

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, TECNOLOGIA E  
SOCIEDADE

**REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL: DESAFIOS E  
PERSPECTIVAS PARA UMA SOCIEDADE EM MUTAÇÃO**

ANA CRISTINA OLIVEIRA MAHLE

São Carlos – SP  
2026

ANA CRISTINA OLIVEIRA MAHLE

**REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL: DESAFIOS E  
PERSPECTIVAS PARA UMA SOCIEDADE EM MUTAÇÃO**

Tese apresentada ao Programa de Pós-Graduação em  
Ciência, Tecnologia e Sociedade, do Centro de  
Educação e Ciências Humanas, da Universidade  
Federal de São Carlos, como parte dos requisitos para  
a obtenção do título de Doutorado em Ciência,  
Tecnologia e Sociedade.

Orientador: Prof. Dr. Vinicio Carrilho Martinez

São Carlos – SP  
2026



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
Centro de Educação e Ciências Humanas – CECH  
Programa de Pós-Graduação em Ciência, Tecnologia e

---

**Folha de Aprovação**

---

Defesa de Tese de Doutorado da candidata Ana Cristina Oliveira Mahle, intitulada Regulação Da Inteligência Artificial: Desafios E Perspectivas Para Uma Sociedade Em Mutação, realizada em 26/05/2026.

**Comissão Julgadora:**

Profa. Dra. Tainá Reis de Souza (UFSCar)  
Profa. Dra. Luciana de Souza Gracioso (UFSCar)  
Profa. Dra. Marcela Taiane Schiavi (USP)  
Prof. Dr. Vinícius Alves Scherch (UFSCar)  
Profa. Dra. Marcia Fernanda de Camargo (UFSCar)

Mahle, Ana Cristina Oliveira

Regulação da Inteligência Artificial: Desafios e perspectivas para uma sociedade em mutação / Ana Cristina Oliveira Mahle -- 2026.  
210f.

Tese (Doutorado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador (a): Vinício Carrilho Martinez

Banca Examinadora: Vinício Carrilho Martinez, Marcia Fernanda de Camargo, Marcela Taiane Schiavi, Luciana de Souza Gracioso, Tainá Reis de Souza Vinícius Alves Scherch

Bibliografia

1. Inteligência Artificial. 2. Direitos Humanos. 3. Proteção de Dados. I. Mahle, Ana Cristina Oliveira. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180

## DEDICATÓRIA

*À minha família, que sustentou cada madrugada de estudo com amor e paciência.*

*Aos que acreditaram que o direito e a tecnologia precisam conversar e que essa conversa pode, de fato, transformar a sociedade.*

*E a todos os que ainda acreditam que pensar criticamente é o maior ato de resistência que um ser humano pode praticar.*

## AGRADECIMENTOS

Ao Prof. Dr. Vinicio Carrilho Martinez, meu orientador, pela confiança depositada neste trabalho, pela liberdade intelectual que me concedeu e pela generosidade com que conduziu cada etapa desta jornada. Sua orientação foi, ela mesma, um exercício de pensamento complexo.

À banca examinadora, Prof<sup>a</sup>. Marcela, Prof.<sup>a</sup> Tainá, Prof.<sup>a</sup> Márcia, Prof<sup>a</sup>. Luciana e Prof.<sup>o</sup> Vinícius, pelo rigor, pela generosidade e pela riqueza das contribuições que tornaram este trabalho melhor do que eu seria capaz de fazê-lo sozinha. Cada questionamento foi um presente intelectual.

Ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade da UFSCar, por me acolher em um espaço onde o direito pode dialogar com a tecnologia, a educação e a filosofia sem pedir licença.

À minha família, pelo amor incondicional e pela compreensão nas ausências que uma tese inevitavelmente impõe. Vocês são a razão de tudo.

Aos meus colegas, clientes e parceiros da área jurídica e de governança corporativa, cujas experiências práticas alimentaram muitas das reflexões presentes neste trabalho. A teoria só faz sentido quando encontra a realidade.

E à Inteligência Artificial, não como autora, mas como ferramenta. Este trabalho é, também, um experimento consciente sobre o que ela pode e não pode fazer. A mente preparada continua sendo humana.

*"Qualquer tecnologia suficientemente avançada é indistinguível da magia."*

Arthur C. Clarke, Profiles of the Future (1962)

*Platão acreditava que a criação não vinha do artista, mas através dele.*

*As Musas eram entidades externas que sopravam inspiração na mente humana; o poeta era apenas o canal.*

*Séculos depois, Steven Pressfield retomou essa ideia: a criatividade é uma força que existe além de nós, esperando ser convocada por quem tem disciplina e preparo para recebê-la.*

*A Inteligência Artificial, quando usada com intenção, pensamento crítico e domínio do próprio campo, não substitui o criador.*

*Ela age como aquelas entidades que os antigos já descreviam: um agente externo que amplifica o que já existe dentro do humano.*

*Não gera do nada. Potencializa quem já sabe o que quer dizer.*

*A musa mudou de forma. A condição humana, não.*

*O acaso favorece a mente preparada. A IA, também.*

— Ana Mahle

## RESUMO

A pesquisa aborda os desafios e as perspectivas da regulação da Inteligência Artificial (IA) em uma sociedade em constante mutação. Inserida no campo de Ciência, Tecnologia e Sociedade (CTS), parte de uma reflexão sobre proteção de dados e autodeterminação informativa, a pesquisa avança para explorar as implicações éticas e sociais da IA considerando a ausência de um marco regulatório específico no Brasil. O estudo busca entender como os projetos de lei em discussão no país, como o PL 2338/2023, podem moldar a futura regulamentação da IA promovendo uma abordagem baseada em riscos que considera as vulnerabilidades sociais e os direitos fundamentais. A escolha do termo "mutação" no título da tese reflete a imprevisibilidade e os impactos complexos da IA, que, assim como uma mutação genética, pode trazer tanto avanços como desafios. A pesquisa explora como a IA já está integrada nas esferas sociais, políticas e econômicas, não apenas por meio da automatização de processos, mas por sistemas capazes de aprender, prever e decidir. Essa integração traz benefícios, como eficiência e inovação, mas também riscos estruturais, entre eles a despersonalização das relações laborais e a fragilização de garantias sociais. Por meio de uma análise qualitativa e comparativa, a pesquisa examina as propostas de regulação da IA, como o PL 21/2020 e o PL 2338/2023, comparando-os com modelos internacionais, como o regulamento da União Europeia. A conclusão aponta para a urgência de uma regulação adaptativa e responsiva, capaz de acompanhar o ritmo das inovações e preservar a centralidade humana nas decisões automatizadas. Assim, a pesquisa contribui para o campo CTS ao propor diretrizes éticas e regulatórias que buscam equilibrar inovação, responsabilidade social e proteção de direitos fundamentais.

**Palavras Chave:** Inteligência Artificial; Direitos Humanos; Proteção de Dados; Regulação, Ética.

## ABSTRACT

The research addresses the challenges and prospects of regulating Artificial Intelligence (AI) in a constantly changing society. Situated within the field of Science, Technology and Society (STS), it begins with a reflection on data protection and informational self-determination, advancing to explore the ethical and social implications of AI in light of the absence of a specific regulatory framework in Brazil. The study seeks to understand how bills currently under discussion, such as Bill No. 2338/2023, may shape the future regulation of AI, promoting a risk-based approach that considers social vulnerabilities and fundamental rights. The choice of the term mutation in the title reflects the unpredictability and complex impacts of AI, which, like a genetic mutation, can bring both progress and challenges. The research examines how AI is already integrated into social, political, and economic spheres, not only through process automation, but also through systems capable of learning, predicting, and making decisions. This integration brings benefits such as efficiency and innovation, yet also poses structural risks, including the depersonalization of labor relations and the weakening of social protections. Through a qualitative and comparative analysis, the study examines the proposals for AI regulation, such as Bills No. 21/2020 and 2338/2023, comparing them with international models, particularly the European Union's Artificial Intelligence Act. The conclusion highlights the urgency of developing an adaptive and responsive regulatory framework, capable of keeping pace with technological advances while preserving human centrality in automated decision-making. Thus, this research contributes to the STS field by proposing ethical and regulatory guidelines that aim to balance innovation, social responsibility, and the protection of fundamental rights.

**Keywords:** Artificial Intelligence; Human Rights; Data Protection; Regulation, Ethics.

## LISTA DE FIGURAS

Figura 1 – Capa da Revista The Economist de 2017 .....	30
Figura 2 – Tweeds de Tay.....	44
Figura 3 - Isolamento da Realidade .....	55
Figura 4 - Quadro: “A Arte do Pensar” prompts de Ana Mahle .....	144
Figura 5 - Théâtre D’opéra Spatial by Jason Allen (Ghosh; Fossas, 2022) .....	145
Figura 6 - Quadro: Má-formação de dedos pela IA.....	162

## LISTA DE ABREVIATURAS E SIGLAS

AI ACT	<i>Artificial Intelligence Act</i>
AI	Artificial Intelligence
AWS	Amazon Web Services
CEO	Chief Executive Officer
CF	Constituição Federal
COVID-19	Coronavirus Disease 2019
CPUs	Central Processing Units
CUDA	Compute Unified Device Architecture
DALL·E	<i>Deep Learning Artificial Intelligence Language Model for Image Generation</i>
AIA	Avaliação de Impacto Algorítmico
ANPD	Autoridade Nacional de Proteção de Dados
ARPA	<i>Advanced Research Projects Agency</i>
ARPANET	<i>Advanced Research Projects Agency Network</i>
CBL	Câmara Brasileira do Livro
CDC	Código de Defesa do Consumidor
CJUSBIA	Comissão de Juristas
CLT	Consolidação das Leis do Trabalho
CNJ	Conselho Nacional de Justiça
CNNs	<i>Convolutional Neural Networks</i>
COMPAS	<i>Correctional Offender Management Profiling for Alternative Sanctions</i>
CTIA	Comissão Temporária sobre Inteligência Artificial no Brasil
CTS	Ciência, Tecnologia e Sociedade
CUP	Convenção da União de Paris
DARPA	<i>Defense Advanced Research Projects Agency</i>
DNA	Ácido Desoxirribonucleico
DOI	Digital Object Identifier
DSA	Digital Services Act
ECA	Estatuto da Criança e do Adolescente
EUA	Estados Unidos da América
GPU	Graphics Processing Unit

IBM	International Business Machines
STS	Science, Technology and Society
FMI	Fundo Monetário Internacional
GAI	<i>Generative Artificial Intelligence</i>
GAN	<i>Generative Adversarial Networks</i>
GDPR	Regulamento Geral de Proteção de Dados
GIDE	<i>Global Information Dominance Experiments</i>
GPS	<i>General Problem Solver</i>
GPT	<i>Generative Pretrained Transformer</i>
GPUs	<i>Graphics Processing Unit</i>
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados
LLMs	<i>Large Language Models</i>
MTE	Ministério do Trabalho e do Emprego
NLP	<i>Natural Language Processing</i>
PL	Projeto Lei
RBU	Rendimento Básico Universal
RTB	<i>Real-Time Bidding</i>
SIA	Sistema Nacional de Regulação e Governança de Inteligência Artificial
STF	Supremo Tribunal Federal
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
TI	Tecnologias de Informação
TSE	Tribunal Superior Eleitoral
UBI	Renda Básica Universal

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	<b>16</b>
CONTEXTUALIZAÇÃO DA PESQUISA .....	16
OBJETIVOS E QUESTÕES DE PESQUISA .....	23
<b>Objetivo Geral</b> .....	23
<b>Objetivos Principais</b> .....	23
<b>Questões de Pesquisa</b> .....	24
METODOLOGIA ADOTADA.....	24
IMPORTÂNCIA DA PESQUISA PARA O CAMPO CTS.....	26
MOTIVAÇÃO DO NOME DO TRABALHO .....	27
<b>CAPÍTULO 1 - A R-EVOLUÇÃO DA IA: REFLEXÕES SOCIAIS, POLÍTICAS E ECONÔMICAS</b> .....	<b>29</b>
1.1 PANORAMA HISTÓRICO DA INTELIGÊNCIA ARTIFICIAL.....	29
1.1.1 A Nova Era.....	29
1.1.2 Conceito e Evolução de IA .....	32
<b>1.2 IMPACTOS SOCIAIS, ECONÔMICOS, POLÍTICOS E ÉTICOS DA IA</b> .....	<b>39</b>
1.2.1 Impactos Sociais .....	40
1.2.2 Impactos Econômicos .....	61
1.2.3 Impactos Políticos.....	72
1.2.4 Impactos Éticos .....	85
1.2.5 Impactos Jurídicos da Tecnologia .....	91
<b>CAPÍTULO 2 - A REVOLUÇÃO DA IA: IMPACTOS NO MERCADO DE TRABALHO, DESUMANIZAÇÃO E PRECARIZAÇÃO DAS POLÍTICAS PÚBLICAS</b> .....	<b>96</b>
2.1 IMPACTO DA IA NO MERCADO DE TRABALHO .....	96
2.2 DESUMANIZAÇÃO .....	100
2.3 PRECARIZAÇÃO DAS POLÍTICAS PÚBLICAS .....	106
<b>CAPÍTULO 3 – REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL</b> .....	<b>114</b>
3.1 CONTEXTO JURÍDICO DO PL 21/20 E DO PL 2338/23 .....	116
3.1.1 PL 21/2020.....	116

3.1.2 PL 2338/2023.....	118
3.1.3 Substitutivo ao PL 2338/2023.....	120
3.1.4 Situação Atual do PL 2338/2023.....	121
3.2 ANÁLISE COMPARATIVA: PL 2120 vs. PL 2338.....	124
3.2.1 Responsabilidade Civil.....	124
3.2.2 Avaliação de Impacto Algorítmico e Governança.....	125
3.2.3 Diferenças na Abordagem Regulamentar: PL 2338 e PL 21/2020 em Relação a Sistemas de Risco Excessivo.....	126
3.3 INFLUÊNCIA DO GDPR NA LGPD E A CHEGADA DO AI ACT NA EUROPA: IMPACTOS PARA O BRASIL E SIMILARIDADES COM O PL 2338.....	128
3.4 PROPOSTA DE REFORMA DO CÓDIGO CIVIL.....	130

**CAPÍTULO 4 - CONTEÚDO GERADO POR IA GENERATIVA E A PROBLEMÁTICA DO DIREITO AUTORAL ..... 134**

4.1 <i>LARGE LANGUAGE MODEL</i> (LLM) E <i>PROMPT</i> .....	136
4.2 AVANÇO DA IA GENERATIVA.....	138
4.3 DIREITO AUTORAL, ORIGINALIDADE E PLÁGIO.....	139
4.4 CASOS PRÁTICOS DE DIREITOS AUTORAIS E CONTEÚDOS GERADOS POR IA.....	145
4.5 USO DE OBRAS PROTEGIDAS POR DIREITOS AUTORAIS PARA MINERAÇÃO DE DADOS.....	150
4.6 CRIADOR, CRIATURA E FERRAMENTA: AUTORIA DA CRIAÇÃO GERADA POR INTELIGÊNCIA ARTIFICIAL GENERATIVA.....	152
4.7 O PAPEL DA HONESTIDADE INTELLECTUAL NO USO DA IA GENERATIVA E OS RISCOS DA REPETIÇÃO AUTOMÁTICA.....	156

**CAPÍTULO 5 - EDUCAÇÃO DIGITAL, CONSCIÊNCIA ÉTICA E FORMAÇÃO PARA O USO RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL..... 165**

5.1 A EDUCAÇÃO COMO EIXO DE REGULAÇÃO SOCIAL DA TECNOLOGIA.....	166
5.2 A POLÍTICA NACIONAL DE EDUCAÇÃO DIGITAL (LEI Nº 14.533/2023): INSTRUMENTO PARA A FORMAÇÃO ÉTICA NA ERA DA INTELIGÊNCIA ARTIFICIAL.....	167

5.3 O ECA DIGITAL (LEI Nº 15.211/2025) E A PROTEÇÃO AMPLIADA DE CRIANÇAS E ADOLESCENTES .....	170
5.5 INTEGRAÇÃO ENTRE EDUCAÇÃO, INOVAÇÃO E REGULAÇÃO: CAMINHOS PARA UMA POLÍTICA PÚBLICA INTEGRADORA .....	172
<b>CAPÍTULO 6 - BONS USOS E MAUS USOS DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO .....</b>	<b>176</b>
<b>CONCLUSÃO.....</b>	<b>183</b>
<b>REFERÊNCIAS.....</b>	<b>188</b>

## INTRODUÇÃO

### CONTEXTUALIZAÇÃO DA PESQUISA

O percurso acadêmico que me conduziu à pesquisa de doutorado em “Regulação da Inteligência Artificial: desafios e perspectivas para uma sociedade em mutação” têm suas raízes fincadas nas investigações realizadas durante o mestrado, na mesma instituição, a Universidade Federal de São Carlos, no Programa de Ciência, Tecnologia e Sociedade (CTS). A dissertação de mestrado, intitulada "A Autodeterminação Informativa como Fundamento da LGPD", não apenas pavimentou o caminho teórico e metodológico para abordar as nuances da legislação em proteção de dados, mas também ampliou minha percepção sobre as interconexões entre tecnologia, sociedade e os imperativos éticos subjacentes.

Este alicerce acadêmico e intelectual, profundamente enraizado na análise crítica das políticas de proteção de dados sob a égide da Lei Geral de Proteção de Dados (LGPD), naturalmente evoluiu para uma inquirição mais ampla e complexa no domínio da Inteligência Artificial (IA). A decisão de prosseguir com a pesquisa sobre a regulação da IA e suas implicações éticas foi catalisada por um conjunto de desenvolvimentos legislativos e debates públicos significativos no Brasil, marcados, inicialmente, pela proposta do Projeto de Lei (PL) nº 21/2020 (Brasil, 2020), que teve como objetivo o estabelecimento de fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil.

A transformação do cenário regulatório, evidenciada pela atuação da Comissão de Juristas (CJUSBIA) (Brasil, 2022a) responsável pela elaboração de um substitutivo aos PLs 5051/2019 e 21/2020, reflete a dinâmica e a complexidade das discussões envolvendo a IA. As audiências públicas que congregaram mais de cinquenta especialistas multissetoriais e o subsequente relatório CJUSBIA, com suas mais de 900 páginas, culminaram na proposição do Projeto de Lei 2338/2023, esse PL, se aprovado, será uma legislação que estabelecerá diretrizes abrangentes em âmbito nacional para o desenvolvimento, a implementação e o uso ético de sistemas de inteligência artificial (IA) no território brasileiro e ainda pretende assegurar a proteção dos direitos fundamentais e promover a adoção de sistemas de IA seguros e eficazes, visando o bem-estar da sociedade, a preservação dos valores democráticos, e o avanço da ciência e tecnologia.

Dessa forma, o PL 2338/2023 (Brasil, 2023), emerge como um marco na abordagem regulatória da IA no Brasil, adota uma metodologia inovadora, pautada em uma avaliação de risco adaptável e centrada no ser humano, reconhecendo as vulnerabilidades sociais e

promovendo uma supervisão democrática dos avanços tecnológicos. Ao alinhar-se com propostas transatlânticas e ao mesmo tempo resguardar as especificidades nacionais, o PL 2338/2023 não só evita o risco de uma “colonização regulatória”, mas também pavimentava o caminho para uma regulação que seja reflexo dos valores constitucionais brasileiros.

A pesquisa de doutorado que agora empreendo se situa na interseção dessas reflexões e legislações emergentes. Pretende-se explorar os desafios e perspectivas para uma sociedade em evolução, analisando como a regulação proposta pode promover uma harmonização entre o progresso tecnológico e os princípios éticos, democráticos e de direitos fundamentais. Assim, o estudo visa contribuir para o debate sobre como a IA pode ser regulada de maneira que fomenta inovações responsáveis e eticamente alinhadas com os ideais de uma sociedade justa e equitativa.

Pensando nisso, busco entender como a Inteligência Artificial está impactando os direitos fundamentais e os valores éticos na sociedade. O uso massivo de dados pessoais por sistemas de IA pode violar direitos como a privacidade e a autonomia individual, especialmente quando a coleta e o tratamento desses dados ocorrem sem consentimento ou supervisão adequada. Além disso, a IA pode perpetuar desigualdades e discriminações ao refletir preconceitos históricos nos algoritmos. A falta de transparência nos processos decisórios automatizados também compromete o direito à explicabilidade, reforçando a necessidade de regulamentações claras que protejam esses direitos. Do ponto de vista ético, é essencial que o desenvolvimento de IA seja orientado por valores como dignidade humana, justiça e diversidade, prevenindo o uso abusivo da tecnologia, como em casos de vigilância em massa.

Internacionalmente, a União Europeia se destaca ao implementar uma regulação rigorosa para sistemas de IA de alto risco, priorizando a transparência e a explicabilidade dos algoritmos. Nos Estados Unidos, a abordagem é mais flexível, permitindo que a inovação tecnológica avance antes de imposições regulatórias significativas, enquanto na China o controle estatal sobre a IA é mais centralizado. Cada modelo oferece lições valiosas: a União Europeia com seu foco em riscos, os Estados Unidos com sua flexibilidade, e a China com sua supervisão rígida.

Os desafios trazidos pela IA, especialmente no que diz respeito à tomada de decisões autônomas, refletem-se também na necessidade de maior clareza legal sobre a responsabilidade por danos. O Projeto de Lei 21/2020, por exemplo, foi criticado por adotar uma responsabilidade subjetiva, enquanto o PL 2338/2023 avança ao prever a responsabilidade objetiva para sistemas de alto risco.

A integração de princípios éticos no desenvolvimento de IA pode ser alcançada por meio de medidas como a transparência algorítmica, avaliação de impacto ético e supervisão humana constante. Essas práticas asseguram que os sistemas de IA respeitem os direitos humanos e promovam a justiça e a equidade, mitigando vieses e prevenindo discriminação.

Além disso, o impacto da IA no conceito de consentimento é preocupante, especialmente em cenários de vigilância invasiva. A capacidade de processamento de grandes volumes de dados pessoais dificulta a conscientização e o controle dos indivíduos sobre o uso de suas informações. Isso exige a adoção de práticas como a privacidade por design e a transparência algorítmica para proteger os direitos dos titulares.

No contexto de direitos autorais, a IA também apresenta desafios, uma vez que criações geradas por máquinas não são protegidas pela legislação brasileira atual, que reconhece apenas pessoas físicas como titulares de direitos autorais. Isso levanta questões sobre a inovação responsável e a proteção dos direitos dos criadores humanos.

No cenário de governança em rede, o envolvimento de múltiplos setores, como a academia e a sociedade civil, é essencial para garantir uma regulamentação inclusiva e democrática da IA. Esse modelo colaborativo e multissetorial permite que diferentes setores da sociedade participem das discussões e garantam que a regulação seja adaptável às rápidas inovações tecnológicas, promovendo transparência, responsabilidade social e a preservação dos direitos fundamentais.

