ministro Roberto Barroso. Data de julg.: 27/09/2017.

Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=15085915>. Acesso em: 8 mar. 2025.

BRASIL. Supremo Tribunal Federal. **Agravo Regimental no Recurso Extraordinário com Agravo no 639337-SP**, Relator(a): Min. Celso de Mello.

Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=627428>. Acesso em: 8 mar. 2025.

BRAZ, Matheus V. “Heteromação e microtrabalho no Brasil”. **Sociologias**, Porto Alegre, ano 23, n. 57, mai-ago 2021, p. 134-172.

Disponível em: <http://doi.org/10.1590/15174522-111017>. Acesso em: 8 mar. 2025.

CANADÁ. **Algorithmic Impact Assessment tool**.

Disponível em: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>. Acesso em: 8 mar. 2025.

CANOTILHO, J. J. Gomes. **Direito Constitucional e Teoria da Constituição**. 7a ed., 11ª reimpressão. Coimbra: Almedina, 2012.

CARLSMITH, Joseph. “Is Power-Seeking AI an Existential Risk?”. **Open Philanthropy**, 2022.

Disponível em: <https://arxiv.org/pdf/2206.13353>. Acesso em: 8 mar. 2025.

CARTA CAPITAL. “Deputada denuncia ‘racismo algorítmico’ após IA gerar imagem com arma em uma favela”. Publ. em 26/10/2023.

Disponível em: <https://www.cartacapital.com.br/cartaexpressa/deputada-denuncia-racismo-algoritmico-apos-ia-gerar-imagem-com-arma-em-uma-favela/>. Acesso em: 8 mar. 2025.

CEBRIAN, Fabiana S. P. Faraco; PRUDENTE, Gustavo do Amaral; GUEDES, Marcelo Santiago; SILVA, Maria Carolina Ferreira da; SÁ, Maria Luiza Duarte; MORAES, Thiago Guimarães. **Biometria e reconhecimento facial**: estudos preliminares. Radar Tecnológico nº 2. ANPD: Brasília, 2024.

COELHO, Henrique; NASCIMENTO, Rafael; ALVES, Raoni. “Mulher presa após reconhecimento facial é solta; mandado de prisão já tinha sido cumprido”. **G1-Globo**. Publicado em 04/01/2024.

Disponível em: <https://g1.globo.com/rj/rio-de-janeiro/noticia/2024/01/04/mulher-presa-apos-reconhecimento-facial-e-solta-mandado-de-prisao-ja-tinha-sido-cumprido.ghtml>.

Acesso em: 8 mar. 2025.

COLLING, Ana Maria. **Tempos diferentes**, discursos iguais: a construção do corpo feminino na história. Dourados: Ed. UFGD, 2014.

CORTE INTERAMERICANA DE DIREITOS HUMANOS: **Caso Márcia Barbosa e Outros Vs. Brasil, Exceções Preliminares, Mérito, Reparações e Custas**, Sentença de 7 de setembro de 2021, par. 138-141.

Disponível em: <http://cepia.org.br/wp-content/uploads/2021/12/Sentenc%CC%A7a-marcia-barbosa.pdf>. Acesso em: 8 mar. 2025.

COZMAN, Fabio Gagliardi, KAUFMAN, Dora. “Viés no aprendizado de máquina em sistemas de Inteligência Artificial: a diversidade de origens e os caminhos de mitigação”. **Revista USP**. São Paulo, n. 135, p. 195-210, outubro/novembro/dezembro 2022.

DANTAS, Maria Eduarda Borba. “Dimensões da violência contra mulheres defensoras de direitos humanos no Brasil”. **ONU MULHERES BRASIL**.

Disponível em: <http://www.onumulheres.org.br/wp-content/uploads/2021/11/Relatorio-Defensoras-Violencia1.pdf>. Acesso em: 8 mar. 2025.

DA SILVA, Flávio S. Corrêa, HIRATA, Nina S. T. “Inteligência Ética. Ética e Regulação na Inteligência Artificial”. **Revista da Sociedade Brasileira de Computação**, julho/2022, nº 47.

DEEP LEARNING BOOK. “Introdução às Redes Neurais Convolucionais” – Capítulo 40. **Data Science Academy**.

Disponível em: <https://www.deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/>. Acesso em: 8 mar. 2025.