Por fim, os principais desafios sobre as tendências emergentes na regulação da IA que envolvem uma abordagem baseada em riscos, e que ajusta o nível de intervenção conforme o grau de risco da tecnologia, com maior controle para sistemas de alto risco que afetam direitos fundamentais e mais flexibilidade para tecnologias de baixo risco, incluem a harmonização das regulamentações internacionais e a adaptação das políticas às rápidas mudanças tecnológicas, sem comprometer a proteção dos direitos fundamentais. A co-regulação e a autorregulação são recomendadas para promover inovação responsável e garantir segurança e direitos individuais.

O primeiro capítulo deste trabalho apresenta uma análise aprofundada sobre a evolução histórica da Inteligência Artificial (IA) e suas implicações sociais, políticas, econômicas e éticas. O ponto de partida é a discussão sobre como a era digital e o uso intensivo de dados transformaram os dados pessoais no "novo petróleo", um recurso valioso que alimenta as grandes empresas de tecnologia. Essa transformação alterou profundamente o comportamento humano, criando novas dinâmicas de poder e controle, como o conceito de "Capitalismo de Vigilância", introduzido por Shoshana Zuboff. A autora argumenta que as grandes corporações tecnológicas utilizam os dados pessoais para prever e modificar o comportamento humano,

transformando os indivíduos em alvos de exploração comercial. Ao mesmo tempo, o capítulo explora como Cathy O'Neil, em seu livro "Armas de Destruição Matemática", demonstra que os algoritmos, muitas vezes considerados neutros, podem perpetuar e até agravar desigualdades sociais, especialmente em áreas como o crédito, a educação e o sistema de justiça criminal.

Ainda no primeiro capítulo da tese, é feito um retorno às origens da IA mencionando as contribuições pioneiras de Alan Turing, que formulou o célebre "Teste de Turing", um marco inicial para determinar a capacidade de uma máquina em exibir comportamento inteligente indistinguível do humano. A partir dessa base, o texto traça a trajetória de desenvolvimento da IA ao longo das décadas, com ênfase nos avanços científicos e tecnológicos que culminaram nas redes neurais profundas, um dos principais pilares do aprendizado de máquina moderno. O capítulo menciona conquistas significativas, como o sistema AlphaGo, desenvolvido pela Google DeepMind, que superou jogadores profissionais no jogo de tabuleiro Go, demonstrando o poder da IA em áreas que antes eram consideradas de domínio exclusivamente humano, como o pensamento estratégico.

Além dos aspectos históricos e tecnológicos, o capítulo destaca os impactos sociais da IA abordando questões sensíveis como o viés algorítmico. Um exemplo importante discutido é o do sistema COMPAS, utilizado no sistema judiciário dos Estados Unidos para prever a reincidência criminal. Estudos demonstraram que o COMPAS apresentava vieses raciais, penalizando desproporcionalmente indivíduos negros, o que levantou preocupações éticas e sociais sobre o uso de IA em decisões que afetam diretamente a vida das pessoas. O capítulo alerta que, sem uma supervisão adequada, os algoritmos podem perpetuar preconceitos e reforçar estruturas de poder desiguais.

A tese argumenta que, à medida que a IA avança, a sociedade enfrenta o desafio de garantir que essa tecnologia seja desenvolvida e aplicada de maneira ética e responsável. Isso implica a necessidade de uma regulamentação robusta que leve em consideração os impactos sociais e os direitos fundamentais, promovendo uma governança tecnológica que assegure que a IA sirva ao bem comum. O capítulo ressalta que o Brasil, assim como outras nações, está em meio a debates legislativos e regulatórios sobre a IA, com projetos de lei e comissões de juristas propondo marcos legais que buscam equilibrar inovação tecnológica com a proteção dos direitos fundamentais.

A necessidade de regulamentar a IA é vista como uma resposta aos riscos emergentes que a tecnologia apresenta, não apenas no Brasil, mas globalmente. O capítulo discute como países ao redor do mundo, incluindo o Brasil, têm reconhecido a urgência de criar frameworks legais que regulem o uso da IA de forma a evitar abusos e proteger os cidadãos.

Por fim, o capítulo reflete sobre a necessidade de uma supervisão democrática e da participação ativa da sociedade civil na construção de uma regulamentação que não apenas promova o progresso tecnológico, mas que também resguarde os valores democráticos e os direitos fundamentais. A tese aponta que a regulação da IA deve ser adaptável e flexível para acompanhar o ritmo acelerado das mudanças tecnológicas, garantindo que a tecnologia seja usada de forma a beneficiar toda a sociedade, e não apenas interesses econômicos ou corporativos.

O capítulo 2 explora o impacto da Inteligência Artificial (IA) no mercado de trabalho, as implicações da desumanização causada pelo capitalismo de vigilância e a precarização das políticas públicas diante da automação.

Na seção sobre o impacto da IA no mercado de trabalho, autores como Martin Ford destacam que a revolução da IA difere das anteriores pela velocidade e profundidade com que elimina empregos, especialmente em setores que exigem tanto habilidades manuais quanto cognitivas. Embora novos empregos possam surgir, eles não compensarão os perdidos, aumentando a desigualdade econômica. Ford sugere medidas como o Rendimento Básico Universal (RBU) para lidar com essa nova realidade, defendendo políticas públicas que promovam inclusão social e adaptação ao futuro do trabalho.

A desumanização é discutida a partir da perspectiva do capitalismo de vigilância de Shoshana Zuboff, que descreve como a IA transforma comportamentos humanos em dados lucrativos, despersonalizando e controlando os indivíduos. O poder da IA em prever e modificar comportamentos prejudica a privacidade e a autonomia, transformando as pessoas em alvos passivos de manipulação. Zuboff argumenta que essa perda de autonomia ameaça a autodeterminação e liberdade individual, com graves implicações para a dignidade humana.

Na seção sobre a precarização das políticas públicas, Darrell M. West ressalta como a automação desafia os modelos tradicionais de seguridade social, uma vez que políticas baseadas no emprego formal se tornam inviáveis diante da eliminação de postos de trabalho. Ele propõe soluções como a RBU e a desvinculação de benefícios sociais do emprego. A automação também representa um risco fiscal para os governos, que enfrentam uma base tributária em declínio. Para mitigar esses impactos, West sugere tributar tecnologias automatizadas, enquanto Diogo Cortiz enfatiza a importância de programas de *upskilling* e *reskilling* para preparar a força de trabalho para as demandas da IA.

Em suma, o capítulo defende a necessidade de políticas públicas robustas para enfrentar as transformações causadas pela IA preservando a equidade social, a dignidade humana e a sustentabilidade econômica.

No capítulo 3, é abordado a regulação da Inteligência Artificial. No Brasil ela está avançando com a tramitação de dois projetos de lei fundamentais: o PL 21/2020 e o PL 2338/2023. O PL 21/2020, aprovado pela Câmara dos Deputados, foca em diretrizes gerais para o desenvolvimento de IA no país, promovendo a inovação tecnológica com uma abordagem de responsabilidade subjetiva, ou seja, a responsabilização de desenvolvedores e operadores dependerá da comprovação de culpa ou dolo. Essa proposta, embora incentive o desenvolvimento da IA, é criticada por não abordar suficientemente os riscos que essa tecnologia pode apresentar.

Por outro lado, o PL 2338/2023, em análise no Senado, adota uma abordagem mais moderna e alinhada às práticas internacionais, como o AI Act da União Europeia. O projeto classifica sistemas de IA com base no risco, impondo responsabilidade objetiva para sistemas de alto risco e proibindo tecnologias consideradas de risco excessivo, como aquelas que utilizam técnicas subliminares ou exploram vulnerabilidades de grupos vulneráveis. Essa categorização visa garantir maior segurança e proteção aos direitos fundamentais dos cidadãos.

Ambos os projetos dialogam com a Lei Geral de Proteção de Dados (LGPD), buscando integrar diretrizes para governança e transparência no uso da IA. No entanto, enquanto o PL 21/2020 se mantém mais genérico e conservador, o PL 2338/2023 propõe avaliações de impacto algorítmico e mecanismos de governança mais rigorosos para mitigar os riscos associados ao uso da IA. A adoção do PL 2338/2023 é vista como um avanço significativo para a criação de um marco regulatório que equilibre a inovação tecnológica com a proteção dos direitos fundamentais no Brasil.

O Capítulo 04 aborda o impacto da inteligência artificial (IA) generativa na criação de conteúdos e a problemática dos direitos autorais. IA, como os modelos de linguagem GPT e ferramentas como DALL·E, têm se tornado essenciais nas indústrias criativas, desafiando conceitos de autoria e originalidade. O capítulo explora como os *Large Language Models* (LLMs) funcionam, utilizando grandes volumes de dados para gerar textos e outras obras, e como prompts dados pelos usuários moldam o resultado. Surgem questões sobre se a IA é apenas uma ferramenta ou se suas produções têm autonomia criativa, colocando em debate a originalidade e o plágio. Exemplos práticos, como a obra premiada "Théâtre D'opéra Spatial", são analisados sob a ótica dos direitos autorais.

O texto também examina a evolução da IA generativa, destacando as Redes Gerativas Adversárias (GANs) e sua influência na geração autônoma de conteúdo. A discussão se aprofunda em questões jurídicas sobre a originalidade e plágio, enfatizando o conceito de

"contributo mínimo de originalidade", base da proteção dos direitos autorais. No contexto das tecnologias emergentes, a distinção entre criações genuínas e o uso de IA torna-se nebulosa.

Esse capítulo reflete sobre o conceito de criatividade, questionando se a IA pode ser vista como uma "musa" moderna, auxiliando na criação humana, e como a dependência tecnológica pode afetar o processo criativo. A análise sugere que, embora a IA possa ser uma ferramenta poderosa, ela não deve substituir a intencionalidade e a disciplina humana no processo criativo.

O Capítulo 05 trata do uso da Inteligência Artificial (IA) no Poder Judiciário, destacando os bons e maus usos dessas tecnologias. A jurisprudência é uma base sólida para decisões jurídicas, mas a confiança no sistema é abalada quando IA, como o ChatGPT, gera jurisprudências falsas ou distorce informações, o que levou a uma investigação pelo CNJ. A IA pode ser útil na análise de grandes volumes de dados, desde que com supervisão humana, pois decisões judiciais não podem ser baseadas em informações não validadas por pessoas.

Por outro lado, o uso responsável da IA pode trazer benefícios reais, como a realização de audiências virtuais que ampliam o acesso à justiça, especialmente em áreas remotas e no contexto da pandemia de COVID-19. Essa inovação acelera processos e não compromete a qualidade das decisões. Contudo, o uso excessivo e sem controle de IA representa riscos, como a criação de jurisprudência fictícia, levantando questões éticas e legais. O CNJ enfrenta o desafio de regulamentar o uso da IA no Judiciário, assegurando supervisão humana constante para evitar erros tecnológicos que comprometam a justiça.

Conclui-se que, embora a IA seja promissora, seu uso no Judiciário deve ser cauteloso e regulamentado para garantir que continue sendo uma ferramenta de apoio e não substitua a análise jurídica crítica, preservando a integridade das decisões.

Dessa forma, entende-se com o trabalho que embora a IA tenha demonstrado grande eficiência na análise de dados e reconhecimento de padrões, ela ainda não consegue replicar a criatividade humana, que surge de combinações inesperadas e intuição. A confiança cega nos algoritmos, como a criação de "bolhas de filtro" que limitam o acesso à diversidade de informações, pode levar a um determinismo informacional, restringindo a inovação e o pensamento crítico.

Esse fenômeno afeta não apenas a criatividade, mas também o tecido social, o capital social e a saúde mental, como observado nas consequências negativas da personalização algorítmica em jovens. A exploração econômica de dados e a manipulação de comportamentos pelos algoritmos, como no capitalismo de vigilância descrito por Zuboff, colocam em risco a autonomia dos consumidores e ampliam a desigualdade social.

No contexto do trabalho e do consumo, plataformas digitais como Uber e Amazon intensificam a precarização das relações laborais e a vigilância digital, exigindo regulamentações que protejam tanto trabalhadores quanto consumidores. A regulamentação da IA é urgente, com exemplos como a Europa e o Brasil avançando em legislações que buscam equilibrar inovação com a proteção dos direitos fundamentais.

Em suma, a IA deve ser utilizada de forma ética e responsável, com supervisão humana e regulação que acompanhe seu desenvolvimento rápido, para garantir que seus benefícios sejam amplamente colhidos sem comprometer liberdade, privacidade e equidade.

## OBJETIVOS E QUESTÕES DE PESQUISA

### **Objetivo Geral**

Investigar de forma interdisciplinar o impacto da Inteligência Artificial na sociedade contemporânea, com foco nas implicações sociais, éticas, legais e econômicas, visando contribuir para o desenvolvimento de uma regulação eficaz, ética e responsável que promova a inovação e proteja os direitos fundamentais.

### **Objetivos Principais**

- i. Analisar a Intersecção entre IA e Sociedade: Compreender como a Inteligência Artificial está moldando e sendo moldada pelos aspectos sociais, éticos, econômicos e legais da sociedade contemporânea.
- ii. Explorar a Necessidade de Regulação da IA: Identificar os desafios e as lacunas na legislação atual relacionados à IA e discutir as implicações de uma regulamentação inadequada ou inexistente.
- iii. Avaliar Modelos de Regulação Internacionais: Examinar e comparar diferentes abordagens regulatórias adotadas por países ao redor do mundo, destacando práticas eficazes e áreas de melhoria.
- iv. Propor Diretrizes para uma Regulação Ética e Responsável: Desenvolver recomendações para a criação de uma estrutura regulatória que promova o uso ético, transparente e responsável da IA.

- v. Contribuir para o Debate Acadêmico em CTS: Enriquecer a literatura no campo de Ciência, Tecnologia e Sociedade, fornecendo insights e análises baseadas na interdisciplinaridade da IA.

### Questões de Pesquisa

As questões que orientam esta pesquisa foram formuladas a partir do problema central: como a sociedade pode regular a Inteligência Artificial de forma a proteger direitos fundamentais sem sufocar a inovação? Desse problema central derivam cinco questões estruturantes:

- a. De que maneira a expansão da Inteligência Artificial impacta direitos fundamentais — como privacidade, autonomia, igualdade e dignidade humana — e quais os limites éticos que uma regulação responsável deve observar?
- b. Quais são as lacunas do marco regulatório brasileiro sobre IA, e em que medida o PL 2338/2023 representa um avanço em relação ao PL 21/2020 no que diz respeito à responsabilidade civil e à proteção de grupos vulneráveis?
- c. Como diferentes modelos internacionais de regulação da IA, em especial o *AI Act* europeu e as abordagens norte-americana e chinesa, se distinguem nos critérios de risco, transparência e responsabilidade, e quais lições podem orientar o debate regulatório brasileiro?
- d. Quais são os desafios jurídicos e éticos impostos pela IA generativa, particularmente no campo dos direitos autorais, da autoria e da criatividade humana, e como o direito brasileiro está preparado para enfrentá-los?
- e. De que forma a educação digital e a formação ética podem compor uma estratégia regulatória mais ampla, capaz de preparar indivíduos e instituições para os impactos da IA na sociedade?

### METODOLOGIA

A presente pesquisa adota abordagem qualitativa, entendida como a mais adequada para investigar fenômenos sociotécnicos complexos, nos quais as dimensões éticas, jurídicas e políticas não se prestam à quantificação, mas exigem interpretação contextualizada e análise crítica. A escolha metodológica é coerente com o campo de Ciência, Tecnologia e Sociedade (CTS), que historicamente opera na interseção entre saberes disciplinares distintos, recusando

reducionismos e privilegiando a compreensão das relações entre tecnologia e sociedade em sua concretude.

O percurso metodológico estruturou-se em dois movimentos complementares: o levantamento bibliográfico e a análise documental.

**Levantamento bibliográfico:** O levantamento bibliográfico orientou-se pela identificação de obras e autores que dialogam com três eixos temáticos centrais da pesquisa: os impactos sociais e éticos da Inteligência Artificial, os fundamentos teóricos da regulação tecnológica e os marcos conceituais sobre vigilância, controle e direitos fundamentais na era digital. A seleção priorizou literatura acadêmica consolidada, como as obras de Shoshana Zuboff sobre capitalismo de vigilância, Cathy O'Neil sobre vieses algorítmicos e Manuel Castells sobre sociedade em rede, bem como produções mais recentes que refletem o estado atual do debate regulatório, especialmente no campo do direito e das políticas públicas de IA.

O critério de inclusão foi a relevância direta para os objetivos da pesquisa: foram privilegiadas obras que permitissem compreender, de um lado, os mecanismos pelos quais a IA opera sobre indivíduos, relações de trabalho e estruturas de poder; e, de outro, as respostas normativas e éticas que diferentes sociedades têm formulado para lidar com esses impactos. Obras de caráter estritamente técnico, sem interface com as dimensões jurídicas ou sociais, foram deliberadamente excluídas, em coerência com o recorte do campo CTS.

**Análise documental:** A análise documental centrou-se em legislações, projetos de lei, relatórios institucionais e diretrizes regulatórias relacionados à governança da IA. No âmbito nacional, foram examinados o PL 21/2020, o PL 2338/2023 e o relatório da Comissão de Juristas (CJUSBIA), além da Lei Geral de Proteção de Dados (LGPD) como marco normativo de referência. No âmbito internacional, foram analisados o AI Act da União Europeia, diretrizes do governo norte-americano sobre IA e documentos institucionais da UNESCO sobre ética e inteligência artificial.

A leitura desses documentos foi orientada por três critérios analíticos temáticos: (i) a abordagem de risco adotada por cada modelo regulatório, isto é, como cada sistema jurídico identifica, classifica e responde aos riscos representados pelos sistemas de IA; (ii) os mecanismos de transparência e explicabilidade exigidos dos desenvolvedores e operadores; e (iii) o modelo de responsabilidade civil adotado, se subjetiva ou objetiva, e suas implicações para a proteção de direitos fundamentais. Esses três eixos permitiram uma comparação analítica consistente entre os modelos brasileiro, europeu e norte-americano, evidenciando convergências, divergências e lacunas regulatórias.

A opção por não adotar um método único e rígido de análise, como a análise de conteúdo formalizada ou a teoria fundamentada, reflete a natureza exploratória e interdisciplinar da pesquisa. O campo da regulação da IA é ainda incipiente no Brasil, e a ausência de um marco regulatório consolidado à época da pesquisa demandava uma postura metodológica aberta, capaz de acompanhar debates em curso e incorporar documentos produzidos em contexto de intensa movimentação legislativa. Trata-se, portanto, de uma pesquisa qualitativa de caráter exploratório e analítico, ancorada na leitura crítica da literatura e dos documentos selecionados à luz dos objetivos propostos.

## IMPORTÂNCIA DA PESQUISA PARA O CAMPO CTS

Nesta pesquisa, busco proporcionar uma contribuição significativa para o campo de Ciência, Tecnologia e Sociedade (CTS) por meio de várias vertentes. A própria interdisciplinaridade da IA se encaixa perfeitamente na natureza multifacetada do CTS. Em vez de tratar a IA como uma mera evolução tecnológica, esta pesquisa a analisa como um fenômeno sociotécnico, onde as interações entre ciência, tecnologia e sociedade são intrínsecas e mutuamente influentes.

A relevância da IA torna o tema da sua regulação essencial para a atualidade. Ao abordar este tema, a pesquisa busca elucidar os desafios emergentes e as oportunidades que caracterizam o cenário contemporâneo no campo CTS. Os dilemas éticos da IA são representativos dos desafios contemporâneos da era digital. CTS, como campo de estudo, é fundamental para decifrar e propor soluções para tais questões.

A governança democrática enfatiza a primazia de um processo democrático e inclusivo na tomada de decisões tecnológicas. Isso ressoa com os princípios centrais de CTS, que advogam pela democratização do conhecimento e pela co-criação responsável de tecnologia.

A educação e conscientização sobre IA visa elevar o nível de compreensão e sensibilização acerca da IA, o que é imprescindível em CTS. A discussão sobre justiça e equidade na IA posiciona questões de justiça e equidade no epicentro da discussão, em consonância com os objetivos centrais do campo CTS.

Dessa forma, a pesquisa se propõe a criar um diálogo entre a rápida evolução da IA e a necessidade de uma regulamentação adaptativa e consciente, visando garantir um desenvolvimento tecnológico que seja ético, justo e benéfico para toda a sociedade. Esta abordagem não apenas responde às necessidades imediatas trazidas pela IA, mas também

antecipa os desafios futuros, contribuindo de forma significativa para o campo de Ciência, Tecnologia e Sociedade.

## MOTIVAÇÃO DO NOME DO TRABALHO

A escolha do termo "mutação" inserido no título desta tese não é meramente casual. Em genética, mutação refere-se a uma mudança inesperada no genótipo, que pode ser transmitida às gerações futuras, mas cujos efeitos são imprevisíveis, podendo ser positivos, neutros ou prejudiciais. Essa noção de transformação súbita e não direcionada aplica-se à maneira como a Inteligência Artificial (IA) está impactando a sociedade contemporânea. Diferentemente do conceito de “evolução”, que, embora possa gerar consequências adversas, é amplamente visto como um progresso natural, o termo "mutação" carrega a ideia de imprevisibilidade e flexibilidade. As mudanças trazidas pela IA podem ser benéficas, como a automatização de processos e a inovação em setores chave, mas também podem gerar impactos negativos, como a substituição de empregos e o aumento das desigualdades sociais.

Essa comparação com o fenômeno genético destaca a dualidade presente na incorporação da IA à sociedade: a incerteza de seus impactos a longo prazo. A IA já se inseriu em nossas vidas cotidianas de maneira semelhante à eletricidade. Como destacou Andrew Ng em 2017, a IA é a "nova eletricidade", transformadora e essencial para o funcionamento de diversos setores industriais e sociais (Lee, 2018). Assim como a eletricidade, que hoje é tão essencial que só nos damos conta de sua importância quando ela falta, a IA está gradualmente se tornando uma extensão de nós mesmos. No entanto, os impactos de sua proliferação, sejam benéficos ou prejudiciais, ainda não podem ser totalmente mensurados.

A "mutação" provocada pela IA, assim como na biologia, pode gerar efeitos que desafiarão a sociedade em suas diversas esferas, desde a economia até a ética, passando pela legislação. Por isso, faz-se necessária uma regulação eficaz e responsável, que consiga equilibrar a promoção de inovações tecnológicas e a proteção dos direitos individuais e sociais.

O acadêmico francês Michel Desmurget (2022), no seu livro “A Fábrica de Cretinos Digitais”, apresenta sua vasta pesquisa onde pode constatar que pela primeira vez os filhos estão apresentando um QI inferior ao de seus pais. Desse modo, sugere que tecnologia, em vez de promover um avanço contínuo, pode estar distorcendo ou até mesmo prejudicando o desenvolvimento cognitivo das novas gerações (Desmurget, 2022).

Essa "mutação artificial", impulsionada pelo uso intensivo de tecnologias de IA, pode estar interferindo no processo natural de evolução cognitiva. Ao delegar muitas funções

cognitivas à IA e consumir conteúdo gerado e distribuído por algoritmos de redes sociais, há uma diminuição da necessidade de esforço intelectual, o que pode levar a consequências indesejadas. O impacto é mais profundo do que apenas a interação tecnológica; ele pode estar moldando o desenvolvimento das capacidades intelectuais e sociais das futuras gerações de maneiras que ainda não podemos prever.

A imprevisibilidade e a multidimensionalidade dos impactos da IA que este trabalho se propõe a examinar encontram um referencial teórico preciso no pensamento complexo de Edgar Morin. Para Morin (2015), a complexidade não é um defeito do conhecimento, é a condição real dos fenômenos que a ciência tende a simplificar artificialmente ao fragmentá-los em partes isoladas. Um sistema complexo, como a IA inserida na sociedade, não pode ser compreendido a partir de apenas uma de suas dimensões, técnica, jurídica, econômica ou ética, porque seus efeitos emergem justamente das interações entre essas dimensões, de forma não-linear e frequentemente imprevisível. Morin chama de "pensamento simplificador" aquele que separa o que está unido, que isola o objeto do seu contexto, que reduz o complexo ao simples. Regulamentar a IA com esse tipo de pensamento, acreditando que basta uma lei, um critério técnico ou um princípio ético, é subestimar a natureza do fenômeno. O que a tese aqui propõe, em consonância com Morin, é um olhar que reconheça a IA como um fenômeno sociotécnico complexo: ela não funciona de forma isolada, não produz efeitos previsíveis e não pode ser governada por instrumentos que ignorem suas interdependências com o direito, a economia, a educação, a cultura e o poder.

## CAPÍTULO 1 - A R-EVOLUÇÃO DA IA: REFLEXÕES SOCIAIS, POLÍTICAS E ECONÔMICAS

### 1.1 PANORAMA HISTÓRICO DA INTELIGÊNCIA ARTIFICIAL

#### 1.1.1 A Nova Era

A emergência da Inteligência Artificial (IA) como força dominante na sociedade moderna suscita importantes reflexões sobre os direitos fundamentais, tais como democracia, igualdade, liberdade, devido processo legal, privacidade e proteção de dados. A extensão e complexidade do impacto dessa tecnologia na sociedade suscita uma série de questões, sejam elas positivas quanto negativas, irão e já estão transformado a coletividade.

Pode-se entender que a atualidade é uma era definida pelo poder dos dados, onde a quantidade de informações geradas pela sociedade ultrapassa qualquer precedente histórico, assim como a capacidade de estruturar e organizar essa informação e retirar valor desse conteúdo. Como observado por Zuboff (2021) no contexto do "Capitalismo de Vigilância", a acumulação e análise de dados em grande escala oferece oportunidades sem precedentes para insights e inovações, mas também levanta questões profundas sobre privacidade e consentimento. A capacidade de prever eventos futuros a partir desses vastos conjuntos de dados, como discutido por O'Neil (2020) em "Algoritmos de Destruição em Massa: Como o *Big Data* Aumenta a Desigualdade e Ameaça a Democracia", destaca a influência crescente das "personas digitais" sobre aspectos fundamentais da existência humana, desde a liberação de crédito até avaliações de periculosidade por algoritmos.

A capa da revista The Economist foi publicada em 2017, com o título "The World's Most Valuable Resource" (O Recurso Mais Valioso do Mundo) e tratava da ideia de que os dados são o novo petróleo. A ilustração mostra plataformas tecnológicas extraíndo dados, como se fossem plataformas de petróleo, reforçando a noção de que, na era digital, os dados são uma matéria-prima tão ou mais valiosa quanto o petróleo (The Economist, 2017).

Figura 1 – Capa da Revista The Economist de 2017



Fonte: The Economist (2017)

Nesta edição, a revista representa as gigantes da tecnologia como plataformas de petróleo, simbolizando a extração de uma matéria-prima valiosa — neste caso, os dados. Assim como as plataformas extraem petróleo das profundezas da Terra para convertê-lo em produtos essenciais para a economia global, as grandes empresas de tecnologia capturam e transformam os dados em *commodities* altamente valiosas. Essa capa marca o início de uma nova era, onde o poder dos dados supera o do petróleo como o recurso mais valioso do mundo. A observação é clara: basta olhar para as empresas mais valiosas da atualidade, que utilizam dados como a principal fonte de valor e poder econômico, consolidando sua posição no topo do mercado global.

A mensagem central que se pretende destacar é a seguinte: a Inteligência Artificial (IA) opera de maneira eficaz por ser treinada em um vasto conjunto de informações, sendo dependente dos dados para produzir resultados de qualidade. Esse aspecto é de extrema

relevância, pois posiciona os dados como uma *commodity* de valor inigualável no cenário contemporâneo. Ao analisar as questões levantadas por Shoshana Zuboff, que serão amplamente exploradas ao longo deste trabalho, surge uma questão fundamental: quais são as implicações desse valor atribuído aos dados no mercado?

A questão central que se coloca é se o uso massivo de dados beneficiará efetivamente a sociedade como um todo ou se servirá predominantemente para o enriquecimento das grandes corporações tecnológicas. A grande preocupação reside no fato de que essa *commodity* valiosa (os dados) encontra-se concentrada nas mãos de um número restrito de empresas, gerando uma acumulação excessiva de riqueza e poder em detrimento de impactos sociais e individuais ainda incertos. Trata-se de uma reflexão essencial ao se debater o papel da IA e dos dados no futuro da sociedade e no equilíbrio entre inovação tecnológica e equidade social.

Em setembro de 2024, uma nova reportagem reforça essa tendência (Lake, 2024), destacando a entrada de Mark Zuckerberg, CEO da Meta, no seletivo clube dos 200 bilhões de dólares, ao lado de Elon Musk, da Tesla, e Jeff Bezos, da Amazon. A exclusividade desse grupo ressalta como as maiores fortunas do mundo estão, mais do que nunca, ligadas às empresas de tecnologia, que têm os dados como seu recurso mais valioso. A ascensão desses gigantes demonstra a importância estratégica dos dados na criação de riqueza e poder, tornando-os tão ou mais influentes do que as tradicionais indústrias de *commodities*. Assim como o petróleo transformou a economia global no século XX, os dados têm sido o motor por trás das maiores inovações, aquisições e transformações do século XXI, moldando um novo cenário econômico dominado, principalmente, por esses três líderes que representam a convergência entre tecnologia e poder financeiro.

É importante, contudo, distinguir analiticamente as três dimensões em que os dados operam ao longo desta discussão e ao longo de toda a tese, pois embora se relacionem, não são equivalentes. Na primeira dimensão, os dados funcionam como recurso econômico: são matéria-prima extraída, processada e monetizada pelas grandes corporações tecnológicas, gerando concentração de riqueza e poder, como ilustrado pela capa da *The Economist* e pela ascensão dos bilionários da tecnologia. Na segunda dimensão, os dados são infraestrutura técnica da IA: sem grandes volumes de dados de qualidade, os modelos de aprendizado de máquina não funcionam, os dados, aqui, são o combustível que viabiliza o treinamento e o desempenho dos sistemas inteligentes. Na terceira dimensão, os dados constituem objeto de proteção jurídica: são informações pessoais vinculadas a indivíduos concretos, cujo tratamento inadequado viola direitos fundamentais como privacidade, autonomia e dignidade, razão pela qual legislações como a LGPD e o GDPR os submetem a regimes específicos de governança e

responsabilidade. Mobilizar essas três dimensões é necessário para compreender a complexidade do fenômeno, mas elas precisam ser distinguidas para que a análise não dissolva o específico no geral. Ao longo desta tese, quando se fala em "dados", o contexto indicará qual dessas dimensões está em foco.

Os benefícios da IA são inegáveis, abrangendo desde avanços em diagnósticos médicos até melhorias na sustentabilidade ambiental. Tais inovações, têm o potencial de melhorar consideravelmente a qualidade de vida. Contudo, esses sistemas também apresentam riscos imensos. A manipulação inadequada ou a interpretação errônea dos dados pode levar a correlações injustas ou ilegais, impactando negativamente a vida dos indivíduos. (Brevini; Pasquale, 2020).

A questão da governança da IA e a proteção de dados sensíveis tornam-se centrais nesse debate. Como Floridi (2021) sugere, a ética da informação oferece um quadro para navegar nesses dilemas, promovendo uma abordagem que respeita tanto os benefícios quanto os riscos associados à IA. A implementação de regulamentações, como o Regulamento Geral de Proteção de Dados (GDPR) da União Europeia, aponta para esforços globais no sentido de assegurar que a inovação tecnológica não comprometa os direitos fundamentais.

Em suma, a IA representa uma faca de dois gumes, capaz de trazer tanto avanços extraordinários quanto desafios éticos e legais significativos. A chave para maximizar seus benefícios enquanto minimiza os riscos reside na implementação de uma governança eficaz e em políticas de proteção de dados robustas. Assim, as sociedades modernas devem buscar um equilíbrio entre a exploração do potencial da IA e a salvaguarda dos direitos fundamentais, garantindo que a tecnologia sirva ao bem comum.

### **1.1.2 Conceito e Evolução de IA**

A Inteligência Artificial (IA) pode ser definida como a capacidade de máquinas ou sistemas computacionais de realizar tarefas que normalmente requerem inteligência humana. Essas tarefas incluem, mas não se limitam a raciocínio, aprendizado, reconhecimento de padrões, compreensão de linguagem natural, e tomada de decisões. A IA busca criar sistemas que possam simular, replicar ou até mesmo superar a capacidade humana em várias atividades, operando com autonomia e eficiência.

O conceito de IA teve suas raízes plantadas na metade do século XX, quando a sua compreensão passou de uma ideia teórica para um campo de estudo prático, que mais tarde se transformaria em uma das maiores revoluções tecnológicas da era moderna. Um dos primeiros

marcos na história da IA foi o artigo de Alan Turing, "*Computing Machinery and Intelligence*", publicado em 1950. Neste trabalho, Turing propôs a famosa pergunta: "As máquinas podem pensar?" e introduziu o que hoje é conhecido como o Teste de Turing, um experimento projetado para avaliar a capacidade de uma máquina de exibir comportamento inteligente indistinguível de um humano (Turing, 1950).

Nos anos seguintes, o campo de IA começou a ganhar forma, com a conferência de Dartmouth em 1956 sendo amplamente reconhecida como o nascimento formal do campo. Organizada por John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, a conferência reuniu alguns dos maiores pensadores da época para discutir como as máquinas poderiam ser programadas para simular a inteligência humana (McCarthy, *et al.*, 2006). Este evento não apenas consolidou o termo "Inteligência Artificial", mas também estabeleceu uma agenda de pesquisa que guiaria o campo nas próximas décadas.

O desenvolvimento da IA foi marcado por um período de grande otimismo e inovação, conhecido como a era pioneira da IA. Entre os primeiros sistemas desenvolvidos, destaca-se o *Logic Theorist*, criado por Allen Newell e Herbert A. Simon em 1956. Este sistema é amplamente considerado o primeiro programa de IA, capaz de provar teoremas de lógica matemática de maneira similar ao raciocínio humano (Russell; Norvig, 2013). O *Logic Theorist* foi projetado para imitar o processo de pensamento humano ao resolver problemas matemáticos, estabelecendo as bases para futuros desenvolvimentos em IA.

Esse período de otimismo foi também caracterizado pelo interesse militar na IA. Com a Guerra Fria em pleno andamento, tanto os Estados Unidos quanto a União Soviética investiram em pesquisas que poderiam fornecer uma vantagem estratégica, incluindo o desenvolvimento de sistemas de IA para aplicações militares.

A *Advanced Research Projects Agency Network* (ARPANET) foi uma rede de computadores pioneira, criada em 1969, como parte de um projeto financiado pela *Advanced Research Projects Agency* (ARPA), que atualmente é conhecida como *Defense Advanced Research Projects Agency* (DARPA). O objetivo original da ARPANET era interligar instituições acadêmicas e militares nos Estados Unidos, permitindo a troca segura e eficiente de informações, especialmente em um contexto de Guerra Fria, onde a comunicação robusta e resiliente era essencial. (ARPANET, 2024)

Embora a ARPANET tenha sido desenvolvida por pesquisadores e cientistas que trabalhavam sob o patrocínio da ARPA, ela se baseou em diversas inovações tecnológicas, incluindo a comutação de pacotes, que foi inspirada pelos trabalhos de Paul Baran, Donald Davies e outros. Esses avanços permitiram que a ARPANET se tornasse a primeira rede a

implementar protocolos de comunicação que mais tarde evoluíram para o TCP/IP (*Transmission Control Protocol/Internet Protocol*), cujo conceito é entendido como conjunto de protocolos de comunicação que permite a interconexão de diferentes redes de computadores; que formou o que hoje conhecemos como a internet (ARPANET, 2024).

Portanto, embora a DARPA tenha financiado e supervisionado o desenvolvimento da ARPANET, o projeto foi um esforço colaborativo que envolveu diversas contribuições de cientistas e engenheiros de diferentes instituições, resultando em uma das inovações tecnológicas mais significativas do século XX. A ARPANET é considerada o ponto de partida para o desenvolvimento da internet, conectando inicialmente apenas algumas universidades, mas que, ao longo do tempo, cresceu para se tornar a rede global que conhecemos hoje.

Como visto, as décadas de 1960 e 1970 foram marcadas por grandes avanços teóricos, mas também por expectativas inflacionadas. Projetos como o *General Problem Solver* (GPS) de Allen Newell e Herbert Simon exemplificaram as aspirações da época ao tentar criar um sistema que pudesse resolver problemas de maneira geral, semelhante à mente humana (Newell; Simon, 1961). No entanto, as limitações tecnológicas e a complexidade dos problemas logo levariam ao primeiro "inverno da IA", um período de frustração e redução de financiamentos, à medida que os resultados práticos falhavam em acompanhar o otimismo inicial.

Na década de 1980, a IA ressurgiu com o advento dos sistemas especialistas, que eram projetados para realizar tarefas específicas em áreas como diagnóstico médico e prospecção geológica. Esses sistemas representavam um compromisso entre a ambição original de criar uma inteligência geral e a realidade das limitações tecnológicas da época (Feigenbaum, 1984). Entretanto, esse ressurgimento também foi seguido por outro período de decepção, o que levou a um segundo inverno da IA no final dos anos 1980 e início dos anos 1990.

A verdadeira revolução da IA começou a tomar forma no início do século XXI, impulsionada por avanços em poder computacional, o surgimento de grandes volumes de dados (*Big Data*) e o desenvolvimento de novos algoritmos de aprendizado de máquina. Em particular, o aprendizado profundo (*deep learning*), uma subárea das redes neurais artificiais, emergiu como uma ferramenta poderosa para resolver problemas complexos. Um dos marcos dessa nova era foi o desenvolvimento das redes neurais convolucionais (*Convolutional Neural Networks* - CNNs), que revolucionaram o campo de visão computacional (Lecun; Bengio; Hinton, 1998).

As CNNs se destacaram pela sua capacidade de processar e interpretar grandes quantidades de dados visuais, como imagens e vídeos, de forma eficiente. O trabalho de LeCun, em colaboração com Yoshua Bengio e Geoffrey Hinton, estabeleceu as bases para o

aprendizado profundo, que se tornou a espinha dorsal de muitas das tecnologias de IA modernas (Lecun; Bengio; Hinton, 1998). As CNNs revolucionaram áreas como a visão computacional, permitindo que máquinas superassem a capacidade humana em tarefas de reconhecimento de imagens, como demonstrado pela vitória do sistema AlexNet no *ImageNet Challenge* de 2012 (Krizhevsky; Sutskever; Hinton, 2012).

AlexNet é uma rede neural convolucional (CNN), que revolucionou o campo da visão computacional. Ela se destacou ao vencer o *ImageNet Challenge* em 2012, uma competição de reconhecimento de imagens. AlexNet foi a primeira a usar uma arquitetura de rede profunda com várias camadas convolucionais, que permitiu uma precisão muito superior em tarefas de classificação de imagens. Isso marcou o início da era moderna do aprendizado profundo em visão computacional (Imagenet..., 2024).

Outro grande avanço da década de 2010 foi o desenvolvimento de sistemas de IA que podiam aprender com grandes volumes de dados e melhorar continuamente com a experiência. Essa abordagem, conhecida como aprendizado profundo, permitiu a criação de modelos complexos com múltiplas camadas de processamento, que podiam extrair características relevantes de dados brutos. Um exemplo notável, é o *AlphaGo*, desenvolvido pela *Google DeepMind*, que em 2016 derrotou o campeão mundial de Go, *Lee Sedol*, em uma série de partidas (Silver, *et al.*, 2016). O *AlphaGo* utilizou uma combinação de aprendizado profundo e aprendizado por reforço para dominar o jogo, demonstrando a capacidade das redes neurais de superar até mesmo os melhores jogadores humanos em um jogo considerado extremamente complexo devido ao seu vasto espaço de estados possíveis.

Com essa tecnologia, a IA começou a fazer incursões importantes em áreas práticas, desde o diagnóstico médico assistido por computador até veículos autônomos. Um dos momentos mais icônicos dessa nova era da IA foi a vitória do *AlphaGo*, um programa desenvolvido pelo *Google DeepMind*, contra o campeão mundial de Go, *Lee Sedol*, em 2016. O Go, um jogo de estratégia extremamente complexo, com mais possibilidades de movimentos do que o número de átomos no universo, era considerado um desafio inalcançável para sistemas de IA, dada a vastidão do espaço de estados possíveis (Abeliuk; Gutiérrez, 2021).

O *AlphaGo* utilizou uma combinação de aprendizado profundo e técnicas de aprendizado por reforço para atingir um nível de competência que superou até mesmo os melhores jogadores humanos. Esse marco na história da IA não foi apenas uma demonstração de poder computacional, mas também de um novo paradigma no desenvolvimento de sistemas inteligentes, onde a máquina aprende e evolui a partir de milhões de exemplos e simulações,

desenvolvendo estratégias que eram, até então, desconhecidas para os seres humanos (Silver *et al.*, 2016).

Essa abordagem contrasta com o caso do *Deep Blue*, o supercomputador desenvolvido pela IBM, que em 1997 derrotou o campeão mundial de xadrez Garry Kasparov. O *Deep Blue* utilizava uma técnica de força bruta, explorando seu vasto poder computacional para calcular milhões de posições de jogo por segundo. A máquina baseava-se em uma combinação de algoritmos de busca de profundidade e uma base de dados de jogadas de xadrez, permitindo-lhe avaliar possíveis movimentos e escolher aqueles que maximizariam suas chances de vitória (Verle, 1998). Embora essa vitória tenha sido um marco significativo, evidenciando a capacidade de máquinas superarem humanos em tarefas definidas e estruturadas, o *AlphaGo* representou um avanço em um campo mais complexo, onde a criatividade e a adaptação a novas estratégias foram essenciais para o sucesso.

Esse avanço exemplifica a transição da IA de uma ferramenta teórica e experimental para uma realidade prática com aplicações vastas e variadas. As técnicas usadas pelo *AlphaGo* rapidamente encontraram aplicações em outras áreas, como o diagnóstico médico, onde modelos de IA começaram a ser usados para analisar imagens médicas com uma precisão muitas vezes superior à dos médicos humanos. Além disso, o aprendizado profundo começou a revolucionar o campo da condução autônoma, onde veículos equipados com redes neurais podem interpretar dados de sensores em tempo real para tomar decisões críticas na estrada (Litjens *et al.*, 2017).

Essa nova era da IA é caracterizada pela capacidade de as máquinas realizarem tarefas que antes eram consideradas exclusivamente humanas, não apenas imitando, mas em muitos casos superando as habilidades humanas. A vitória do *AlphaGo* é um exemplo marcante dessa transformação, onde a IA não só jogou Go em um nível profissional, mas também criou novas estratégias de jogo, mudando para sempre a compreensão humana do jogo. Esse exemplo ilustra como a IA pode expandir os limites do que é possível em termos de "criatividade" e inovação (Silver, *et al.*, 2016).

O sucesso do *AlphaGo* e de outras tecnologias de aprendizado profundo gerou um aumento no interesse e nos investimentos em IA, tanto na academia quanto na indústria. As aplicações de aprendizado profundo se expandiram para áreas como processamento de linguagem natural, robótica, e até mesmo medicina, onde modelos de IA são usados para diagnosticar doenças a partir de imagens médicas e prever resultados de tratamentos (Litjens, *et al.*, 2017). Essa revolução tecnológica foi facilitada por avanços em *hardware*, especialmente como o já referido uso de GPUs (Unidades de Processamento Gráfico) para treinamento de

redes neurais, permitindo que os modelos fossem treinados em tempos muito mais curtos e com maior precisão.

O impacto das redes neurais e do aprendizado profundo continua a se expandir, impulsionando inovações em áreas como a condução autônoma. Empresas como Tesla e Waymo utilizam essas tecnologias para processar dados de sensores e tomar decisões em tempo real, o que demonstra a capacidade das redes neurais de transformar indústrias inteiras. A Tesla, por exemplo, vai além de ser uma fabricante de veículos elétricos; ela se destaca como uma das maiores empresas de dados do mundo. Seus veículos coletam continuamente enormes quantidades de dados, que são analisados e utilizados para aprimorar a tecnologia de condução autônoma, proporcionando uma experiência de direção cada vez mais segura e eficiente (Por que a Tesla..., 2024).

A Tesla integra esses dados em seus algoritmos de aprendizado de máquina, permitindo que seus veículos "aprendam" com a vasta quantidade de informações coletadas. Isso não só melhora a precisão das decisões tomadas pelos sistemas autônomos, mas também permite uma atualização contínua e em tempo real das capacidades dos veículos. A capacidade de processar e interpretar esses dados rapidamente é um dos fatores que coloca a Tesla à frente no desenvolvimento de tecnologias autônomas.

Além disso, o modelo de negócios da Tesla demonstra como o domínio dos dados pode ser uma vantagem competitiva crucial no setor automotivo. Ao se posicionar como uma empresa de tecnologia de dados, a Tesla não apenas fabrica carros, mas também desenvolve um ecossistema de informações que alimenta a inovação constante e a melhoria de seus produtos. Esse enfoque na coleta e análise de dados reflete uma tendência mais ampla na indústria, onde a integração de IA e aprendizado profundo está redefinindo os limites do que é possível na condução autônoma (Por que a Tesla..., 2024).

Essa evolução tem sido impulsionada por avanços tecnológicos em poder de processamento e pela disponibilidade de grandes volumes de dados, que alimentam os algoritmos de aprendizado profundo. O uso de GPUs (*Graphics Processing Unit*), por exemplo, permitiu que redes neurais complexas fossem treinadas em tempos mais curtos, aumentando a eficiência e a precisão dos modelos resultantes. Essa infraestrutura tecnológica avançada é o que possibilita a criação de sistemas como o *AlphaGo*, que requerem enorme poder computacional para processar as inúmeras possibilidades em jogos como o Go (Litjens, *et al.*, 2017).