FABRIS, Alessandro; PURPURA, Alberto; SILVELLO, Gianmaria; SUSTO, Gian Antonio. “Gender Stereotype Reinforcement: Measuring the Gender Bias Conveyed by Ranking Algorithms”. **Department of Information Engineering**, University of Padua, Italy. Disponível em: <https://arxiv.org/pdf/2009.01334>. Acesso em: 8 mar. 2025.

DIMOULIS, Dimitri; MARTINS, Leonardo. **Teoria geral dos direitos fundamentais**. 5. ed. rev., atual. e ampl. São Paulo: Atlas, 2014.

FELTRIN, Fernando (aut.). **Ciência de Dados e Aprendizado de Máquina: Uma Introdução às Redes Neurais Artificiais**. São Paulo-SP: Amazon, 2019. E-book (296p.) (eBook Kindle).

FLORIDI, Luciano. **The fourth revolution: how the infosphere is reshaping human reality**. UK: Oxford University Press, 2014.

FREUD, Sigmund. **Psicologia das Massas e Análise do Eu**. Brasil: Lebooks, 2020.

FREUD, Sigmund. (1912-1914) - **Obras completas volume 11: Totem e tabu, Contribuição à história do movimento psicanalítico e outros textos**. São Paulo: Companhia das Letras, 2012.

FREUD, Sigmund. (1923-1925) **Obras completas volume 16: O Eu e o Id, "Autobiografia" e Outros Textos**. Tradução de Paulo César de Souza. São Paulo: Companhia das Letras, 2011.

FREUD, Sigmund. **Além do princípio de prazer**. Tradução de Renato Zwick. Porto Alegre: L&PM, 2016.

FRIEDMAN, B.; NISSENBAUM, H. (1996). "Bias in computer systems". **ACM Transactions on Information Systems (TOIS)**, 14(3), pp. 330-347.

GEBRU, Timnit; MORGENSTERN, Jamie; VECCHIONE, Briana; VAUGHAN, Jennifer Wortman; WALLACH, Hanna; DAUMÉ III, Hal; CRAWFORD, Kate. "Datasheets for Datasets". **ArXiv:1803.09010**. Dec. 2021.  
Disponível em: <https://arxiv.org/pdf/1803.09010>. Acesso em: 8 mar. 2025.

GILLIGAN, Carol. **In a Different Voice: Psychological Theory and Women's Development**. Boston: Harvard University Press, 1982.

HAN, Byung-Chul. **Psicopolítica e Neoliberalismo**. Belo Horizonte: Ayiné, 2018.

IBM. "O que são redes neurais".  
Disponível em: <https://www.ibm.com/br-pt/topics/neural-networks>. Acesso em: 8 mar. 2025.

INSTITUTO PATRÍCIA GALVÃO. "Cultura e Raízes da Violência contra Mulheres". **Dossiê Violência Contra Mulheres**.  
Disponível em: <https://dossies.agenciapatriciagalvao.org.br/violencia/violencias/cultura-e-raizes-da-violencia/>.

Acesso em: 8 mar. 2025.

KAHN, Jeremy. “Inteligência Artificial de Musk cria fake news com poucas palavras”. **Exame**. Publicado em 16/02/2019.

Disponível em: <https://exame.com/tecnologia/inteligencia-artificial-de-musk-cria-fake-news-com-poucas-palavras/>. Acesso em: 8 mar. 2025.

KELSEN, Hans. **A Democracia**. Tradução de Ivone Castilho Benedetti, Jefferson Luiz Camargo, Marcelo Brandão Cipolla, Vera Barkow. São Paulo: Martins Fontes, 2000.

KELSEN, Hans. **Teoria pura do direito**. 7a ed. Tradução de João Baptista Machado. São Paulo: Martins Fontes, 2006.

KOWALSKI, R., “Algorithm = Logic + Control”, **Communications of the ACM**, 22(7), 1979, pp. 424-436.

Disponível em: <https://www.doc.ic.ac.uk/~rak/papers/algorithm%20=%20logic%20+%20control.pdf>. Acesso em: 8 mar. 2025.

KRAKOVNA, Victoria, UESATO, Jonathan, MIKULIK, Vladimir, RAHTZ, Matthew, EVERITT, Tom, KUMAR, Ramana, KENTON, Zac, LEIKE, Jan, LEGG, Shane. “Specification gaming: the flip side of AI ingenuity”. **DeepMind Safety Research**: 2020.