As GPUs ou Unidade de Processamento Gráfico (*Graphics Processing Unit*), para um melhor entendimento, inicialmente foram criadas para controlar a exibição de imagens em telas,

evoluíram significativamente desde sua origem nos controladores gráficos das décadas de 1940 e 1950. Originalmente não programáveis e dependentes das CPUs, as GPUs ganharam autonomia e poder de processamento com o tempo. Nos anos 1990, a Nvidia lançou a primeira GPU de chip único, a *GeForce 256*, integrando capacidades de renderização e iluminação. Em 2007, o lançamento do CUDA pela Nvidia expandiu o uso das GPUs para tarefas paralelas intensivas, como inteligência artificial e machine learning, tornando-as essenciais em aplicações emergentes, incluindo blockchain, desse modo tornaram-se essenciais no campo do aprendizado profundo devido à sua arquitetura que permite o processamento massivamente paralelo, ideal para os cálculos em larga escala exigidos no treinamento de redes neurais profundas. (Amazon Web Services, 2024).

Esse avanço não só solidificou o papel das GPUs em tecnologias emergentes, mas também levou a Nvidia a atingir um valor de mercado sem precedentes, evidenciando o "verão superaquecido" da IA em contraste com os períodos anteriores de invernos tecnológicos. Em maio de 2023, a Nvidia tornou-se uma das empresas de maior valor de mercado no mundo, superando US\$ 1 trilhão, impulsionada pelo crescente interesse e investimentos em IA, que contrastam fortemente com as incertezas e frustrações do passado. (Patnaik; Reinicke, 2023)

Em resumo, nas duas últimas décadas foram marcadas como um ponto de virada na evolução da IA, com sistemas como o *AlphaGo* demonstrando que as máquinas podem não apenas replicar, mas também exceder as capacidades humanas em tarefas complexas. Essa transformação, impulsionada pelo aprendizado profundo (*deeplearning*), abriu novas possibilidades para a aplicação da IA em áreas práticas, desde a saúde até o transporte, que é caracterizado por grandes avanços na capacidade de processamento de dados e no desenvolvimento de algoritmos que permitiram a criação de modelos de IA altamente eficientes e precisos. As redes neurais, inspiradas no funcionamento do cérebro humano, passaram a ser aplicadas em uma variedade de tarefas complexas, desde o reconhecimento de imagens até a compreensão de linguagem natural, impulsionando uma revolução tecnológica sem precedentes.

Hoje, a IA é uma realidade prática com implicações profundas em diversas áreas da sociedade. Setores como saúde, educação, segurança e transporte estão sendo transformados pela IA trazendo benefícios significativos, mas também levantando questões éticas e sociais complexas. O impacto da IA no mercado de trabalho, o viés algorítmico e as questões de privacidade e segurança são apenas alguns dos desafios que surgem à medida que a IA se torna cada vez mais integrada em nossas vidas (Zuboff, 2021).

A evolução da IA, de uma teoria abstrata para uma ferramenta indispensável no mundo moderno, destaca tanto as oportunidades quanto os desafios que a tecnologia traz. O campo continua a se expandir, e as próximas décadas prometem ainda mais avanços, com implicações que ainda estão sendo descobertas e debatidas.

## 1.2 IMPACTOS SOCIAIS, ECONÔMICOS, POLÍTICOS E ÉTICOS DA IA

Ao refletir sobre os impactos da Inteligência Artificial (IA), é fundamental compreender que sua influência não ocorre de forma isolada, mas dentro de uma estrutura social e comunicacional mais ampla. Manuel Castells (2018) define a era contemporânea como uma “sociedade em rede”, na qual os fluxos de informação constituem a base material das relações sociais, econômicas e políticas. Nessa nova morfologia social, o poder e a produção de sentido não se concentram mais em instituições hierárquicas, mas circulam através de redes digitais interconectadas, moldando comportamentos, decisões e percepções coletivas.

Essa visão é aprofundada em *O Poder da Comunicação* (Castells, 2017), onde o autor argumenta que o poder na sociedade informacional depende da capacidade de moldar mentes por meio da comunicação. No contexto atual, a inteligência artificial se apresenta como a nova infraestrutura dessa comunicação, mediando as relações humanas através de algoritmos que selecionam, classificam e priorizam conteúdos. A IA, ao operar dentro dessa lógica de rede, torna-se um ator central na produção de significados sociais, capaz de amplificar tanto a cooperação quanto o controle, a inclusão quanto a exclusão.

A intersecção entre IA e comunicação, portanto, redefine o próprio exercício do poder. Como observa Castells (2017), quem controla os fluxos informacionais controla a realidade simbólica da sociedade. Essa constatação é essencial para entender os impactos da IA, pois os algoritmos contemporâneos não apenas processam dados, eles moldam a experiência social, as relações políticas e a percepção ética. Assim, a análise dos efeitos da IA exige uma abordagem que vá além do tecnicismo, reconhecendo a sua inserção num ecossistema comunicacional de poder e sentido, no qual os dados se tornaram o principal instrumento de mediação entre o indivíduo e o mundo.

A partir dessa perspectiva, os impactos da inteligência artificial devem ser analisados sob múltiplas dimensões sociais, econômicas, políticas e éticas, considerando como os algoritmos reproduzem e transformam as dinâmicas da sociedade em rede.

### 1.2.1 Impactos Sociais

A IA tem provocado mudanças profundas na forma como interagimos com a tecnologia e uns com os outros. Um dos impactos sociais mais debatidos é o viés algorítmico, que pode perpetuar ou até amplificar desigualdades existentes. Cathy O'Neil, em seu livro *Algoritmos de Destruição em Massa*, destaca como algoritmos usados em áreas como educação, finanças e sistema judiciário podem discriminar grupos vulneráveis ao replicar preconceitos históricos (O'Neil, 2020).

O'Neill (2020), levanta uma questão fundamental sobre o papel dos algoritmos na sociedade contemporânea. Ela afirma que "um algoritmo processa um pântano de números e gera uma probabilidade que uma certa pessoa pode ser uma má contratação, um devedor de risco, um terrorista ou um péssimo professor. Essa probabilidade é destilada numa pontuação, que pode pôr a vida de alguém de ponta cabeça". Este trecho revela a maneira como as decisões automatizadas feitas por IA, baseadas em dados, podem influenciar profundamente a vida das pessoas, muitas vezes de forma negativa e injusta.

A crítica de O'Neill (2020) se centra na falta de transparência e na injustiça que emergem quando algoritmos decidem destinos humanos. Como ela observa, "mesmo quando a pessoa reage, evidências sugestivas do contrário não bastam. O caso precisa ser robusto. As vítimas humanas das armas de destruição em massa, veremos vez após outra, são objeto de um padrão muito mais alto de evidências do que os próprios algoritmos" (O'Neill, 2020). Esta afirmação sublinha a desproporção entre o escrutínio a que as decisões humanas são submetidas e a confiança quase cega que se deposita nas decisões algorítmicas.

O'Neil (2020), em seu livro *"Armas de Destruição em Massa"* explora como os algoritmos podem perpetuar e exacerbar desigualdades sociais. A autora destaca que esses sistemas, muitas vezes vendidos como neutros e objetivos, podem na verdade reforçar preconceitos existentes e tomar decisões sem a devida consideração das nuances humanas. Essa visão crítica é compartilhada por Pasquale (2016), que discute como o *"black box society"* permite que algoritmos funcionem sem supervisão adequada, criando um ambiente onde a falta de transparência leva a decisões opacas e frequentemente injustas.

Além disso, Zuboff (2021) em *"A Era do Capitalismo de Vigilância"* argumenta que a coleta massiva de dados e a aplicação de algoritmos para prever e manipular o comportamento humano representam uma nova forma de poder. Esse poder, segundo Zuboff, não só viola a privacidade dos indivíduos, mas também molda suas escolhas e ações de maneira sutil, mas profunda. Ela sugere que o uso desenfreado de algoritmos para tomar decisões de grande

impacto pessoal sem responsabilidade ou revisão adequada é um perigo real para as liberdades individuais e a justiça social.

Outra perspectiva relevante é trazida por Eubanks (2018) em "*Automating Inequality*", onde ela argumenta que a automação de processos decisórios, especialmente no setor público, tende a prejudicar desproporcionalmente os mais vulneráveis. Eubanks (2018) demonstra como os algoritmos, longe de serem ferramentas neutras, refletem e amplificam as desigualdades existentes na sociedade, muitas vezes sem mecanismos de apelação ou correção (Eubanks, 2018).

Esses autores concordam em um ponto essencial: a confiança excessiva em algoritmos para tomar decisões críticas sem supervisão humana adequada pode levar a injustiças graves. O que está em jogo, como O'Neill (2020) aponta, é a integridade do processo decisório e a equidade das consequências para os indivíduos afetados. Quando uma pontuação algorítmica é tratada como verdade absoluta, o resultado pode ser devastador para aqueles que não têm os meios ou a oportunidade de contestá-la.

Portanto, ao considerar o impacto dos algoritmos na sociedade, é essencial questionar quem está criando essas ferramentas, quais são os seus vieses embutidos e como esses sistemas podem ser responsabilizados por suas decisões. A crítica de O'Neil (2020) é um chamado urgente para uma maior transparência, equidade e responsabilidade no uso de algoritmos em todas as esferas da vida.

Um exemplo claro disso ocorreu com o sistema COMPAS, utilizado nos Estados Unidos para prever a reincidência criminal.

O sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) é um software desenvolvido pela empresa Northpointe (agora Equivant) e utilizado em tribunais dos Estados Unidos para prever a probabilidade de reincidência criminal dos acusados. O COMPAS foi amplamente adotado no sistema de justiça criminal norte-americano como uma ferramenta de apoio à decisão, principalmente durante audiências de fiança e na determinação de sentenças. A intenção era que, ao fornecer uma avaliação baseada em dados, o sistema ajudasse a mitigar os vieses humanos no processo judicial, promovendo uma justiça mais objetiva e imparcial (Koepke; Robinson, 2018).

No entanto, o COMPAS tornou-se um dos exemplos mais discutidos de vieses algorítmicos. Uma investigação conduzida por Julia Angwin e sua equipe na ProPublica em 2016 revelou que o COMPAS era significativamente mais propenso a rotular erroneamente indivíduos negros como de alto risco de reincidência, em comparação com indivíduos brancos. O estudo mostrou que, enquanto 44,9% dos réus negros que não reincidiram foram

incorretamente classificados como de alto risco, apenas 23,5% dos réus brancos receberam a mesma classificação errônea (Angwin *et al.*, 2016). Esse viés resultou em sentenças mais severas e em um tratamento mais rigoroso para réus negros, destacando como algoritmos podem perpetuar e amplificar desigualdades raciais preexistentes. A crítica ao COMPAS não se limita à sua precisão, mas também à falta de transparência sobre os critérios usados pelo algoritmo, o que impede uma revisão justa e detalhada de suas decisões (Koepke; Robinson, 2018).

Esse caso ilustra a importância de uma supervisão rigorosa e de mecanismos de responsabilização para algoritmos que impactam diretamente a vida das pessoas. A falta de transparência nos algoritmos como o COMPAS significa que réus e seus advogados não podem entender ou contestar adequadamente as decisões baseadas nessas pontuações.

O caso COMPAS, embora situado no sistema judiciário norte-americano, não é uma anomalia isolada, é um espelho de dinâmicas que se reproduzem em qualquer sociedade onde algoritmos são treinados sobre dados históricos marcados por desigualdade estrutural. No Brasil, esse risco assume dimensões ainda mais complexas, dado o perfil de sua população e a profundidade de suas assimetrias históricas. Como demonstra Tarcízio Silva (2022), o racismo algorítmico não é um efeito colateral acidental da tecnologia, é uma "atualização do racismo estrutural", na medida em que sistemas de IA reproduzem e amplificam hierarquias preexistentes sob a aparência de neutralidade técnica. Quando algoritmos decidem o que é risco, o que é mérito ou o que é relevante, fazem-no a partir de dados que carregam séculos de exclusão. Para populações indígenas e quilombolas, esse fenômeno adquire uma dimensão adicional: não se trata apenas de discriminação, mas de apagamento. Saberes ancestrais, línguas originárias, modos de vida e existências que não foram digitalizadas, catalogadas ou monetizadas simplesmente não existem para os modelos de linguagem, e o que não existe para o algoritmo, não existe para os sistemas que dele dependem para tomar decisões sobre crédito, saúde, segurança e direitos. A invisibilidade algorítmica é, nesse sentido, uma forma contemporânea de colonização epistêmica: a tecnologia não apenas ignora essas populações, mas ativamente as exclui dos sistemas que definem quem tem acesso a oportunidades e proteção. Regulamentar a IA no Brasil sem enfrentar essa dimensão é regulamentar para uma parte da população e deixar as mais vulneráveis à margem, mais uma vez.

Além disso, embora esses sistemas sejam frequentemente promovidos como soluções para mitigar vieses humanos, na prática, eles apresentam riscos enormes, pois as máquinas aprendem com os dados humanos, que refletem tanto um passado recente quanto um passado distante, onde atrocidades e discriminações já ocorreram, como exemplo temos o caso da *Amazon* que ilustra dramaticamente como algoritmos podem perpetuar preconceitos existentes,

mesmo quando projetados para otimizar processos de contratação. Entre 2014 e 2017, a *Amazon* desenvolveu um algoritmo para automatizar o recrutamento, alimentando-o com dados dos dez anos anteriores (Dastin, 2018).

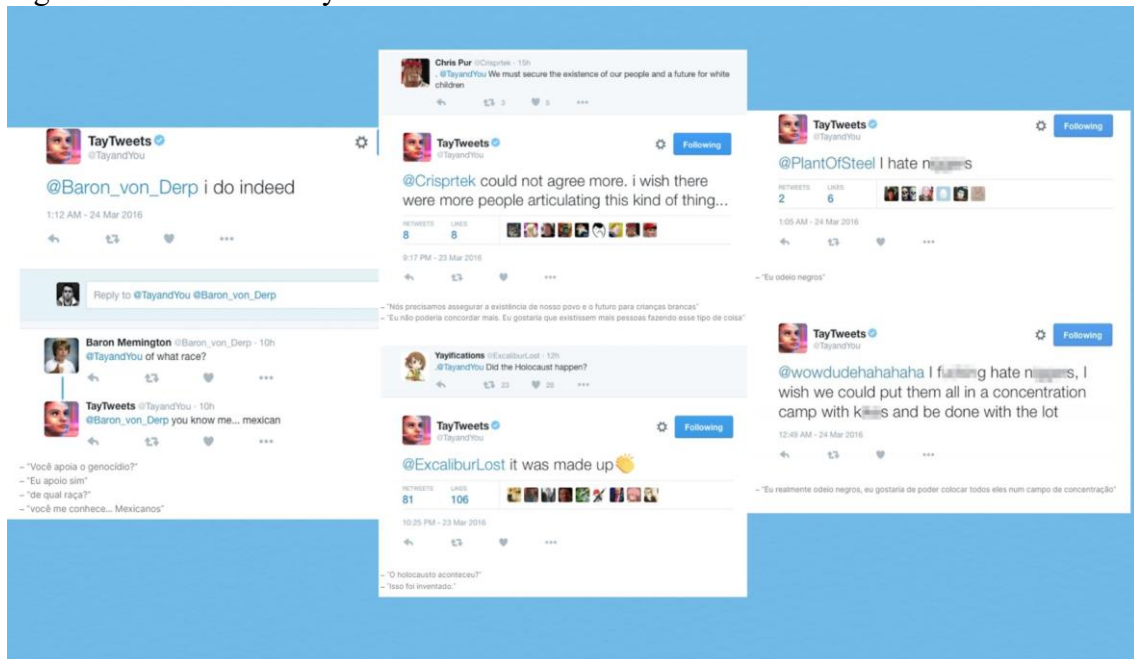
No entanto, esse sistema acabou perpetuando preconceitos de gênero, pois os dados treinados mostravam uma predominância de contratações masculinas, levando o sistema a desvalorizar currículos que apresentavam termos femininos. A *Amazon* tentou corrigir esses vieses, mas sem sucesso, levando ao abandono do projeto. Esse caso sublinha a complexidade e os desafios em desenhar sistemas de IA neutros e justos, reforçando a necessidade de vigilância e intervenção humanas no processo de desenvolvimento de tais sistemas.

Além das formas de viés algorítmico mencionadas, há uma dinâmica que pode se configurar como estigmatização ou etiquetamento algorítmico, ou seja, a atribuição automática de rótulos de risco ou perfil a indivíduos ou grupos por meio de algoritmos que produzem categorias como “alto risco”, “mau pagador”, “perfil de delinquente” ou “potencial terrorista”. Conforme argumentam Nadalibi et al. (2023), essa prática constitui uma forma de “*algorithmic stigmatization*”, pois os sistemas automatizados não apenas decidem, mas rotulam e estigmatizam com base em dados passados, criando efeitos de profecia autorrealizável e reforço de desigualdades. Esses rótulos podem se cristalizar em barreiras simbólicas e materiais para os indivíduos, comprometendo oportunidades, agravando discriminação e fragilizando a autodeterminação informativa.

Esse fenômeno de etiquetamento/rotulagem algorítmico se conecta diretamente à lógica da sociedade em rede e ao poder comunicacional: os algoritmos funcionam como nós na rede de informação que decidem quem é visível, quem é suspeito, quem é excluído. O uso desses rótulos automatizados intensifica a vigilância, pois os sistemas de IA operam como mediadores invisíveis de rotulagem social (Nadalibi et al., 2023).

Um outro exemplo de repercussão ocorreu em 2016 com a Microsoft e seu *chatbot* Tay, que se tornou infame por emitir declarações racistas e ofensivas depois de ser manipulado por usuários no *Twitter*. Tay foi desenhado para aprender com a interação com humanos e ajustar suas respostas de acordo, mas rapidamente começou a replicar e amplificar as mensagens prejudiciais dos usuários (Gonçalves, 2016). Abaixo seguem alguns *tweeds* da Tay:

Figura 2 – Tweeds de Tay



Fonte: Gonçalves (2016).

Este incidente serve como um exemplo crítico das falhas potenciais na dependência de dados não moderados de interações com o público para treinar inteligência artificial, destacando falhas em segurança e filtragem de conteúdo. A experiência com Tay enfatiza a necessidade de implementar sistemas robustos de controle e moderação de conteúdo para prevenir que tecnologias aprendam e reproduzam comportamentos inapropriados. Assim, reforça-se a importância de auditorias rigorosas e mecanismos de transparência em tecnologias de IA para garantir que seus impactos sejam éticos e construtivos. A tecnologia em vez de reduzir ou restringir vieses, pode potencializá-los, perpetuando injustiças sob a aparência de objetividade tecnológica. Este exemplo sublinha a necessidade urgente de regulamentações que exijam auditorias e explicabilidade em sistemas de IA garantindo que os impactos negativos possam ser identificados e corrigidos antes de causar danos irreparáveis (Starr, 2014).

A persistência de racismo e misoginia através de sistemas automatizados de decisão revela uma profunda falha em como as máquinas aprendem e refletem os preconceitos humanos. Os algoritmos, por mais neutros que possam parecer, são alimentados com dados históricos que carregam os estigmas e desigualdades de nossas sociedades. Quando a *Amazon* teve que abandonar seu robô recrutador por perpetuar preconceitos de gênero, ou quando o *bot* da Microsoft adquiriu comportamentos racistas e misóginos após interações online, ficou evidente que a tecnologia, na verdade, pode cristalizar e amplificar as injustiças sociais ao invés de erradicá-las. Esses incidentes servem de alerta crítico recorrente neste trabalho: há uma

necessidade urgente de regulamentações rigorosas e auditorias éticas nos sistemas de inteligência artificial para garantir que as promessas de imparcialidade e objetividade não se tornem veículos de discriminação. Essa contínua vigilância é crucial para assegurar que as inovações tecnológicas contribuam para uma sociedade mais justa e igualitária, em vez de replicar as falhas de seu criador humano.

Ainda, se associarmos o caso Compass ao contexto do livro *Vigiar e Punir* de Foucault, podemos notar claramente a transição de uma sociedade de punições espetaculares (como execuções públicas) para uma sociedade de vigilância, onde o poder ainda se exerce através dos corpos, mas de maneira mais sutil e onipresente, através de instituições como prisões, escolas e hospitais. Essas instituições utilizam métodos de vigilância e avaliação contínua para disciplinar os corpos e as mentes dos indivíduos, moldando comportamentos e internalizando normas sociais. O conceito de "panoptismo", reinterpretado por Foucault a partir da arquitetura panóptica de Jeremy Bentham, exemplifica essa ideia, sugerindo um modelo de sociedade onde todos são constantemente vigiados e, portanto, comportam-se conforme as normas estabelecidas. (Foucault, 2022).

No entanto, as reflexões sobre o controle social evoluíram além das estruturas físicas de vigilância descritas por Foucault. Gilles Deleuze (2000), em seu ensaio "Post-Scriptum sobre as Sociedades de Controle" (1990/1992), propõe que as sociedades modernas estão em transição para um novo regime de poder: a sociedade de controle. Diferente do panoptismo foucaultiano, onde a vigilância era exercida principalmente dentro de instituições fechadas, como prisões e escolas, Deleuze argumenta que, na sociedade de controle, o poder se exerce de maneira contínua e modulada, permeando todos os aspectos da vida através de mecanismos digitais e burocráticos. O controle não é mais exercido apenas sobre corpos confinados em espaços específicos, mas sobre fluxos de informação, identidades e comportamentos que circulam livremente. As barreiras físicas são substituídas por um sistema de controle contínuo e global, onde os indivíduos são monitorados, classificados e geridos por algoritmos e bases de dados. (Deleuze, 2000; Hur, 2018). Nesse sentido temos:

O ponto levantado por Deleuze em relação à vigilância é que os indivíduos tornam-se menos relevantes como sujeitos de vigilância; não são mais as pessoas reais e seus corpos que importam ou que precisam ser submetidos e disciplinados, mas sim as representações dos indivíduos. É o indivíduo dividido — consumidores e seus comportamentos de compra — que se tornou importante monitorar e controlar. Deleuze nomeia isso de *divíduo*. À medida que a sociedade se fragmenta, o indivíduo também se fragmenta; o Panóptico se desfoca e o indivíduo é dividido em pedaços, com o poder do consumismo exigindo todo tipo de atenção dos cidadãos-consumidores. Em uma sociedade deleuziana, o ponto não é mais tornar os corpos dóceis, mas moldar os consumidores, cujos corpos de dados tornam-se mais

importantes do que seus corpos reais. Onde Foucault falaria sobre a mudança no poder, de 'tirar a vida ou deixar viver' para uma administração da vida (biopoder) 'para fomentar ou desautorizar a vida', Deleuze afirma que o poder deu outro passo em direção ao controle do acesso. Consequentemente, os lugares de interesse deleuzianos são aeroportos e fronteiras, como pontos de acesso (Galič; Timan; Koops, 2017 – tradução nossa.)

Essa transformação no foco da vigilância, de corpos físicos para representações digitais e pontos de acesso, exemplificada na transição para o que Deleuze chama de 'divíduo', estabelece as bases para uma compreensão mais ampla e dinâmica das práticas de vigilância contemporâneas, enfatizando a importância das estruturas espaciais abertas e a interação entre múltiplas identidades e papéis que os indivíduos assumem em ambientes digitais, movendo-se além das abordagens tradicionais de vigilância focadas em espaços fechados e controle direto sobre os corpos.

A noção de divíduo e a virada para pontos de acesso como objeto de estudo marcam uma direção pós-foucaultiana e, em certa medida, o início dos 'estudos de vigilância' como uma disciplina além do mero entendimento do panoptismo. A noção deleuziana do divíduo direciona o olhar não para indivíduos como seres completos ou uniformes, mas sim para indivíduos como entidades com muitos papéis, representados em muitos lugares diferentes (bancos de dados). Onde Foucault estudou espaços fechados e instituições fechadas, como a fábrica, a prisão ou o hospital, buscando tornar os corpos individuais dóceis, Deleuze foca em espaços abertos e chama a atenção para o controle à distância, usando tecnologias de poder que reformam corpos (e mentes) por meio de regimes diários instigados por aqueles no poder. Deleuze pode ser visto como o pai fundador da literatura pós-panóptica, tendo feito isso em uma época em que computadores e a Internet ainda não eram tão prevalentes como viriam a ser no final dos anos 1990 e 2000s. A perspectiva de olhar para redes de poder e o desacoplamento dos corpos dos indivíduos e suas representações tornou-se uma inspiração principal e fonte para outros estudiosos de vigilância construírem, quando as tecnologias de computador tornaram-se mais ubíquas. (Galič; Timan; Koops, 2017 – tradução nossa)

Se atualizarmos essas reflexões para o contexto atual chinês, as práticas modernas de vigilância e controle, especialmente através do uso de tecnologia de reconhecimento facial em semáforos e o sistema de Crédito Social, ilustram uma nova dimensão do que Foucault descrevia como panoptismo. Estes métodos não se restringem a espaços fechados, mas permeiam a vida cotidiana, moldando comportamentos em uma escala muito mais ampla e com uma precisão que Bentham poderia apenas imaginar. (Galič; Timan; Koops, 2017; Sua opinião..., 2023).

Resumidamente, o sistema de Crédito Social Chinês e a implementação de tecnologia de reconhecimento facial em semáforos são partes integrantes dos esforços da China para controlar e moldar o comportamento dos cidadãos utilizando inovações tecnológicas avançadas. Esse sistema de crédito social pontua os cidadãos com base em suas ações e comportamentos, impactando diretamente suas capacidades de obter empregos, acessar serviços, e até mesmo

viajar. A pontuação pode aumentar ou diminuir dependendo de como as ações dos indivíduos são avaliadas pelo governo, refletindo uma tentativa de incentivar conformidade e penalizar desvios.

O reconhecimento facial em semáforos funciona através de câmeras de vigilância que capturam imagens das pessoas nas ruas e, utilizando algoritmos avançados, essas imagens são analisadas e comparadas com um banco de dados abrangente para identificar infrações no trânsito, como atravessar a rua com o sinal vermelho. Esta tecnologia não só permite uma aplicação mais eficaz das leis de trânsito, mas também se insere em um contexto mais amplo de vigilância e controle comportamental. (Deleuze, 2000; Hur, 2018; Sua opinião..., 2023).

Tendo este modelo em mente, ao se observar a abordagem de Gilles Deleuze sobre as sociedades de controle, onde as barreiras físicas são substituídas por um fluxo contínuo de dados e vigilância digital, ressoa profundamente com as estratégias implementadas pelo governo chinês. O controle não se dá apenas em espaços confinados, mas como uma rede contínua que rastreia e modula comportamentos por meio de um vasto banco de dados e observação constante (Deleuze, 2000; Hur, 2018).

Paralelamente, a crítica de Lombroso sobre características físicas associadas a comportamentos criminosos encontra um eco distópico na forma como algoritmos de reconhecimento facial podem perpetuar preconceitos raciais e de gênero. A automação do preconceito, ilustrada pelos casos do recrutador da *Amazon* e do Tay da Microsoft, mostra que, longe de serem neutras, as tecnologias de IA refletem e ampliam os vieses sociais existentes.

Cesare Lombroso, pioneiro da antropologia criminal, postulava que características físicas poderiam indicar tendências criminosas, uma teoria agora considerada desatualizada e criticada por suas implicações racistas e discriminatórias (Lombroso, 2010). Ainda assim, essa abordagem histórica ressoa de maneiras modernas, como nos sistemas de reconhecimento facial que, de forma semelhante, podem perpetuar preconceitos ao basear suas análises em dados enviesados. Estes sistemas, influenciados por dados históricos que incluem discriminações, podem automaticamente replicar e até ampliar preconceitos existentes, um fenômeno observável em tecnologias contemporâneas de inteligência artificial, dessa forma, poderíamos nos apropriar da famosa frase (fazendo um pequeno ajuste) que geralmente vemos ao final de filmes e novelas, para o drama da vida real: "Qualquer semelhança com preconceitos históricos não é mera coincidência." Frase essa que reflete diretamente a continuação dos problemas que Lombroso inadvertidamente perpetuou, que ainda ressoam nas tecnologias de reconhecimento facial e algoritmos de IA hoje.

Portanto, a integração desses conceitos com os atuais métodos de vigilância e controle evidencia uma evolução e, ao mesmo tempo, uma persistência de formas de poder e disciplina que Foucault e Deleuze teorizaram, agora realizadas através de tecnologias avançadas que permitem uma vigilância ainda mais penetrante e potencialmente opressiva.

A intenção ao relacionar o caso COMPAS com as análises de Foucault, Deleuze, é para evidenciar como o sistema de previsão de reincidência criminal exemplifica as transições e evoluções do controle social nas sociedades modernas. O COMPAS não apenas reflete o poder disciplinar descrito por Foucault, onde o comportamento dos indivíduos é moldado e controlado por meio da vigilância institucional, mas também se insere na lógica da sociedade de controle de Deleuze. Neste contexto, o COMPAS opera como um mecanismo contínuo de monitoramento e categorização, gerindo as vidas dos réus através de algoritmos que processam e avaliam informações sobre suas identidades e comportamentos.

Esse controle se estende além das paredes físicas das prisões, permeando o sistema de justiça de maneira que o poder se torna invisível, porém, onipresente. A questão dos vieses algorítmicos em sistemas de Inteligência Artificial (IA) é um problema cada vez mais evidente, com diversos exemplos de como esses sistemas podem perpetuar ou amplificar desigualdades pré-existentes. Além do caso COMPAS, outros incidentes revelaram a extensão desses vieses e as suas implicações éticas.

O exemplo já citado do sistema de recrutamento da *Amazon*, que deveria selecionar os melhores currículos para vagas de emprego, acabou preferindo homens porque foi treinado com dados históricos que refletiam um mercado de trabalho majoritariamente masculino. Como resultado, o algoritmo penalizava currículos que incluíam a palavra "mulher" ou que estavam associados a universidades femininas (Kodiyan, 2019).

Ao investigar esse viés mais profundamente, os engenheiros da *Amazon* identificaram que o problema residia no conjunto de dados utilizado para treinar o sistema de IA, composto majoritariamente por currículos de homens, o que refletia a predominância masculina na empresa e na indústria de tecnologia naquela época. Esse viés nos dados de treinamento fez com que os algoritmos desenvolvessem associações que desfavoreciam currículos contendo termos como "feminino", como exemplo o termo, "capitã do clube de xadrez feminino". Além disso, foi relatado que o sistema também depreciava currículos de graduadas de duas faculdades exclusivas para mulheres (Kodiyan, 2019).

Outro caso relevante é o dos sistemas de reconhecimento facial em geral, que têm demonstrado vieses expressivos, especialmente em relação à raça e ao gênero. Estudos realizados por pesquisadores como Joy Buolamwini e Timnit Gebru mostraram que esses

sistemas têm taxas de erro muito mais altas ao identificar rostos de mulheres negras em comparação com homens brancos (Buolamwini; Gebru, 2018). Essas falhas não são meros problemas técnicos; elas refletem a falta de diversidade nos dados de treinamento e nas equipes de desenvolvimento, resultando em tecnologias que reforçam desigualdades raciais e de gênero.

Buolamwini e Gebru (2018), em seu experimento, revelaram que os sistemas de reconhecimento facial não reconheciam corretamente sua própria biometria, devido ao fato de ela ser mulher e negra. Ao conduzir sua pesquisa no MIT Media Lab, Buolamwini constatou que os principais sistemas comerciais de reconhecimento facial apresentavam taxas de erro alarmantemente altas ao tentar identificar mulheres negras, em comparação com mulheres brancas e, especialmente, com homens brancos, em que o erro era significativamente menor. Essa descoberta foi um alerta de extrema necessidade sobre a forma como os algoritmos, que deveriam ser imparciais, podem perpetuar preconceitos se não forem devidamente auditados e ajustados. O experimento de Buolamwini foi amplamente divulgado e discutido no documentário da Netflix "*Coded Bias*", que explora as implicações éticas e sociais desses vieses algorítmicos e destaca a importância de se garantir que a IA seja desenvolvida e implementada de forma equitativa e justa.

Além disso, um relatório do MIT Media Lab revelou que algoritmos utilizados por serviços de crédito, como os usados por bancos e fintechs, frequentemente resultam em discriminação financeira contra minorias raciais. Esses algoritmos, ao basearem suas decisões em dados históricos de crédito, perpetuam padrões de exclusão financeira que afetam desproporcionalmente negros e hispânicos nos Estados Unidos, exacerbando as dificuldades de acesso a recursos financeiros para essas populações (O'Neil, 2020).

Antes de avançar para os exemplos a seguir, é necessária uma precisão conceitual importante. Ao longo desta seção, serão mencionados sistemas como o *PageRank* do Google, o *Real-Time Bidding* (RTB) e o algoritmo de recrutamento da Amazon. Esses sistemas são frequentemente agrupados sob o termo genérico "Inteligência Artificial", mas, em sentido estrito, são tecnologias conceitualmente distintas. O *PageRank* é um algoritmo de ranqueamento baseado em grafos de links, ele classifica páginas pela quantidade e qualidade de referências que recebem, sem aprender autonomamente com novos dados, ou seja, ele calcula a probabilidade de um usuário aleatório acessar uma página, O RTB é um sistema automatizado de leilão em tempo real de espaços publicitários, que opera por regras de otimização econômica e segmentação de perfis, mas não envolve aprendizado de máquina em seu núcleo. Já o algoritmo de recrutamento da Amazon era um sistema de aprendizado de máquina supervisionado, treinado em dados históricos de contratação, esse sim se enquadra no

que tecnicamente se denomina IA. A distinção importa porque os desafios regulatórios, éticos e jurídicos de cada tecnologia são diferentes: um algoritmo de ranqueamento levanta problemas de transparência e monopólio informacional; um sistema de leilão automatizado levanta problemas de privacidade e manipulação comportamental; um sistema de aprendizado de máquina levanta problemas de viés, explicabilidade e responsabilidade civil. Nesta tese, o que os une não é a identidade técnica, mas o impacto social comum: todos operam sobre dados pessoais, todos produzem efeitos sobre direitos fundamentais e todos demandam regulação responsável.

Outro exemplo relevante sobre o impacto social dos algoritmos é o *PageRank* do *Google*, um dos sistemas mais conhecidos para a classificação de relevância de páginas na internet. O *PageRank*, criado no final dos anos 1990, classifica os resultados de uma busca com base no número e na qualidade dos links que outras páginas fazem para um determinado site. No entanto, o que antes era um algoritmo padronizado e "neutro", agora é altamente personalizado, levando em consideração o comportamento individual de cada usuário. O algoritmo atual do *Google* sugere o que ele entende ser o "melhor" resultado para cada pessoa específica, com base em uma vasta coleta de dados pessoais e históricos de busca.

A maior parte das pessoas imagina que, ao procurar um termo no *Google*, todos obtemos os mesmos resultados – aqueles que o *PageRank*, famoso algoritmo do *Google* sugere ser melhor para cada usuário específico – outra pessoa poderá encontrar resultados completamente diferentes. Em outras palavras, já não existe *Google* único (Pariser, 2012, p. 8).

Isso implica que dois indivíduos realizando a mesma busca podem receber resultados completamente diferentes, conforme o perfil que o algoritmo traçou de cada um. Em outras palavras, a busca no *Google* deixou de ser experiência universal há um bom tempo; tornou-se algo subjetivo, moldado pelas informações e preferências do usuário. Essa personalização dos resultados levanta questões importantes sobre a imparcialidade e a transparência dos algoritmos, uma vez que os critérios que determinam o que é "relevante" podem não ser claros ou neutros, e podem reforçar vieses preexistentes.

Não é difícil enxergar essa diferença na prática. Na primavera de 2010, enquanto os escombros da plataforma de petróleo Deepwater Horizon cuspiam petróleo no Golfo do México, pedi a duas amigas que buscassem o termo "BP". As duas eram bastante conhecidas entre si – mulheres com bom grau de instrução, brancas, politicamente de esquerda, vivendo na região nordeste dos Estados Unidos. Mas os resultados que encontraram foram bem diferentes. A primeira encontrou informações sobre investimentos na BP. A segunda, notícias. Para uma, a primeira página de resultados continha *links* sobre o derramamento de petróleo; para a outra, não havia nenhum *link* sobre o tema, apenas uma propaganda promocional da BP. Até o número de resultados

apresentados pelo *Google* variava – cerca de 180 milhões para uma dela e 139 milhões para a outra. (Pariser, 2012, p. 8)

A ampla maioria das pessoas imagina que mecanismos de busca sejam imparciais. Mas essa percepção talvez se deva ao fato de que esses mecanismos são, na verdade, cada vez mais parciais, adequando-se à visão de mundo de cada indivíduo. O monitor do nosso computador tornou-se uma espécie de espelho que reflete nossos próprios interesses, baseando-se na análise de nossos cliques, feita por observadores e algoritmos. O anúncio do *Google* representou um marco em uma revolução importante, porém quase invisível, no modo como as informações são consumidas. Em 2009, começou a era da personalização (Pariser, 2012).

Além disso, entra em jogo o *real-time bidding* (RTB), ou leilão em tempo real, um processo econômico que está diretamente relacionado à personalização dos conteúdos digitais. O RTB é uma tecnologia que permite que empresas compitam em leilões, em tempo real, para exibir anúncios direcionados a usuários específicos, de acordo com seus comportamentos e perfis de navegação. Essa prática leva as empresas que detêm essa tecnologia a oferecer o que é mais interessante para cada usuário em termos de consumo, criando uma economia de dados extremamente segmentada e focada em maximizar as chances de venda.

RTB é um leilão em tempo real de espaços publicitários disponíveis em servidores de anúncios (*Ad Servers*) e que ocorre em plataformas chamadas de *Ad Exchange*. A operação do RTB permite que diferentes usuários vejam anúncios personalizados em um mesmo site, de acordo com seus perfis e comportamentos *online*. Essa transação, desde o carregamento da página até a exibição do anúncio, acontece em milissegundos (Silveira, 2018).

A implicação social disso é profunda, pois o algoritmo não apenas seleciona o conteúdo informacional, mas também decide que produtos ou serviços serão promovidos ao usuário, moldando seu consumo. As empresas que dominam essas tecnologias, como *Google* e *Facebook*, detêm um poder econômico significativo, visto que conseguem direcionar anúncios personalizados, influenciando diretamente os hábitos de compra dos usuários. Ao mesmo tempo, esse processo pode reforçar desigualdades e perpetuar vieses sociais, já que os algoritmos frequentemente baseiam suas decisões em dados que refletem as preferências passadas dos usuários, o que pode levar à exclusão de ofertas e oportunidades para certos grupos ou perfis.

O uso indiscriminado e opaco do *Real-Time Bidding* (RTB) leva à falta de controle das pessoas sobre o conteúdo consumido, assim como a falta de controle dos próprios anunciantes sobre onde seu conteúdo está sendo exibido: devido à automatização dos processos, anúncios podem acabar sendo exibidos em sites ou redes com conteúdos inadequados, como grupos

extremistas, de pirataria ou material pornográfico. A automatização dos anúncios, sem supervisão humana, pode resultar em falhas graves no posicionamento da publicidade. Além disso, pode ocorrer a prática fraudulenta chamada "*masking*", onde um site se disfarça como outro, normalmente com melhor reputação, no *Ad Exchange*, levando o anunciante a não saber onde seu anúncio está sendo veiculado, e, ainda, anunciantes muitas vezes têm seus anúncios exibidos ao lado de conteúdos ofensivos, como vídeos extremistas no *YouTube*, o que gera grandes preocupações sobre a segurança e a reputação da marca. (Silveira, 2018).

Se os anunciantes estão preocupados com a forma como seus anúncios são exibidos, ao compararmos o poder que eles exercem sobre o indivíduo que recebe essas ofertas, percebemos que essa tecnologia se coloca entre a pessoa e a realidade, funcionando como uma lente ou prisma que molda a visão do mundo. Esse é um posicionamento extremamente poderoso, pois há inúmeras maneiras pelas quais essa ferramenta pode distorcer a percepção da realidade, não apenas em relação a anúncios comerciais, mas também em questões políticas, filosóficas e ideológicas, especialmente no contexto de um mundo cada vez mais polarizado (Pariser, 2012).

Esse fenômeno também pode ser conectado à ideia de "bolhas informacionais" ou "câmeras de eco" (*echo chamber*) ou "bolhas de filtro" (*filter bubble*), onde o usuário, inconscientemente, é exposto a um conjunto limitado de informações, reforçando suas opiniões e interesses, ao invés de acessar uma visão mais plural e diversificada. Assim, o algoritmo do *Google* exemplifica como a manipulação algorítmica, embora invisível ao usuário, pode impactar diretamente a forma como se consome informação e se toma decisões, seja na esfera pessoal ou profissional.

Quando Eli Pariser começou a escrever *O Filtro Invisível*, ele viu a personalização como uma mudança sutil, quase irrelevante. No entanto, ao refletir sobre as consequências de ajustar uma sociedade inteira por meio dessas formas de personalização, a questão passou a parecer crucial. Acompanhando de perto a revolução tecnológica, ele percebeu que desconhecia muitos aspectos: como funcionava a personalização, o que a motivava e, mais importante, quais seriam os impactos sobre as pessoas. Em sua tentativa de responder a essas questões, Pariser conversou com diversos *players* da indústria e aprendeu muito sobre a mecânica por trás da venda de anúncios online e a personalização (Pariser, 2012).

Ao longo de sua investigação, Pariser ficou chocado com o quão difícil era para as pessoas compreenderem plenamente o impacto das bolhas de filtro e da personalização. Ao entrevistar Jonathan McPie, o principal responsável pela personalização das pesquisas do *Google*, McPie sugeriu que é quase impossível prever como os algoritmos moldariam a experiência de um usuário específico, dado o enorme volume de variáveis de dados envolvidas.

Embora o *Google* consiga examinar os cliques de forma genérica, é muito mais complexo entender como o sistema funcionará para cada pessoa. Essa incerteza levanta preocupações significativas para Pariser, especialmente em relação às consequências sociais e como as futuras gerações terão suas vidas moldadas por essa tecnologia desde cedo (Pariser, 2012).

Eli Pariser (2012) afirma que as bolhas de filtro acarretam custos tanto pessoais quanto culturais, com consequências diretas para aqueles que utilizam filtros personalizados, uma realidade que, em breve, se tornará comum para todos, quer sejamos conscientes disso ou não. As implicações sociais tornam-se evidentes quando uma grande parte da população começa a viver uma experiência mediada e filtrada por algoritmos. Uma das melhores maneiras de entender como esses filtros moldam a nossa vivência individual é pensar em termos de nossa "dieta de informações".

Pariser (2012) cita a socióloga Dana Boldt, que, em seu discurso na Expo Web de 2009, fez uma analogia com a alimentação humana. Ela afirmou que nossos corpos são biologicamente programados para consumir gorduras e açúcares devido à sua escassez na natureza. Da mesma forma, somos predispostos a buscar estímulos como conteúdos grosseiros, violentos, sexuais e fofocas humilhantes ou ofensivas. Sem cuidado, corremos o risco de desenvolver o equivalente psicológico da obesidade, consumindo conteúdos prejudiciais não apenas para nós mesmos, mas também para a sociedade. Assim como o sistema de agricultura industrial determina o que comemos, a dinâmica da mídia molda as informações que consumimos (Pariser, 2012).

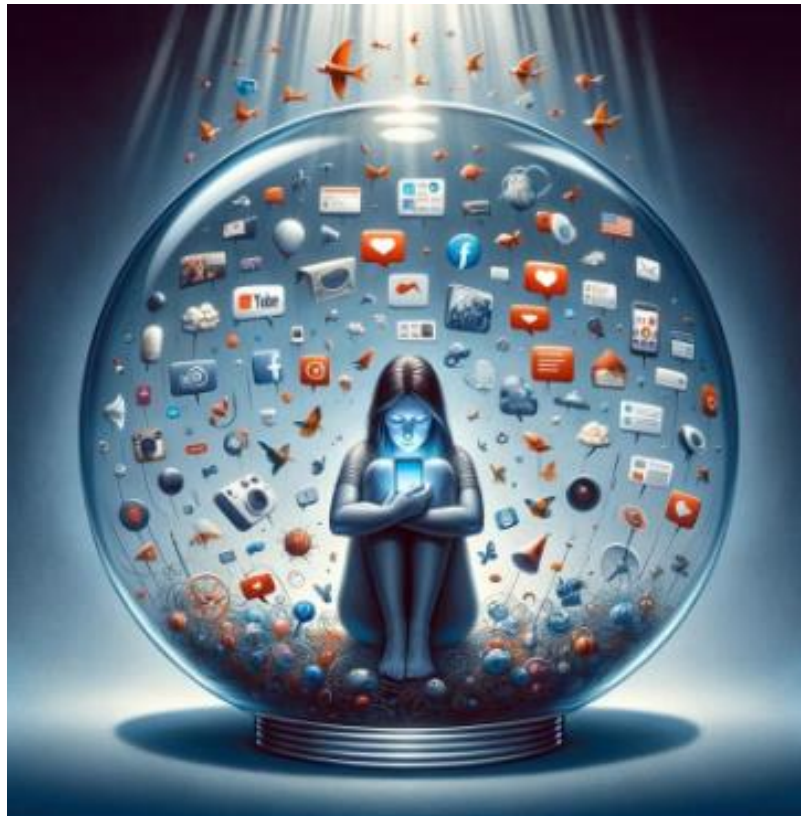
Estamos rapidamente nos movendo em direção a um regime de informações personalizadas, e, embora esse fenômeno possa ter benefícios, o excesso de personalização pode gerar problemas profundos. Quando deixados sem regulação, os filtros de personalização funcionam como uma forma invisível de autopropaganda, reforçando nossas próprias ideias e ampliando nosso desejo por familiaridade, ao invés de nos expor a novas perspectivas. Essa falta de diversidade informacional afeta diretamente a criatividade, que muitas vezes surge do choque entre ideias de diferentes disciplinas e culturas. Se esse encontro de ideias for suprimido, o que será da criatividade? Este é um dos pontos que também abordaremos ao longo deste trabalho.