Disponível em: <https://deepmindsafetyresearch.medium.com/specification-gaming-the-flip-side-of-ai-ingenuity-c85bdb0deeb4>. Acesso em: 8 mar. 2025.

LACAN, Jacques. **A Identificação: seminário 1961 – 1962**, Tradução de Ivan Corrêa e Marcos Bagno. Recife: Centro de Estudos Freudianos do Recife, 2003.

LERNER, Gerda. **A Criação do Patriarcado**: história da opressão das mulheres pelos homens. Tradução: Luiza Sellera. São Paulo: Cultrix, 2019.

LÈVI-STRAUSS, Claude. **As Estruturas Elementares do Parentesco**. Tradução de Mariano Ferreira. Petrópolis: Vozes, 1982.

LUHMANN, Niklas. **Sistemas sociais**: esboço de uma teoria geral. Tradução de Antônio C. Luz Costa, Roberto Dutra Torres, Marco Antônio dos Santos Casanova. Petrópolis: Vozes, 2016.

LUHMANN, Niklas. **La sociedad de la sociedad**. México: Universidad Iberoamericana, 2006.

MacINTYRE, A. **After Virtue – A Study in Moral Theory**. Third Edition. Indiana: University of Notre Dame Press, 2007.

MAZZUOLI, Valerio. **Controle jurisdicional da convencionalidade das leis**. 4a. ed. São Paulo: RT, 2016.

MERLEAU-PONTY, Maurice. **Fenomenologia da percepção**, 2.ed. Tradução: Carlos Alberto Ribeiro de Moura. São Paulo: Martins Fontes, 1999.

MIRBABAIE, M., STIEGLITZ, S. & MARX, J. “Digital Detox”. **Bus Inf Syst Eng** **64**, 239–246 (2022).

Disponível em: <https://doi.org/10.1007/s12599-022-00747-x>. Acesso em: 8 mar. 2025.

MOREIRA, Adílson José. **Tratado de Direito Antidiscriminatório**. São Paulo: Contracorrente, 2020.

MOSS, Emanuel; WATKINS, Elizabeth; SINGH, Ranjit; ELISH, Madeleine Clare; METCALF, Jacob. **Assembling Accountability**: Algorithmic Impact Assessment for the Public Interest (June 29, 2021).

Disponível em: <http://dx.doi.org/10.2139/ssrn.3877437>. Acesso em: 8 mar. 2025.

MOZELLI, Rodrigo. “IA do mal? Gemini ofende usuário e pede que ele morra”. **OLHAR DIGITAL**, publicado em 20/11/2024.

Disponível em: <https://olhardigital.com.br/2024/11/20/pro/ia-do-mal-gemini-ofende-usuario-e-pede-que-ele-morra/>. Acesso em: 8 mar. 2025.

MÜLLER, Léo. “Tay: Twitter conseguiu corromper a IA da Microsoft em menos de 24 horas”. **TECMUNDO**, Publicado em 24/03/2023.

Disponível em: <https://www.tecmundo.com.br/inteligencia-artificial/102782-tay-twitter-conseguiu-corromper-ia-microsoft-24-horas.htm>. Acesso em: 8 mar. 2025.

NUNES, José Coelho D.; MORATO DE ANDRADE, O. “O Uso da Inteligência Artificial Explicável enquanto ferramenta para compreender decisões automatizadas: Possível Caminho para aumentar a legitimidade e confiabilidade dos modelos algorítmicos?”. **Revista Eletrônica do Curso de Direito da UFSM**, [S. l.], v. 18, n. 1, p. e69329, 2023.

Disponível em: <https://periodicos.ufsm.br/revistadireito/article/view/69329>. Acesso em: 8 mar. 2025.

OECD. “OECD Framework for the Classification of AI systems”, **OECD Digital Economy Papers**, No. 323, OECD Publishing, Paris, 2022.

Disponível em: <https://doi.org/10.1787/cb6d9eca-en>. Acesso em: 8 mar. 2025.

O’NEIL, Cathy. **Algoritmos de Destruição em Massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução de Rafael Abraham. Santo André: Rua do Sabão, 2020.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Plataforma do Cairo, 1994. **Relatório da Conferência Internacional sobre População e Desenvolvimento.**

Disponível em: <http://www.unfpa.org.br/Arquivos/relatorio-cairo.pdf>. Acesso em: 8 mar. 2025.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. IV Conferência das Nações Unidas sobre a Mulher, 1995. **Declaração e Plataforma de Ação de Pequim.**

Disponível em:

[https://www.onumulheres.org.br/wp-content/uploads/2013/03/declaracao\\_beijing.pdf](https://www.onumulheres.org.br/wp-content/uploads/2013/03/declaracao_beijing.pdf).