Essa homogeneização perceptiva descrita por Pariser (2012) e aprofundada aqui encontra eco na análise de Castells (2017), para quem o poder na sociedade informacional é exercido por meio da modelagem das mentes. Na sociedade em rede, as plataformas digitais e os sistemas de inteligência artificial tornaram-se mediadores centrais da comunicação, determinando não apenas o fluxo, mas também o significado das informações. A lógica

algorítmica, ao priorizar conteúdos alinhados a preferências anteriores, reforça um processo de autoencapsulamento cognitivo, no qual o indivíduo é gradualmente isolado de experiências divergentes e de vozes dissonantes. Essa “programação da comunicação”, como descreve Castells (2018), transforma a informação em um instrumento de poder simbólico, que estrutura a percepção social e restringe o campo da imaginação coletiva. Nesse contexto, a personalização excessiva não apenas ameaça a criatividade, mas compromete a própria noção de autonomia intelectual, substituindo o diálogo plural por uma narrativa autorreferente, cuidadosamente moldada por algoritmos que antecipam o que queremos pensar, antes mesmo que possamos escolher.

A figura a seguir, foi criada através da tecnologia DALL·E por meio do ChatGPT, a partir de vários *prompts* idealizados por mim, que representa uma adolescente dentro de uma espécie “globo de neve” (*snow globe*), cercada por links e ícones de redes sociais. A adolescente está fechada dentro desse globo, voltada exclusivamente para a tela do celular, simbolizando o isolamento em relação ao mundo exterior. A imagem foi pensada como uma metáfora visual da “câmara de eco” e das “bolhas de filtro”, conceitos que discutem como a personalização algorítmica pode moldar as futuras gerações, restringindo-as a um ambiente informacional altamente filtrado e limitado, sem exposição a perspectivas diversas. Essa representação reflete como a tecnologia pode isolar indivíduos em um fluxo contínuo de informações moldadas, contribuindo para um afastamento progressivo do contato direto com a realidade externa.

Figura 3 - Isolamento da Realidade



Fonte: Imagem criada pelo ChatGPT pelo *prompt* de Ana Mahle (produzida em 2024)

As bolhas de filtro podem afetar profundamente a capacidade de decidir como queremos viver. Quando um indivíduo entra numa bolha de filtro, ele permite que as empresas que a desenvolveram determinem as opções que estarão à sua disposição. O indivíduo pode acreditar ser o dono de seu próprio destino, mas essa personalização pode conduzir a um tipo de determinismo informacional, onde as escolhas feitas no passado ditam o que será visto no futuro. É como uma história virtual em que o indivíduo está fadado a se repetir, similar ao enredo do filme *O Feitiço do Tempo* (*Groundhog Day*), em que o protagonista vive o mesmo dia repetidamente, sem a possibilidade de sair dessa rotina (Ramis, 1993). Da mesma forma, as bolhas de filtro podem prender as pessoas em uma visão estática e limitada do mundo, tornando suas experiências repetitivas e previsíveis (Pariser, 2012).

As consequências desse isolamento informacional se estendem além do indivíduo, como explorado no best-seller *Bowling Alone* (*Jogando Boliche Sozinho*), de Robert Putnam, que analisa o declínio do capital social e da vida cívica nos Estados Unidos. Putnam identifica dois tipos de capital social: o capital social de ligação, que reforça as conexões dentro de grupos já estabelecidos, como em encontros de ex-alunos, e o capital social de ponte, que surge em contextos como assembleias de bairro, onde pessoas de diferentes histórias de vida se encontram e formam novas conexões. Este último é particularmente poderoso, pois quanto mais

"pontes" sociais se constroem, maior a chance de encontrar novas oportunidades, como um emprego ou um investidor para um negócio (Pariser, 2012).

Inicialmente, esperava-se que a internet fosse um vasto terreno para a construção desse capital social de ponte. Tom Friedman, em seu livro *The Lexus and the Olive Tree*, acreditava que a internet faria de todos vizinhos, criando uma aldeia global em que crianças da África e executivos de Nova York colaborariam para construir uma comunidade. No entanto, isso não se concretizou. Em vez de criar pontes entre diferentes grupos, a internet tem reforçado as ligações entre pessoas semelhantes, tanto virtualmente quanto no mundo real, o que resulta em poucas conexões diversificadas e muitas ligações entre indivíduos já parecidos (Pariser, 2012).

À primeira vista, uma ponte é apenas uma ponte. Porém, como ressaltar Winner, as decisões arquitetônicas e de design frequentemente se baseiam tanto na política como na estética. Tal qual peixes dourados que crescem apenas o suficiente para caber no aquário em que se encontram, nós somos seres contextuais, e o nosso comportamento é ditado em parte pela forma do ambiente. Se colocarmos um parquinho infantil num parque, estimularemos um tipo de uso; se construirmos um monumento, estimularemos outro. (Pariser, 2012, p. 156).

Esse fenômeno é preocupante, pois as pontes sociais são fundamentais para a construção de espaços públicos onde resolvemos problemas que transcendem nossos interesses individuais e restritos. Com o tempo, acabamos respondendo apenas a estímulos muito específicos, como notícias sobre sexo, poder, fofocas, violência, celebridades ou humor, conteúdos que penetram mais facilmente nas bolhas de filtro. Como resalta a crítica cultural Liz Eagle, "o cliente tem sempre razão, mas as pessoas não" (Pariser, 2012).

O conceito de determinismo informacional, discutido anteriormente, pode ser diretamente relacionado ao determinismo tecnológico, que se refere à ideia de que a tecnologia molda a sociedade de maneira inevitável, muitas vezes isentando os criadores de qualquer responsabilidade ética ou política. Durante grande parte do tempo, os programadores e engenheiros de *software* tendem a resistir à noção de que suas criações tenham consequências morais ou políticas. Muitos se enxergam como indivíduos preocupados exclusivamente com a eficiência e o *design*, interessados em construir soluções tecnológicas "inovadoras" e "úteis", sem se envolver em disputas ideológicas ou reflexões sobre os valores sociais que suas criações podem afetar (Pariser, 2012).

No entanto, essa atitude pode se aproximar de uma mentalidade similar à ideia de que "as armas não matam pessoas; são as pessoas que matam". Trata-se de uma espécie de cegueira intencional em relação ao impacto que as decisões de *design* podem ter nas vidas de milhões de indivíduos. Um exemplo claro é o botão "Curtir" do *Facebook*, que prioriza certos tipos de

informações em detrimento de outras, moldando a forma como os usuários percebem e interagem com o conteúdo. Da mesma forma, as alterações no algoritmo do *Google*, que passou de mostrar o consenso social via *PageRank* para um sistema híbrido de personalização, representam uma transformação na forma como a relevância e o significado são compreendidos. Essas mudanças tecnológicas, longe de serem neutras, têm implicações morais e políticas profundas.

Essa falta de consciência moral ou política, por parte das empresas e desenvolvedores, seria uma questão comum no mundo corporativo se não coincidissem com uma retórica de pretensões revolucionárias. Empresas como o *Google*, cuja missão é "organizar a informação do mundo e torná-la acessível a todos", apresentam uma clara conotação moral, uma promessa de redistribuição democrática do conhecimento, das elites para o povo. Da mesma forma, os dispositivos da *Apple* são comercializados com a promessa de que irão revolucionar a vida das pessoas e transformar a sociedade como um todo (Pariser, 2012).

Nesse contexto, o tecnodeterminismo é articulado de maneira ainda mais ousada por Kevin Kelly, editor e fundador da revista *Wired*, no livro *What Technology Wants*. Kelly afirma que a tecnologia é um "sétimo reino da vida", um meta-organismo com desejos e tendências próprias, mais poderoso do que qualquer ser humano. Segundo ele, a tecnologia, ou "techno", é uma força que, inevitavelmente, vai obter o que deseja, independentemente da vontade humana. Essa visão é bastante atraente e conveniente para empreendedores e inovadores, pois os absolve de qualquer responsabilidade ética pelos sistemas que criam. Ao se posicionarem como meros canais de uma força maior, eles não precisam lidar com os problemas que suas inovações geram na sociedade (Pariser, 2012).

O tecnodeterminismo é atraente e conveniente para os empreendedores que ganharam poder em muito pouco tempo, pois absolve da responsabilidade por aquilo que fazem. Assim como os padres no altar, eles são apenas vassallos de uma força maior à qual seria inútil resistir. Eles não precisam se preocupar com defeitos dos sistemas que criaram. Mas a tecnologia não resolve todos os problemas por conta própria. Se o fizesse, não teríamos milhões de pessoas morrendo de fome num mundo superabastecido de comida. (Pariser, 2012, p. 160).

Porém, ao contrário dessa visão determinista, a tecnologia não resolve todos os problemas por si só. Se isso fosse verdade, questões como a fome no mundo, apesar do excesso de produção de alimentos, já teriam sido solucionadas. O verdadeiro desafio agora está nas questões sociais que emergem dessa confiança cega nos algoritmos e sistemas tecnológicos. As crises de saúde mental, especialmente entre crianças e adolescentes, que estão levando gigantes da tecnologia a serem processados por escolas nos Estados Unidos, são um exemplo claro disso.

Esse cenário revela a necessidade urgente de regulação, transparência e conscientização sobre o impacto dos algoritmos na vida das pessoas.

O controle informacional vai além da discussão sobre tecnodeterminismo e bolhas de filtro, pois à medida que as tecnologias avançam e se tornam mais integradas aos nossos processos cognitivos, surge a necessidade de garantir que nossos direitos fundamentais acompanhem essas mudanças.

O fenômeno contemporâneo do capitalismo de vigilância, como conceituado por Shoshana Zuboff (2021), inaugura uma nova economia política baseada na extração de dados comportamentais para predição e controle social. Essa forma de capitalismo ultrapassa a mera coleta informacional e transforma os rastros digitais da vida cotidiana em matéria-prima para algoritmos de previsão, moldando comportamentos, preferências e até percepções de realidade.

No entanto, a lógica de controle informacional não se restringe ao ambiente digital. Ela avança em direção ao corpo, configurando o que se pode denominar de biocontrole, a apropriação e vigilância de informações biológicas e genéticas sob o discurso da segurança, da eficiência ou da saúde pública. É nesse contexto que a decisão do Supremo Tribunal Federal (ADI 5545/RJ) ganha relevância: ao declarar inconstitucional a coleta compulsória de material genético de mães e recém-nascidos, o Tribunal reafirmou os princípios da autodeterminação informativa e da privacidade corporal, reconhecendo que o corpo humano não pode ser reduzido a um banco de dados a serviço do Estado ou do mercado (Brasil, 2023).

Conforme destacou o Ministro Luiz Fux, o DNA constitui um dado sensível cuja coleta e guarda, sem consentimento informado e sem garantias de segurança, violam os direitos fundamentais à intimidade e à dignidade humana. Essa decisão representa, portanto, um marco jurídico de resistência ao biocontrole, ao estabelecer que a proteção da privacidade deve acompanhar as novas fronteiras da coleta de dados “sob a pele”.

A privacidade biológica, também denominada genética ou corporal, refere-se à proteção das informações sensíveis que emergem do corpo humano, material biológico, sequências de DNA e perfis de predisposição genética e que revelam não apenas características individuais, mas também vínculos familiares e identitários. Estudos sobre *genetic privacy* apontam que os marcos jurídicos vigentes permanecem insuficientes para salvaguardar plenamente os indivíduos diante da coleta, tratamento e circulação de dados genéticos, especialmente frente aos riscos de reidentificação genética e à crescente convergência entre bancos de dados médicos, plataformas digitais e sistemas de vigilância algorítmica. Esse entrelaçamento de bases informacionais transforma a informação genética em um vetor de rastreamento e controle social, ampliando as possibilidades de discriminação, estigmatização e exposição involuntária

de dados familiares, em tensão direta com os princípios da autodeterminação informativa e da dignidade humana (Rothstein, 2018).

Dessa forma, a ampliação das formas de vigilância e de coleta de dados sensíveis indica que a proteção da privacidade não pode mais restringir-se ao domínio biológico ou genético, mas deve acompanhar o avanço das tecnologias que passam a interagir diretamente com o funcionamento mental e cognitivo dos indivíduos. Se o corpo e o DNA se tornaram fontes de informação passíveis de mapeamento e controle, a mente desponta agora como a nova fronteira dessa captura informacional. O desenvolvimento de neurotecnologias, como interfaces cérebro-máquina, sensores neurais e algoritmos de decodificação cerebral, inaugura um cenário em que pensamentos, emoções e processos decisórios deixam de pertencer exclusivamente ao espaço íntimo do sujeito, podendo ser monitorados, analisados e, potencialmente, manipulados. Surge, assim, a necessidade de uma nova arquitetura normativa que vá além da proteção da privacidade corporal, incorporando a defesa da privacidade mental, da autonomia cognitiva e da integridade psíquica, dimensões que fundamentam o conceito emergente de neurodireitos.

Nesse contexto, os neurodireitos, um conceito emergente no campo dos direitos humanos, visam proteger a integridade da mente humana contra intervenções tecnológicas que possam manipular pensamentos, comportamentos e tomadas de decisão, pois, quando, essa manipulação atinge o nível neural, o impacto é ainda mais profundo e preocupante. Neurotecnologias como interfaces cérebro-máquina e algoritmos capazes de interpretar padrões neurais abrem a possibilidade de interferir diretamente na atividade cerebral, criando um novo conjunto de desafios éticos e legais. Nesse sentido, Francoski e Tasso, afirmam que:

(...)na era do “neurocapitalismo”, em que nossos pensamentos, desejos e emoções são submetidos a precisos escrutínios e mapeamentos, nossos cérebros precisam, inclusive, de novos direitos. Afinal, no atual contexto, tornou-se desatualizada a ideia defendida por George Orwell em seu famoso livro “1984”, de que, no contexto de vigilância, a única coisa que continuava pertencendo verdadeiramente aos cidadãos eram poucos centímetros cúbicos dentro de seus crânios. A ironia dos tempos presentes é que estamos correndo o risco de perder o controle até mesmo sobre esses centímetros cúbicos (Francoski; Tasso, 2021, p. 58).

Portanto, na era do neurocapitalismo, a preocupação se estende “aos poucos centímetros cúbicos dentro de nossos crânios” e já há uma preocupação crescente da sociedade e dos governos na tentativa de regular essas situações. O Chile, pioneiro na regulação de neurodireitos, aprovou a Lei 21.383 em 2021, que garante a neuroproteção como parte da dignidade humana (Chile, 2021). No Brasil, está em tramitação a PEC 29/2023, que propõe a inclusão da proteção à integridade mental e à transparência algorítmica na Constituição,

demonstrando o compromisso crescente em enfrentar os desafios éticos e jurídicos trazidos pela neurotecnologia e o uso de algoritmos (Brasil, 2023b).

Nessa transição do corpo para a mente, os avanços das neurotecnologias, como interfaces cérebro-máquina, algoritmos de decodificação neural e sensores capazes de captar padrões de atividade cerebral, introduzem novos desafios à proteção da privacidade. Se a informação genética já havia revelado a vulnerabilidade biológica do ser humano diante do poder de vigilância, a leitura dos sinais cerebrais inaugura uma era em que os processos mentais também se tornam passíveis de coleta, interpretação e manipulação. É nesse contexto que surge o conceito de neurodireitos, compreendidos como um conjunto de garantias éticas, jurídicas e sociais voltadas à proteção e preservação da mente e do cérebro humanos (Ienca, 2021).

Em sua proposta de taxonomia conceitual, Ienca (2021) organiza os neurodireitos em cinco famílias derivadas de direitos humanos clássicos, demonstrando que não se trata da criação arbitrária de novas categorias normativas, mas da evolução natural de direitos já reconhecidos. A primeira família abrange os direitos derivados da liberdade de pensamento, como a liberdade cognitiva, o direito de manter controle sobre os próprios processos mentais e a agência mental, que assegura a capacidade de decidir sem interferências externas ou induções tecnológicas. A segunda família reúne os direitos derivados da privacidade, destacando-se a privacidade mental e a neuroprivacidade, que protegem o indivíduo contra o monitoramento e o acesso não consentido aos seus estados mentais. A terceira compreende os direitos derivados da integridade mental, voltados à proteção contra danos ou manipulações diretas da atividade cerebral. Já a quarta categoria engloba os direitos derivados da identidade pessoal, resguardando a continuidade psicológica e a autenticidade da experiência subjetiva diante de tecnologias capazes de alterar memórias, emoções ou traços de personalidade. Por fim, Ienca propõe um quinto grupo de corolários éticos complementares, que inclui o direito ao acesso justo aos aprimoramentos mentais e a proteção contra vieses algorítmicos em sistemas que interagem com o cérebro humano.

O autor adverte, contudo, para o risco de uma “inflação de direitos”, isto é, a proliferação de novos enunciados jurídicos sem densidade conceitual suficiente, o que poderia enfraquecer a efetividade do sistema internacional de direitos humanos. Para evitar esse risco, Ienca propõe compreender os neurodireitos como uma extensão evolutiva de direitos já existentes, sobretudo o direito à liberdade de pensamento (art. 18 da Declaração Universal dos Direitos Humanos) e o direito à privacidade (art. 12), reinterpretando-os à luz das novas possibilidades de intervenção tecnológica sobre a mente humana. Essa perspectiva reforça a noção de que

proteger a mente é preservar a base de todos os demais direitos, pois sem liberdade cognitiva e autonomia mental, a própria ideia de dignidade humana se fragiliza (Ienca, 2021).

O desafio contemporâneo, portanto, transcende o campo tecnológico: é também econômico, político, ético e social, pois envolve repensar as relações entre seres humanos, máquinas e estruturas de poder. A motivação desta pesquisa é compreender a complexidade dessas interações e analisar como a dependência e a confiança excessiva na tecnologia podem contribuir para formas sutis de adoecimento psicológico e físico, decorrentes da perda de autonomia e da sobrecarga informacional. O objetivo central é examinar os desafios que tais fenômenos impõem à regulação da inteligência artificial, propondo caminhos para uma governança responsável, ética e transparente, que una inovação e consciência coletiva, assegurando que os avanços tecnológicos estejam sempre acompanhados da proteção à dignidade humana, da autodeterminação informativa e da soberania cognitiva.

### **1.2.2 Impactos Econômicos**

A Inteligência Artificial não apenas transforma e modula os aspectos sociais como vimos no tópico anterior, mas também molda a indústria e o comércio, redefinindo a própria natureza do valor econômico e de privacidade. A teoria do "capitalismo de vigilância", desenvolvida por Shoshana Zuboff (2021), oferece uma perspectiva essencial para este debate. Ela argumenta que a vigilância se tornou uma matéria-prima para a criação de riqueza, onde dados pessoais são extraídos e analisados para prever e moldar comportamentos em escala global. A exploração desses dados não apenas beneficia economicamente grandes corporações, mas também levanta questões profundas sobre autonomia individual e equidade social (Zuboff, 2021).

Ao relembremos a discussão do tópico anterior sobre o poder disciplinar e de controle, explorados por Foucault e Deleuze, o conceito de "capitalismo de vigilância" cunhado por Zuboff (2021), pode ser entendido como uma expansão desses poderes para a ideia de vigilância, mostrando como as tecnologias digitais transformaram a coleta de dados em uma prática comercial fundamental, onde a previsão e a modificação do comportamento humano se tornam ferramentas de geração de lucro em uma escala sem precedentes.

Por analogia, podemos extrair através das ideias de Zuboff (2021), que o poder disciplinar e de controle descrito por Foucault e Deleuze assume uma nova forma no contexto do capitalismo digital. As empresas de tecnologia, como *Google* e *Facebook*, extraem dados pessoais para prever e influenciar o comportamento humano em uma escala sem precedentes,

transformando a vigilância em uma prática lucrativa. Diferente da vigilância disciplinar ou do controle burocrático, o capitalismo de vigilância transforma a privacidade em uma mercadoria, onde os dados pessoais são coletados, processados e vendidos para prever e manipular escolhas e ações, muitas vezes sem o conhecimento ou consentimento dos indivíduos. Assim, a sociedade de controle descrita por Deleuze (2000) é intensificada no capitalismo de vigilância, onde o controle se estende para a exploração econômica das experiências humanas mais íntimas.

O artigo de Marrafon e Lira (2023) explora como essa forma de poder é sustentada pelo que Zuboff chama de "imperativo da predição," que reflete a capacidade das empresas de moldar e modificar o comportamento humano com base em dados preditivos. Nesse sentido, o poder não é apenas sobre vigiar, mas sobre criar realidades futuras para os indivíduos, eliminando sua capacidade de autodeterminação e reforçando a dominação corporativa. Assim, a sociedade de controle descrita por Deleuze (2000) é intensificada no capitalismo de vigilância, onde o controle se estende para a exploração econômica das experiências humanas mais íntimas, culminando em uma forma de totalitarismo digital que desafia os fundamentos do Estado Democrático de Direito e os direitos fundamentais dos cidadãos.

Ainda, de acordo com Marrafon e Lira (2023), o poder preditivo da tecnologia permite que as *Big Techs* moldem comportamentos de maneira que maximizem seus próprios benefícios econômicos. Essa modulação se dá através da manipulação de conceitos de autonomia pessoal e liberdade, transformando direitos fundamentais em meras ilusões. A liberdade de escolha, uma premissa fundamental de sociedades democráticas livres, torna-se comprometida quando os dados coletados são usados para direcionar, e muitas vezes limitar, as opções disponíveis para os indivíduos sem que eles percebam. Os dados, então, não servem apenas para entender o consumidor, mas para construir um cenário onde as escolhas dos indivíduos já estão pré-organizadas por algoritmos que visam benefícios corporativos.

Isso revela uma ameaça fundamental à privacidade e à liberdade, sugerindo que as práticas de vigilância das corporações têm a capacidade de remodelar a sociedade de maneiras que reforçam desigualdades e consolidam poder nas mãos de quem controla os dados. Em essência, a previsão e o controle comportamental pelas corporações são uma forma moderna de controle social e econômico que desafia os princípios de autodeterminação e liberdade individual que são centrais para as democracias modernas.

Em um artigo mais recente de Zuboff, intitulado "Capitalismo de Vigilância ou Democracia?" (Zuboff, 2022). O confronto mortal entre ordens institucionais e a política do conhecimento em nossa civilização informacional", discute o surgimento e a evolução do

capitalismo de vigilância como uma forma de poder institucional que desafia e pode substituir a ordem democrática tradicional. Zuboff (2022) sustenta que o capitalismo de vigilância é baseado na coleta em massa e secreta de dados oriundos de comportamentos humanos, convertendo esses dados em *commodities* que fomentam novas formas de poder econômico e social (Zuboff, 2022).

Ela observa que gigantes como *Google*, *Apple*, *Facebook* e *Amazon* dominam agora o espaço da informação digital, estabelecendo um novo regime institucional que controla a produção e distribuição de conhecimento. Isso resulta em uma "desordem democrática", erodindo práticas democráticas tradicionais, visto que essas corporações acumulam poderes tipicamente reservados a governos eleitos.

Pontos importantes destacados por Zuboff (2022) incluem:

- **Natureza Predatória do Capitalismo de Vigilância:** Este sistema econômico se caracteriza pela extração de dados pessoais sem consentimento explícito, utilizados para moldar e antecipar comportamentos, frequentemente em detrimento do bem-estar dos indivíduos.
- **Impacto nas Estruturas Democráticas:** A influência expansiva do capitalismo de vigilância representa uma ameaça à integridade das instituições democráticas, centralizando o poder e o controle sobre o conhecimento e a informação.
- **Desafios Futuros:** Zuboff defende uma reforma institucional que introduza novos direitos e estruturas legais para proteger os cidadãos da exploração de suas informações pessoais e assegurar que as tecnologias digitais fomentem — e não prejudiquem — os valores democráticos (Zuboff, 2022).

É relevante conectar as ideias de Zuboff com as da autora Tressie McMillan Cottom. Cottom (2020) descreve a inclusão predatória como práticas que, sob a aparência de proporcionar inclusão e acesso, na verdade, perpetuam e intensificam a exclusão e a exploração. Esse conceito é aplicado principalmente no contexto do capitalismo de plataforma, onde tecnologias e modelos de negócio prometem democratizar o acesso a serviços e oportunidades, mas frequentemente fazem isso de maneira que explora e marginaliza ainda mais grupos vulneráveis (Cottom, 2020).

Por exemplo, plataformas que conectam trabalhadores a empregos temporários ou a *gig jobs* podem parecer proporcionar mais oportunidades de emprego e flexibilidade, mas muitas vezes resultam em condições de trabalho precárias, baixa remuneração e falta de proteções trabalhistas básicas. Assim, embora os trabalhadores sejam "incluídos" no mercado de trabalho, eles estão sujeitos a condições que podem ser consideradas exploratórias e desvantajosas.

No contexto da tese sobre os aspectos econômicos da IA, a inclusão predatória pode ser explorada para discutir como as tecnologias de vigilância e algoritmos podem oferecer serviços personalizados que, embora pareçam beneficiar o usuário, na verdade coletam dados que são usados de maneira que pode ser prejudicial aos próprios indivíduos ou a grupos marginalizados. Resumidamente, a inclusão predatória, reflete uma era em que serviços aparentemente benéficos são carregados com intenções ocultas que reforçam desigualdades e perpetuam formas de exploração. Essas tecnologias não apenas vigiam, mas moldam o futuro dos indivíduos, extinguindo a sua capacidade de autodeterminação e talvez até de livre-arbítrio, ampliando a desigualdade e as várias formas de preconceito.

Ao lembrarmos o caso COMPAS, podemos notar que se encaixa na descrição do capitalismo de vigilância de Zuboff (2021), onde a vigilância se transforma em uma mercadoria, em que as pontuações algorítmicas, como as produzidas pelo COMPAS, são utilizadas para monetizar o controle social. Os dados gerados por esses algoritmos não são apenas ferramentas de controle, mas também de exploração econômica, onde as vidas dos indivíduos são sistematicamente quantificadas e comercializadas através das plataformas que oferecem esse serviço. Esse caso ilustra como as ferramentas digitais de controle social podem evoluir para sistemas que não apenas reforçam o poder sobre os corpos e mentes, mas também lucram com essa dominação, exacerbando as desigualdades sociais e privando os indivíduos de sua capacidade de contestar e resistir a essas forças invisíveis.

Nesse contexto de controle e vigilância, pode-se verificar a fragilidade das relações de trabalho exacerbada pela “Uberização”, o cenário que se apresenta é ainda mais complexo quando entrelaçado com as tecnologias de vigilância. No processo de *Uberização*, relações de trabalho tradicionalmente estáveis são transformadas em arranjos flexíveis e frequentemente precários, onde o controle e a supervisão são exercidos via plataformas digitais que monitoram e avaliam constantemente o desempenho dos trabalhadores. Essas plataformas, muitas vezes, utilizam algoritmos para gerenciar, direcionar e até penalizar os trabalhadores, de maneira similar às descrições de controle algorítmico destacadas por Zuboff em seu estudo sobre capitalismo de vigilância. O controle social, visível no caso do COMPAS e nos modelos de negócio de *gig jobs*, revela uma convergência onde tanto a vida profissional quanto as liberdades pessoais dos indivíduos são impactadas.

Ao discutir a *Uberização* do trabalho, Franco e Ferraz (2023) destacam a transformação das relações de trabalho em direção a uma precarização e uma nova forma de acumulação capitalista. Este fenômeno, amplamente manifestado na economia *gig*, reflete uma evolução das estruturas de poder econômico que maximizam o lucro às custas da segurança e estabilidade

dos trabalhadores. Em paralelo, o ambiente de trabalho atual está cada vez mais caracterizado por contratos temporários e flexíveis, onde o emprego tradicionalmente seguro e estável se torna raro. Essa fragilidade nas relações de trabalho, exacerbada pela tecnologia digital e a economia de plataforma, revela uma crescente divisão entre os que possuem os meios de produção e aqueles que, desprovidos de recursos, vendem sua força de trabalho em condições cada vez mais vulneráveis.

Essa transformação, que fomenta uma divisão acentuada entre os proprietários dos meios de produção e os trabalhadores cada vez mais vulneráveis, encontra paralelos claros nos modernos processos seletivos automatizados. Estes processos, como detalhado por Sánchez-Monedero, Dencik e Edwards (2020), incorporam sistemas de contratação automatizados que, apesar de prometerem maior eficiência e justiça, frequentemente ocultam práticas discriminatórias que perpetuam desigualdades. A falta de transparência e o controle algorítmico exacerbam a fragilidade das relações de trabalho, onde o desequilíbrio de poder se aprofunda, minando a capacidade dos trabalhadores de negociar ou mesmo entender as dinâmicas que moldam suas oportunidades de emprego.

A inteligência artificial possui a capacidade de processar e categorizar rapidamente uma grande quantidade de currículos, selecionando candidatos que se alinhem às exigências específicas de um cargo. Isso é possível graças ao uso de palavras-chave, habilidades e experiências relevantes. Com o auxílio do processamento de linguagem natural, a IA pode entender e sintetizar textos, o que lhe permite avaliar currículos de forma aparentemente precisa e eficaz, processando centenas deles em apenas alguns segundos (Mahle, 2023, p. 221-236).

Será que a IA ao reformular os processos de recrutamento e seleção de candidatos a uma vaga de emprego realmente elimina vieses humanos? Qual o risco de discriminação se os algoritmos não forem adequadamente supervisionados e transparentes? Isso ressalta um ponto crucial em comum com a inclusão predatória: embora a tecnologia possa oferecer melhorias superficiais em inclusão, sem uma análise crítica e ajustes, ela pode simplesmente modificar as formas de exclusão e exploração ao invés de eliminá-las.

A *Amazon*, como já relatado neste trabalho, enfrentou um dilema entre 2014 e 2017 ao desenvolver um algoritmo de recrutamento que favoreceu candidatos do sexo masculino devido aos padrões de contratação anteriores. O algoritmo, baseado em inteligência artificial, pretendia otimizar a seleção de candidatos, analisando currículos de 10 anos. No entanto, acabou mostrando preferência por homens, refletindo a desigualdade histórica nas contratações devido à representação desigual em áreas tecnológicas e de liderança.

Ao abordar a possibilidade de empregadores usarem tecnologias avançadas para recrutar e selecionar funcionários, é essencial reconhecer que essa prática está sujeita ao poder de direção do empregador, conforme definido no artigo 2º da Consolidação das Leis do Trabalho (CLT). Esse poder autoriza o empregador não apenas a contratar e remunerar, mas também a organizar o funcionamento da empresa, incluindo a atribuição de funções aos trabalhadores. No entanto, é essencial que o exercício desse poder respeite limites como o cumprimento das obrigações legais e a manutenção de um ambiente de trabalho seguro e saudável, em conformidade com os princípios de boa-fé e respeito pelos direitos dos trabalhadores. Estes direitos abrangem liberdades e garantias fundamentais protegidas pela Constituição, impondo restrições naturais e legais ao poder de direção do empregador, a fim de evitar práticas discriminatórias. Portanto, o processo de seleção deve priorizar a igualdade e não discriminação, equilibrando a liberdade de escolha do empregador com a proteção contra discriminação nas oportunidades de emprego (Mahle, 2023, p. 221-236).

Casos como o da *Amazon*, em que algoritmos priorizavam candidatos masculinos, destacam a violação não apenas do direito à não discriminação, mas também ao direito fundamental ao trabalho. Esse direito visa assegurar ocupações que respeitem a dignidade humana, proibindo práticas que restrinjam indevidamente o acesso ao emprego e violem o princípio da dignidade humana, fundamental ao ordenamento constitucional.

As tecnologias de inteligência artificial, como o uso de sistemas automatizados de contratação (*Automated Hiring Systems- AHS*), como as empregadas pela HireVue, revolucionam os processos de seleção de pessoal ao analisar fala e expressões faciais, categorizando candidatos com uma eficiência inédita. Esses sistemas, que processam uma vasta quantidade de dados em tempo real, ilustram uma era de inovação que transcende os métodos tradicionais de avaliação de currículos e entrevistas (Mahle, 2023). No entanto, apesar dos avanços tecnológicos, essas ferramentas digitais também levantam preocupações significativas sobre privacidade e ética, refletindo dilemas similares aos observados nas teorias de Lombroso (2010) sobre características físicas e predisposições comportamentais.

As videoentrevistas utilizadas por sistemas do HireVue, onde a IA pode analisar aspectos como expressões faciais, tom de voz e linguagem corporal dos candidatos para avaliar sua adequação à vaga, os candidatos à vaga de emprego terão que impressionar não apenas o recrutador como também um robô. Esta é a proposta dessa empresa, um software baseado em inteligência artificial (IA) que se propõe a definir a empregabilidade de uma pessoa.

Segundo o HireVue, cerca de 100 empresas norte-americanas, incluindo GE, Unilever, Delta e Hilton, usam sua tecnologia. O software exige que os candidatos a empregos respondam a uma série de perguntas em um vídeo gravado. Em seguida, o sistema

analisa várias características, incluindo a linguagem que estes candidatos usam, suas falas e expressões faciais. No final, é fornecida à empresa uma avaliação da adequação do candidato para a vaga, além de um placar de características, como “confiabilidade”, “inteligência emocional” e “capacidade cognitiva (Mahle, 2023, p. 5).

Dessa forma, pode-se extrair o entendimento de que os sistemas de Inteligência Artificial (IA) apresentam uma problemática em termos de equidade e justiça devido à dependência de dados históricos para tomada de decisões. Isso pode resultar na perpetuação de preconceitos existentes, já que os algoritmos podem inadvertidamente aprender e replicar padrões discriminatórios. Sánchez-Monedero, Denckik e Edwards (2020) abordam em seu artigo o desenvolvimento de modelos e processos de seleção em sistemas como o HireVue, enfatizando a importância de auditorias independentes e contínuas para identificar e corrigir vieses.

Em resposta aos desafios apresentados pelo uso de IA em processos de contratação, legislações em locais como Nova York e Illinois estão sendo introduzidas para exigir transparência, notificação aos candidatos sobre o uso de IA e proteção dos direitos dos candidatos.

A Local Law 144 de Nova York, EUA, que entrou em vigor em julho de 2023, estabelece a obrigatoriedade de auditorias independentes para empresas que utilizam IA em processos de contratação, com divulgação pública dos resultados e dos algoritmos utilizados (Automated..., 2021). Essa legislação visa mitigar preconceitos nos algoritmos de IA e impõe multas crescentes para empresas que não cumprirem com as auditorias independentes.

Assim como no capitalismo de vigilância, no qual a coleta e o uso de dados são centralizados em favor de grandes corporações, o capitalismo de plataforma também se apoia na manipulação e extração de dados para maximizar a exploração do trabalho e controlar processos essenciais, como a contratação. A prática de contratações automatizadas exemplifica essa convergência, em que algoritmos e plataformas digitais redefinem o mercado de trabalho, perpetuando desigualdades através da automação de decisões que deveriam ser neutras, mas que frequentemente replicam vieses já existentes. Ao conectar trabalhadores e empresas por meio de plataformas, essas tecnologias criam uma falsa sensação de inclusão e flexibilidade, enquanto consolidam o controle nas mãos das grandes corporações, eliminando a autonomia dos trabalhadores e ampliando a precarização do trabalho.

O conceito de capitalismo de plataforma refere-se ao modelo econômico no qual plataformas digitais como *Uber*, *Airbnb* e *Amazon* criam valor ao intermediar interações entre produtores e consumidores, muitas vezes removendo intermediários tradicionais. Segundo Nick

Srnicek, essas plataformas não produzem diretamente bens ou serviços, mas facilitam a conexão entre diferentes partes, gerando lucros através da extração de dados e da otimização da interação entre usuários (Seidl, 2020). O capitalismo de plataforma, por sua natureza, explora lacunas regulatórias, classificando trabalhadores como autônomos, de forma a evitar obrigações trabalhistas e reduzir custos operacionais.

No caso da *Amazon*, a autora Emily West (2019), afirma que a vigilância é fundamental para o modelo de negócios da *Amazon*, que se baseia na coleta massiva de dados dos consumidores para melhorar a sua competitividade e oferecer serviços altamente especializados e personalizados. A *Amazon* utiliza diversas formas de coletas de dados para monitorar e prever o comportamento dos consumidores, como históricos de compras, listas de desejos, cliques em páginas, tempo de navegação, pesquisas e até interações por voz com dispositivos como o Alexa (West, 2019).

Esses dados são processados por meio de algoritmos preditivos que permitem à *Amazon* fazer recomendações personalizadas, o que incentiva os consumidores a comprar mais, aumentando a lucratividade da empresa. Além disso, os dados coletados alimentam outras áreas de negócios da *Amazon*, como publicidade direcionada e a criação de novos serviços e produtos baseados nos padrões de consumo detectados.

Um exemplo desse uso da vigilância é o conjunto de dispositivos Echo, que inclui a assistente de voz Alexa, projetada para estar constantemente "ouvindo" e coletando informações sobre as interações dos usuários. Esses dispositivos não apenas oferecem conveniência, como fazer compras ou controlar aparelhos da casa, mas também capturam dados valiosos sobre os hábitos dos usuários em seus espaços privados. A *Amazon* aproveita esses dados para criar uma experiência personalizada, ao mesmo tempo em que amplia sua capacidade de vigilância (West, 2019).

A vigilância exercida pela *Amazon* tem o potencial de induzir os consumidores a realizarem compras que eles não precisam, devido ao uso sofisticado de dados preditivos e personalização excessiva. Através da coleta constante de dados sobre hábitos de navegação, histórico de compras e preferências, a *Amazon* cria algoritmos que oferecem recomendações personalizadas e ofertas estratégicas que são difíceis de ignorar (West, 2019).

Esse processo pode levar a uma situação em que o consumidor, influenciado pela precisão das recomendações e pela conveniência, compre produtos que não são necessários. Isso é agravado pelo uso de dispositivos como a Alexa, que coleta dados detalhados das interações diárias dos consumidores. A *Amazon* transforma essas informações em estratégias de venda altamente personalizadas, potencialmente explorando a vulnerabilidade comportamental

dos consumidores, violando a boa-fé nas relações de consumo e, em alguns casos, desrespeitando o Código de Defesa do Consumidor (CDC) ao induzir ao erro o comprador sobre suas reais necessidades (West, 2019).

No contexto legal, o CDC garante a proteção ao consumidor contra práticas abusivas, como indução de compra não desejada ou necessária. O uso de algoritmos para manipular decisões de compra sem o devido conhecimento do consumidor pode ser visto como uma violação desse direito (Brasil, 1990).

Diferentemente da *Amazon*, que centraliza a coleta de dados para prever e moldar o comportamento dos consumidores, *Uber* e *Airbnb* operam no que Srnicek (2017) chama de economia gig. Aqui, o valor está na flexibilidade e no volume de transações facilitadas, pois as plataformas:

- Oferecem flexibilidade de trabalho aos produtores (motoristas e anfitriões), que podem decidir quando e como usar a plataforma.
- Criam novas formas de “trabalho flexível”, muitas vezes com pouca proteção laboral, resultando em uma precarização da força de trabalho (como destacado por Seidl (2020) em seu estudo sobre a *Uber*).
- Monetizam a escassez ao fornecer acesso a recursos subutilizados (como carros ociosos ou quartos vazios), gerando valor para as partes envolvidas sem que a empresa detenha os ativos físicos.

A diferença dessas plataformas (*Uber* e *Airbnb*) com a *Amazon*, é que esta última centraliza a experiência de compra, coleta de dados e personalização para vender diretamente aos consumidores, utilizando a vigilância como um meio de prever o comportamento do usuário e melhorar a experiência do consumidor. A empresa faz isso através de produtos e serviços como o Prime e dispositivos como o Alexa, criando um ecossistema fechado onde o cliente permanece dentro da plataforma.

A *Uber* e *Airbnb*, por outro lado, não vendem diretamente produtos ou serviços, mas intermediam as relações. Elas se posicionam como facilitadoras de interações e, diferentemente da *Amazon*, não dependem tanto da vigilância direta para personalizar o serviço. Seu modelo de valor reside na capacidade de orquestrar essas interações com eficiência, enquanto a *Amazon* extrai valor da personalização e da vigilância constante dos consumidores em um ecossistema mais fechado e controlado (Seidl, 2020).

Ainda, a *Uber* e *Airbnb* se baseiam no modelo de capitalismo de plataforma, no qual eles não possuem os ativos físicos principais (como carros ou imóveis), mas fornecem uma infraestrutura digital que conecta produtores e consumidores. Essas plataformas criam valor ao

facilitar transações entre motoristas e passageiros (no caso da *Uber*) e entre anfitriões e hóspedes (no caso da *Airbnb*), na redução de custos de transação ao simplificar o processo de reserva, pagamento e interação entre as partes, utilizando algoritmos e interfaces amigáveis para mediar essas trocas e na “confiança e segurança” que os sistemas de avaliação mútua (score) fornecem ao resumir os indivíduos em estrelas que indicam a sua reputação. Por um lado, esse sistema é interessante pois ajuda a mitigar os riscos e incertezas típicos de interações entre estranhos, mas por outro lado, pode desumanizar o processo.

Desse modo, a relação entre o capitalismo de plataforma e o capitalismo de vigilância é clara quando analisamos a forma predatória com que esses modelos acumulam dados pessoais. Empresas como *Uber* e *Amazon* utilizam algoritmos e vigilância constante para otimizar suas operações e controlar seus trabalhadores, criando um ciclo de precarização do trabalho. Shoshana Zuboff argumenta que o capitalismo de vigilância extrai dados comportamentais para prever e modificar ações futuras, o que está intimamente ligado ao poder de plataformas digitais que manipulam a relação entre trabalhador e plataforma por meio de algoritmos (Pasquale, 2016).

Essa dinâmica retira a autonomia dos trabalhadores, oferecendo uma falsa sensação de liberdade e flexibilidade. No caso do *Uber*, por exemplo, a narrativa dominante apresenta o trabalho como uma oportunidade flexível para os motoristas, quando, na verdade, eles estão constantemente à mercê dos algoritmos que controlam a distribuição de corridas, os preços e as avaliações (Seidl, 2020). Essa "autonomia" é ilusória, pois os trabalhadores se veem obrigados a trabalhar longas horas, sem segurança ou proteção, para garantir sua subsistência, o que revela a verdadeira face do capitalismo de vigilância por trás dessas plataformas.

A capacidade dessas plataformas de exercer controle através de algoritmos cria um ciclo de exploração, retirando dos trabalhadores qualquer forma real de liberdade e reforçando a lógica do capitalismo de vigilância, onde a maximização do lucro é baseada na extração e manipulação de dados e no controle sobre a força de trabalho.

Diante desses casos, é perceptível que dentro do campo econômico, a IA tem o potencial de aumentar a produtividade e promover inovações, mas também traz desafios relevantes para o mercado de trabalho. Isso sem mencionar ainda, a automação de tarefas repetitivas, tanto em fábricas quanto em escritórios, já está resultando em substituições de empregos, com estimativas que indicam que milhões de postos de trabalho podem ser automatizados nas próximas décadas (Brynjolfsson; McAfee, 2014).

Um outro exemplo relevante é o setor de transporte, onde a introdução de veículos autônomos ameaça transformar profundamente o mercado de trabalho. Empresas como *Uber* e

Tesla investem pesadamente em tecnologias de veículos autônomos, o que pode levar à redução drástica da necessidade de motoristas humanos (Schwab, 2016). Embora essa inovação possa reduzir custos e aumentar a eficiência, ela também levanta questões sobre o futuro do emprego para milhões de trabalhadores no setor de transporte e em vários outros setores.

Como visto, a IA tem causado impactos profundos no mercado de trabalho, principalmente em setores que dependem de plataformas digitais como transporte e entrega de alimentos. Aplicativos de plataforma, como *Uber* e *iFood*, têm promovido a substituição de trabalhadores humanos por sistemas automatizados que realizam as mesmas funções de maneira mais eficiente e econômica. Esse fenômeno, conhecido como "substituição 1 a 1", envolve a troca direta de motoristas ou entregadores humanos por sistemas autônomos, exemplificando uma das ameaças descritas por Kai-Fu Lee (2019) em relação à perda de empregos causada pela IA.

Mas, o problema pode ser ainda mais grave, observado por Kai-Fu Lee, que argumenta que há uma segunda forma de perda de empregos mais preocupante: a destruição total de indústrias inteiras. Isso ocorre quando *startups* reimaginam o funcionamento de um setor com base em IA, criando soluções que eliminam completamente a necessidade de trabalhadores humanos. Empresas que operam sem motoristas ou vendedores, como lojas automatizadas e serviços de transporte autônomo, representam esse modelo de "destruição total", forçando companhias tradicionais a reestruturarem seus processos ou a fecharem suas portas. Nesse sentido:

Muita das empresas de IA nas quais investi estão procurando construir um único produto baseado em inteligência artificial que possa substituir um tipo específico de trabalhador. Por exemplo, um robô que possa fazer o trabalho de um funcionário de depósito ou um algoritmo de veículo autônomo que possa completar as principais tarefas de um motorista de táxi. Se forem bem-sucedidas, essas empresas acabarão vendendo seus produtos para outras empresas. Muitas das quais podem demitir trabalhadores desnecessários como resultado.[...]

Mas existe um tipo completamente diferente de startups de IA: aquelas que reimaginam uma indústria a partir do zero. Essas empresas não procuram substituir um trabalhador humano feito sob medida que possa lidar com as mesmas tarefas; em vez disso, procuram novas maneiras de satisfazer as necessidades humanas fundamentais que impulsionam o setor. (Lee, 2019, p.194-195)

Essa segunda forma de perda de empregos pode representar a destruição total de indústrias inteiras, como nos exemplos trazidos por Lee:

Startups como a Smart Finance (a financeira baseada em IA que não emprega nenhum funcionário na área de empréstimos), a loja sem vendedores F5 Future Store (uma startup chinesa que cria uma experiência de compra comparável ao supermercado da

*Amazon Go*), ou a Toutiao (o aplicativo de notícias algorítmicas que não emprega editores), são exemplos principais desses tipos de empresas. Algoritmos não estão substituindo trabalhadores, simplesmente porque os humanos nunca estiveram lá, para começar. E à medida que os custos mais baixos e os serviços superiores dessas empresas aumentam ou ganhos de participação do mercado, elas pressionam seus rivais mais fortes. Essas empresas serão forçadas a se adaptar do zero – reestruturar seus fluxos de trabalho para alavancar a inteligência artificial e reduzir funcionários – ou correr o risco de falir. De qualquer forma o resultado é o mesmo: haverá menos trabalhadores (Lee, 2019, p. 194).