Acesso em: 8 mar. 2025.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS, Comitê CEDAW: **Recomendação Geral nº 33** sobre o acesso das mulheres à justiça, 3 de agosto de 2015, CEDAW/C/GC/33, par. 26, p. 14.

Disponível em: <https://assets-compromissoeatitude-ipg.sfo2.digitaloceanspaces.com/2016/02/Recomendacao-Geral-n33-Comite-CEDAW.pdf>. Acesso em: 8 mar. 2025.

PATEMAN, Carole. **O Contrato Sexual.** Paz e Terra, São Paulo, 1993.

PETTIT, Philip. **Republicanism: A Theory of Freedom and Government.** Oxford: Clarendon Press, 1997.

PEREZ, Caroline Criado. **Invisible Women: Data Bias in a World Designed for Men.** New Yor: Adam Press, 2019. [recurso eletrônico AZW3].

REBELLO, Aiuri. “Da placa de carro ao CPF: Conheça o Córtex, sistema de vigilância do governo que integra de placa de carro a dados de emprego”. **Intercept Brasil.**

Disponível em: <https://www.intercept.com.br/2020/09/21/governo-vigilancia-cortex/>. Acesso em: 8 mar. 2025.

REINO UNIDO. DataKind UK. “Examining Tools for assessing algorithmic systems the Black Box: Tools for assessing algorithmic systems”. **Ada Lovelace Institute.**

Disponível em: <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf>. Acesso em: 8 mar. 2025.

REIS, Carolina; ALMEIDA, Eduarda; DA SILVA, Felipe; DOURADO, Fernando. “Relatório sobre o uso de tecnologias de reconhecimento facial e câmeras de vigilância pela administração pública no Brasil”. Brasília: **Laboratório de Políticas Públicas e Internet**, 2021.

REUTERS. “Amazon ditched AI recruiting tool that favored men for technical jobs”. **The Guardian.** 11 out 2018.

Disponível em: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>. Acesso em: 8 mar. 2025.

ROCHA, Rosely; MUNIZ, Marise. **Central Única dos Trabalhadores-CUT**. “INSS usa robôs para analisar pedidos de benefícios e milhares são negados”.

Disponível em: <https://www.cut.org.br/noticias/inss-usa-robos-para-analisar-pedidos-de-beneficios-e-mais-de-300-mil-sao-negados-1cd3>. Acesso em: 8 mar. 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. Tradução de Regina Célia Simille. Rio de Janeiro: Elsevier, 2013.

SAMPAIO, José Adércio Leite (Coord.); FURBINO, Meire; BOCCHINO, Lavínia Assis; LIMA, Maria Jocélia Nogueira (Org.). **A Inteligência Artificial: a (des)serviço do estado de direito**. Belo Horizonte: CAPES: Programa de Pós-graduação em Direito-PUC Minas: RTM, 2023.

SANTIAGO, Eloisa Samy. “Gênero e tecnologia: por que a maioria dos assistentes de IA são mulheres?”. **Prêmio Elisa Frota Pessoa 2023** [livro eletrônico] / [organizadores Vitória Holz, Nina Pougy. 1. ed. Rio de Janeiro: Museu do Amanhã/Instituto de Desenvolvimento e Gestão - IDG, 2024, PDF.

SARLET, Ingo Wolf. **Dignidade da Pessoa Humana e Direitos Fundamentais na Constituição Federal de 1988**. 4.ed. rev. atual. Porto Alegre: Livraria do Advogado, 2006.

SCIAMMARELLA, Ana Paula. **Magistratura e gênero: uma análise da condição profissional feminina no Judiciário Fluminense**. Rio de Janeiro: Autografia, 2020. [recurso eletrônico EPUB].

SCHREIBER, Anderson. **Direitos da personalidade**. 3.ed. São Paulo: Atlas, 2014.

SCOTT, Joan. “Gênero: Uma Categoria Útil de Análise Histórica”. **Educação & Realidade**, v.15, n. 2, jul./dez., 1990.

SOUZA, Antônio Carlos Zambroni; PINHEIRO, Carlos Alerto Murari. **Introdução à modelagem, análise e simulação de sistemas dinâmicos**. Rio de Janeiro: Interciência, 2008.

SPENCER, E.A.; HENEGHAN C.; “Confirmation bias”. **Catalogue Of Bias** 2018. Disponível em: [www.catalogueofbiases.org/biases/confirmationbias](http://www.catalogueofbiases.org/biases/confirmationbias). Acesso em: 8 mar. 2025.

STEINBRUCH, David. **Um Estudo de Algoritmos para Classificação Automática de Textos Utilizando Naive-Bayes** (Tese). Colaborador(es): Daniel Schwabe – Orientador Ruy Luiz Milidiu – Coorientador. Sistema Maxwell, Coleção Digital; Pontifícia Universidade Católica do Rio de Janeiro – PUC-RIO: 2007, pp 20-39.

Disponível em: <https://www.maxwell.vrac.puc-rio.br/colecao.php?strSecao=resultado&nrSeq=9637@1>. Acesso em: 8 mar. 2025.

TELES, Edson. “Ação Política Híbrida e a Dissolução da Cidadania”. **Revista de Filosofia Moderna e Contemporânea**, Brasília, v.8, n.3, dez. 2020, p. 81-103.

Disponível em: <https://doi.org/10.26512/rfmc.v8i3.34494>. Acesso em: 8 mar. 2025.

THOMPSON, G. **Introducing functional grammar**. London: Arnold, 1996.

UNESCO. **Declaração Universal sobre Bioética de Direitos Humanos**. Paris, 2006.

Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000146180\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000146180_por). Acesso em: 8 mar. 2025.

UNESCO. **Report Of The Social And Human Sciences Commission (SHS)**. Paris, 2021.

Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>. Acesso em: 8 mar. 2025.

UNIÃO EUROPEIA. European Commission. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. **Communication: Building Trust in Human Centric Artificial Intelligence**. Brussels. 8.4.2019.

Disponível em: <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>. Acesso em: 8 mar. 2025.

UNIÃO EUROPEIA. **Regulamento 2024/1689** do Parlamento Europeu e do Conselho de 13 de junho de 2024 (Regulamento da Inteligência Artificial).

Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689>. Acesso em: 8 mar. 2025.

VALENTE, Jonas. “Tecnologias de reconhecimento facial são usadas em 37 cidades no país”. **Agência Brasil**.

Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-09/tecnologias-de-reconhecimento-facial-sao-usadas-em-37-cidades-no-pais>. Acesso em: 8 mar. 2025.

VAN DEN HOVEN, Jeroen. “Computer Ethics And Moral Methodology”. **Metaphilosophy**. Vol. 28, N. 3, July, 1997. Cambridge: Blackwell Publishers, pp. 234-248.

VENTURA, Layse. “Número de acidentes com piloto-automático da Tesla é maior do que se sabia”. **OLHAR DIGITAL**, publicado em 10/06/2023.

Disponível em: <https://olhardigital.com.br/2023/06/10/carros-e-tecnologia/numero-de-acidentes-com-piloto-automatico-da-tesla-e-maior-do-que-se-sabia/>. Acesso em: 8 mar. 2025.

VIEIRA, James Batista; BARRETO, Rodrigo T. de Souza. **Governança, gestão de riscos e integridade**. Brasília: ENAP, 2019.

WOJCIECHWSKI, Paola Bianchi; DA ROSA, Alexandre Morais. **Vieses da Justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva**. 2a.ed. Florianópolis: Ematis, 2021.

YEUNG, K. (2018), “Algorithmic regulation: A critical interrogation”. **Regulation & Governance, 12: 505-523**.

Disponível em: <https://doi.org/10.1111/rego.12158>. Acesso em: 8 mar. 2025.

ZHANG, Baobao; DREKSLER, Noemi; ANDERLJUNG, Markus; KAHN, Lauren; GIATTINO, Charlie; DAFOE, Allan; HOROWITZ, Michael C. “Forecasting AI Progress: Evidence from a Survey of Machine Learning Researchers”. **Cornell University: 2022**; Seccion 4.1: Human-level Machine Intelligence Forecasts from 2019 Cross-sectional Sample.

Disponível em: <https://arxiv.org/abs/2206.04132>. Acesso em: 8 mar. 2025.

ZUBOFF, Shoshana. **A Era do capitalismo de Vigilância**. Tradução de George Schlesinger. Rio de Janeiro: Intrínseca, 2021[Edição digital].