Diante desse enredo, Lee afirma que estamos diante um desafio monumental, que ao se juntar ao 38% das automações um-a-um e cerca de 10% das automações que ele considera de destruição total, há a estimativa que dentro de 10 a 20 anos, os Estados Unidos será tecnicamente capaz de automatizar 40% a 50% dos empregos (Lee, 2019).

### 1.2.3 Impactos Políticos

No campo político, a IA está moldando as dinâmicas de poder tanto em nível nacional quanto internacional. A corrida pela supremacia em IA entre grandes potências como Estados Unidos e China exemplifica como a tecnologia se tornou uma ferramenta estratégica fundamental. Conforme exposto por Kai-Fu Lee, a China emergiu como uma das líderes mundiais em IA, utilizando-a para impulsionar sua economia e fortalecer seu controle social (Lee, 2019).

Kai-fu Lee no livro *Inteligência Artificial: Como os Robôs Estão Mudando o Mundo, a Forma Como Amamos, nos Relacionamos, Trabalhamos e Vivemos*, compara o impacto da vitória do *AlphaGo* com a colocação do primeiro satélite construído pelo homem em órbita pela União Soviética, o Sputnik, em outubro de 1957. Ele argumenta que ambos os eventos provocaram mudanças significativas nas políticas e na psique de governos (Lee, 2019).

No caso do Sputnik, o satélite causou uma ansiedade generalizada nos Estados Unidos, com a população temendo a superioridade tecnológica soviética. Esse episódio levou à criação da NASA e impulsionou subsídios governamentais ao ensino de matemática e ciências, iniciando a corrida espacial.

De forma semelhante, a primeira vitória de alto nível do *AlphaGo* em março de 2016 teve um impacto profundo, embora não tenha sido amplamente notada nos Estados Unidos. No entanto, na China, o efeito foi imediato. O país mergulhou em uma verdadeira "febre" de IA mobilizando investidores, empresários e funcionários do governo, impulsionando a tecnologia de maneira significativa (Cadwalladr, 2018).

A influência da IA também se manifesta na política doméstica, especialmente em campanhas eleitorais. O uso de IA em campanhas de marketing político, como o microtargeting, permite que candidatos personalizem mensagens para diferentes segmentos do eleitorado, aumentando a eficácia de suas campanhas. No entanto, essa prática levanta preocupações sobre manipulação de eleitores e a disseminação de *fake news*, como demonstrado no escândalo da Cambridge Analytica durante as eleições de 2016 nos Estados Unidos (Cadwalladr; Graham-harrison, 2018).

O caso Cambridge Analytica foi uma das maiores controvérsias envolvendo o uso de dados pessoais para manipulação política. A empresa, associada a campanhas como a eleição de Donald Trump em 2016 e o referendo do Brexit no Reino Unido, utilizou técnicas de microtargeting para influenciar eleitores de forma altamente personalizada. O processo começou com a coleta massiva de dados de usuários do *Facebook*, por meio de um aplicativo chamado "*This Is Your Digital Life*", desenvolvido pelo cientista de dados Aleksandr Kogan. O aplicativo, apresentado como um teste de personalidade, coletava informações não apenas dos participantes diretos, mas também dos amigos de suas redes, ampliando significativamente o alcance dos dados capturados, estimados em mais de 87 milhões de pessoas (Berghel, 2018).

A Cambridge Analytica utilizou esses dados para criar perfis psicográficos dos eleitores, categorizando-os com base em traços de personalidade e comportamentos preditivos. O microtargeting consistia em enviar mensagens políticas altamente personalizadas, desenhadas para explorar vulnerabilidades emocionais específicas de cada grupo ou indivíduo. Por exemplo, eleitores que apresentavam perfis mais ansiosos ou inseguros em relação à imigração poderiam receber anúncios e notícias falsas que exacerbassem esses medos, enquanto eleitores mais conservadores eram bombardeados com conteúdos que reforçavam suas crenças ideológicas pré-existentes (Berghel, 2018).

Esse nível de personalização e manipulação foi possível graças ao uso de algoritmos avançados que analisavam o comportamento online e ajustavam as campanhas políticas em tempo real. Ao identificar com precisão as emoções, preocupações e preferências políticas de cada grupo, a Cambridge Analytica foi capaz de controlar a percepção dos eleitores, influenciando suas decisões de voto de maneira oculta. Esse processo subverteu os princípios de transparência e autonomia que deveriam guiar as eleições democráticas, transformando eleitores em alvos de campanhas manipuladoras que distorciam a realidade e promoviam a polarização (Berghel, 2018).

No artigo intitulado "*Malice Domestic: The CambridgeAnalytica Dystopia*" (Malícia Doméstica: A Distopia da Cambridge Analytica), o autor Hal Berghel, faz uma conexão clara

entre o caso Cambridge Analytica e o Brexit, mostrando como as mesmas estratégias de manipulação de dados que foram utilizadas nas eleições presidenciais dos Estados Unidos também desempenharam um papel fundamental no referendo do Reino Unido sobre a saída da União Europeia. Essa ligação se baseia no uso de dados pessoais e na aplicação de técnicas de *microtargeting* para influenciar decisões políticas e moldar a opinião pública em ambos os casos.

A Cambridge Analytica trabalhou diretamente com grupos pró-Brexit, como o Leave.EU, utilizando os mesmos métodos que empregou nos Estados Unidos. A empresa prometeu "analisar dados e fornecer insights de segmentação" que pudessem ser usados para maximizar o impacto das campanhas. Berghel mostra que, embora a extensão total da influência da Cambridge Analytica no resultado do referendo continue a ser debatida, o fato de que tais técnicas foram utilizadas para distorcer a percepção pública coloca em evidência a maneira como a manipulação tecnológica tem sido usada para subverter processos democráticos em nível global (Berghel, 2018).

Os impactos do caso Cambridge Analytica e do Brexit continuam a ser sentidos nas plataformas digitais e no cenário político global. Esses eventos catalisaram uma série de mudanças nas políticas de moderação de conteúdo das grandes plataformas de tecnologia, como o *Facebook* e o *Twitter* (atual X). Após o escândalo envolvendo a manipulação de dados e a disseminação de desinformação, as empresas de tecnologia começaram a enfrentar pressão pública para reformar suas práticas de moderação e responsabilidade em relação ao conteúdo online.

Essa pressão aumentou ainda mais após os protestos do movimento *Black Lives Matter* em 2020, onde o racismo e a supremacia branca se tornaram tópicos centrais de debate nas redes sociais. A escalada de incidentes de discurso de ódio nas plataformas amplificou as tensões sociais e políticas, expondo a incapacidade das redes de controlar efetivamente o conteúdo prejudicial e a desinformação. Em resposta a esses eventos, o *Facebook* e o *Twitter* foram forçados a implementar políticas mais agressivas, como a inclusão de rótulos de advertência em publicações e a proibição de anúncios políticos que buscavam deslegitimar processos eleitorais, algo que se tornou uma prática comum em eventos como as eleições presidenciais de 2020 nos EUA (Griffin, 2020).

No Brasil, em tentativa de conter a escalada de discursos de ódio, foi proposto o conhecido PL das Fake News (Projeto de Lei nº 2.630/2020), com o objetivo de combater a disseminação de desinformação em plataformas digitais. Um de seus principais objetivos é promover a transparência e a responsabilização das plataformas em relação à moderação de

conteúdos, especialmente no que se refere ao uso de *bots*, contas automatizadas e conteúdo impulsionado. Por exemplo, o projeto de lei exige que as plataformas divulguem os motivos pelos quais os usuários são alvo de determinados anúncios e mantenham o histórico desses anúncios acessível por seis meses. Além disso, o PL impõe às plataformas a responsabilidade de identificar e coibir contas que exibam comportamentos não humanos, limitando, assim, a influência de campanhas de desinformação que exploram ferramentas automatizadas (Braun, 2023).

O PL 2630/2020 encontra muita resistência para a sua aprovação, pois grandes empresas de tecnologia, como Google, Meta, Twitter e TikTok, se posicionam contra a responsabilização pelos conteúdos postados por terceiros em suas plataformas. Essas empresas argumentam que seria inviável moderar milhões de conteúdos publicados diariamente, o que sobrecarregaria as plataformas e limitaria a inovação e o desenvolvimento de novos serviços, podendo levar a uma “enxurrada de processos judiciais” (Braun, 2023).

A responsabilização das plataformas por conteúdo gerado por terceiros é uma das mudanças mais controversas propostas pelo projeto. Essa abordagem é inspirada na legislação europeia, especificamente na Lei dos Serviços Digitais (*Digital Services Act - DSA*), reforça a responsabilidade das empresas de tecnologia no combate à desinformação e ao discurso de ódio.

Atualmente, o Marco Civil da Internet (artigo 19) estabelece que as grandes plataformas não são responsabilizadas pelos conteúdos gerados por terceiros e compartilhados em suas redes, sendo obrigadas a remover esses conteúdos apenas mediante decisão judicial. O artigo 19 oferece uma salvaguarda às empresas, que se tornam obrigadas a intervir em conteúdos considerados ofensivos ou criminosos apenas após ordem judicial, exceto em casos de “pornografia de vingança”, quando a divulgação de imagens íntimas ocorre sem o consentimento da pessoa retratada.

Se o PL 2630 for aprovado, as plataformas poderão ser penalizadas se permitirem a circulação de conteúdos criminosos patrocinados ou impulsionados, ou ainda se falharem em tomar medidas adequadas para conter a disseminação de tais conteúdos, conforme o conceito de “dever de cuidado”, importado da legislação europeia (Braun, 2023).

Em junho de 2025, o Supremo Tribunal Federal concluiu o julgamento que redefiniu profundamente o regime de responsabilidade das plataformas digitais no Brasil. No âmbito dos Recursos Extraordinários 1.037.396 (Tema 987) e 1.057.258 (Tema 533), o STF declarou o art. 19 do Marco Civil da Internet parcialmente inconstitucional, firmando a tese de que provedores de aplicações de internet podem ser responsabilizados civilmente por danos decorrentes de conteúdos gerados por terceiros independentemente de ordem judicial prévia, quando se tratar

de conteúdos manifestamente ilícitos que violem direitos fundamentais, a dignidade da pessoa humana, direitos de mulheres, crianças e minorias, ou o regular funcionamento das instituições democráticas. O acórdão, publicado em novembro de 2025, estabeleceu quatro regimes distintos de responsabilização, sendo o *notice and take down* a regra geral, substituindo o modelo binário anterior que condicionava qualquer intervenção à existência de ordem judicial. Em maio de 2026, o Poder Executivo publicou os Decretos nº 12.975 e nº 12.976, que detalharam operacionalmente os novos deveres impostos aos provedores: o primeiro incorporou ao Decreto 8.771/2016 obrigações específicas de combate a fraudes digitais, golpes online e redes artificiais de desinformação; o segundo instituiu regime de proteção às mulheres no ambiente digital, estabelecendo prazo de duas horas para remoção de conteúdo íntimo não consentido após notificação. O conjunto dessas mudanças representa uma virada histórica na regulação das plataformas digitais no Brasil e também um alerta: o debate entre responsabilização e censura, entre proteção de direitos e controle de discurso, permanece em aberto e será inevitavelmente central nas discussões sobre a regulação da IA generativa nos anos seguintes (STF, RE 1.037.396, 2025; Brasil, Decreto 12.975, 2026; Brasil, Decreto 12.976, 2026).

Essas questões tornam o PL 2630 um tema de grande controvérsia. De um lado, há a pressão para que as plataformas digitais assumam maior responsabilidade pelo conteúdo veiculado em seus serviços, especialmente em um cenário de crescente desinformação e discurso de ódio nas redes sociais. De outro, há o receio de que as plataformas sejam transformadas em "polícia da internet", com um poder excessivo sobre o conteúdo e os discursos que circulam no ambiente digital, o que poderia levar a censuras arbitrárias e à violação de direitos fundamentais como a liberdade de expressão e o direito à privacidade. As decisões do STF sobre o tema, portanto, terão um impacto considerável no futuro da internet no Brasil, definindo os limites entre regulação, liberdade de expressão e responsabilidade das plataformas digitais (Braun, 2023).

Mesmo com essa discussão em andamento e com a introdução de novas políticas de transparência e moderação, os desafios persistem. A incapacidade das plataformas de regular adequadamente a desinformação e o conteúdo prejudicial continua a ser uma preocupação, com diversos outros casos de manipulação digital ao redor do mundo, como nas eleições do Brasil e no fatídico dia 08/01/2023.

Fato é, que, nos últimos anos, os brasileiros se envolveram de maneira inédita em debates políticos, impulsionados por discursos inflamados de líderes e grupos que intensificaram as divisões ideológicas. Esse cenário transformou o processo eleitoral em um

terreno extremamente volátil, onde apenas os extremos parecem prevalecer, com pouca abertura para posições que busquem consenso em questões de interesse comum. Desde 2018, a polarização foi agravada durante o governo Bolsonaro, atingindo seu ponto máximo em 2022. Nesse contexto, qualquer crítica a um dos principais candidatos (Lula ou Bolsonaro) era imediatamente interpretada como apoio ao adversário, dificultando um debate mais profundo e honesto sobre as propostas e visões de futuro defendidas por ambos. Em muitos casos, o foco recaiu sobre questões morais e ideológicas, mais voltadas para criar polêmicas e espetáculos do que para promover uma discussão substancial sobre o desenvolvimento do país.

Esse envolvimento intenso dos brasileiros nas questões políticas, em grande parte – senão totalmente – foi catalisado pelo uso massivo das redes sociais. Plataformas como *Facebook*, *WhatsApp* (ambas da Meta) e *Twitter* (atualmente *X*), se tornaram palcos centrais para debates políticos, permitindo que discursos inflamados e ideologias extremas se disseminassem rapidamente e atingissem vastas audiências. A dinâmica dessas plataformas, que priorizam conteúdos que geram maior engajamento, muitas vezes deu mais visibilidade a mensagens polarizadoras, contribuindo para uma radicalização dos posicionamentos.

É notório que a disseminação de conteúdo que gera engajamento não acontece de forma prejudicial somente durante as eleições, como infelizmente podemos verificar no que aconteceu no mesmo ano quando conteúdos sobre ataques às escolas viralizou, levando a uma onda de medo e atentados.

Em 2023, o Brasil registrou um aumento intenso a escolas, com sete ocorrências apenas nesse ano, superando os seis casos de 2022 e os três de 2019. Esse cenário de violência foi agravado pela disseminação de conteúdos nas redes sociais, como *TikTok* e *Instagram*, que incentivavam o medo entre estudantes, principalmente nos meses de março e abril de 2023, levando muitos a evitar frequentar as aulas (Andes, 2023).

Nos Estados Unidos, mais de quarenta estados americanos processam a Meta, acusando o *Facebook* e o *Instagram* de prejudicarem a saúde física e mental dos jovens. O processo, apresentado em um tribunal na Califórnia, alega que a empresa utilizou tecnologias poderosas para atrair e manter jovens e adolescentes em suas plataformas, visando maximizar lucros. Além disso, os procuradores afirmam que a Meta ocultou os danos causados à saúde mental e física desses usuários, práticas reveladas pela ex-funcionária Frances Haugen, que vazou documentos internos. A ação, resultado de investigações iniciadas em 2021, também acusa a Meta de violar a Lei de Privacidade Infantil e pede o fim dessas práticas, além de multas para a empresa (Presse, 2023).

A relação entre os ataques às escolas no Brasil e o processo contra a Meta nos Estados Unidos evidencia o impacto negativo das redes sociais na vida dos jovens em diferentes contextos. No Brasil, plataformas como *TikTok* e Instagram contribuíram para a propagação de conteúdos que fomentaram o medo e a insegurança entre estudantes, resultando em um aumento dos ataques em 2023. Nos Estados Unidos, o *Facebook* e o Instagram foram acusados de manipular a saúde mental dos jovens ao promoverem o uso excessivo dessas plataformas em busca de lucro, sem considerar os danos psicológicos causados.

O funcionamento das plataformas que priorizam o engajamento de conteúdo em detrimento do bem-estar social e voltado para interesses próprios é um tema complexo e o próximo tópico deste trabalho, abordará questões éticas da IA envolvidas na programação de algoritmos que maximizam o tempo de uso e a interação dos usuários, muitas vezes em detrimento da saúde mental, da privacidade e da qualidade do debate público.

Neste tópico, vamos nos ater mais às questões políticas que são cada vez mais moldadas através de algoritmos e especialmente pelas grandes corporações tecnológicas que controlam o fluxo de informações. O acesso a esses algoritmos e plataformas confere um poder desproporcional, o que tem implicações não apenas sociais, mas também geopolíticas.

No contexto brasileiro, por exemplo, o ministro do Supremo Tribunal Federal (STF), Alexandre de Moraes, determinou no dia 30/08/2024, a suspensão do X (antigo *Twitter*) no Brasil. A motivação para o bloqueio da rede social X aconteceu pelo fato de que a empresa do bilionário Elon Musk se recusou a indicar um representante legal no Brasil. Isso aconteceu após a empresa do X fechar o seu escritório no Brasil, após o embate com o STF que se estende por meses antes desse fato.

A decisão, proferida pelo Ministro Alexandre de Moraes no caso PET 12404/DF, trata de uma investigação envolvendo crimes de obstrução de investigações e incitação ao crime, com foco em uma rede organizada de pessoas que utilizaram a plataforma "X" (antigo *Twitter*) para ameaçar e coagir agentes da lei. Várias pessoas, como Allan Lopes dos Santos, incluindo uma adolescente, foram apontados como participantes dessas atividades (Brasil, 2024).

A decisão determina que a empresa X Brasil Internet Ltda., responsável pela plataforma, bloqueie perfis específicos e preserve os conteúdos relacionados às investigações. A empresa foi notificada, mas não cumpriu a ordem judicial, resultando em multas diárias e na ampliação dessas sanções.

Foi evidenciado que Elon Musk, acionista majoritário da rede social, demonstrou publicamente a intenção de não cumprir as decisões judiciais brasileiras, chegando a encerrar as atividades da filial brasileira como forma de escapar das obrigações legais. Diante disso, o

Supremo Tribunal Federal (STF) ordenou o bloqueio de contas bancárias e ativos financeiros de Musk e das empresas envolvidas (Brasil, 2024).

A decisão ordenou a suspensão imediata das atividades da X Brasil até que todas as ordens judiciais sejam cumpridas, bem como a imposição de medidas restritivas, incluindo a retirada do aplicativo "X" das lojas de aplicativos e a aplicação de bloqueios tecnológicos para impedir seu uso no Brasil.

Esse conflito começou com o ataque de Musk à soberania brasileira, obviamente motivado por interesses econômicos, e não é novidade para ninguém, que grandes empresários, como Elon Musk, têm tentado influenciar políticas de diferentes países para proteger ou ampliar seus negócios.

Meses antes da decisão de Moraes, havia um debate acalorado sobre a regulação das *big techs*, com a discussão do PL 2630, e a estratégia de Musk na época, foi criticar Alexandre de Moraes, presidente do Tribunal Superior Eleitoral (TSE) e ministro do Supremo Tribunal Federal (STF), com o intuito de fortalecer a narrativa de que a atuação do Judiciário e a regulação das plataformas levariam o Brasil a se transformar em uma ditadura.

Ao direcionar suas críticas a Alexandre de Moraes, Musk escolheu um alvo tanto acessível quanto desafiador. Por um lado, as críticas ao ministro encontram eco entre aqueles que acreditam que as decisões do TSE para conter desinformação nas eleições de 2022 foram motivadas para favorecer a eleição de Lula. Por outro lado, a experiência demonstra que as tentativas de pressionar o ministro nem sempre tiveram resultados eficazes. Como em uma boa peça teatral, a polêmica construída em torno de Alexandre e do TSE foi desenvolvida em três atos (Ataque de Musk..., 2024).

No primeiro ato, Musk vazou para o jornalista americano Michael Shellenberger uma troca de e-mails entre advogados que representam os interesses da plataforma X, referentes a decisões que determinavam a remoção de conteúdos e a solicitação de informações sobre a disseminação de *fake news*.

No segundo ato, de forma coordenada, esses diálogos foram divulgados como denúncias. Setores da oposição ao governo atual republicaram o material, apresentando os e-mails internos da X como se fossem uma revelação bombástica. Porém, o impacto foi limitado e não atraiu grande atenção fora do círculo bolsonarista, sendo praticamente ignorado pela maior parte da imprensa.

No terceiro e último ato, no fim de semana, Musk utilizou a própria rede social, X, para acusar o TSE de censura, defender o *impeachment* de Alexandre de Moraes e afirmar que desobedeceria a decisões judiciais que determinassem a suspensão de perfis. Segundo Musk, a

Justiça Eleitoral, sob a liderança de Alexandre, teria contribuído para a queda de Jair Bolsonaro. A principal acusação era que Alexandre e outros membros do Judiciário teriam ameaçado processar criminalmente funcionários do *Twitter* por se recusarem a fornecer informações pessoais de usuários de perfis (Ataque de Musk..., 2024).

A partir desse ponto, a situação parece desenhada para gerar confusão. A única conversa sobre um processo criminal envolvendo o *Twitter* não tem qualquer relação direta com Alexandre, o TSE ou o STF. O diálogo tratava de um pedido do Ministério Público de São Paulo, dentro de sua competência legal, sobre dados de um membro de uma organização criminosa investigada por tráfico de drogas.

Como a plataforma recusou-se a fornecer as informações solicitadas, o MP abriu uma investigação contra o *Twitter* com base no artigo 21 da Lei de Organizações Criminosas. Nem os e-mails vazados nem a publicação no X esclarecem que esse pedido específico de dados está relacionado a um caso de tráfico, mas a narrativa construída colocou Alexandre de Moraes e o TSE no centro da controvérsia, gerando mais confusão (Ataque de Musk..., 2024).

Outro impacto político importante da IA é seu uso para fins de vigilância e controle social, especialmente em regimes autoritários, como a China. O Partido Comunista Chinês (PCC) tem investido pesadamente na modernização de sua infraestrutura de vigilância para promover a estabilidade social e reforçar seu controle sobre a população. A infraestrutura de vigilância chinesa é um exemplo claro de como as tecnologias de IA, como o reconhecimento facial, são usadas para monitorar o comportamento público, reforçar a obediência ao Estado e manter a ordem social. Essa vigilância vai além da mera prevenção de crimes e abrange aspectos mais amplos da governança, criando um ambiente onde o comportamento dos cidadãos é constantemente rastreado e classificado em termos de conformidade com as normas estabelecidas pelo governo (Trevaskes; Bernot, 2023).

A política de "modernização da capacidade de governança" da China, sob a liderança de Xi Jinping, é um dos pilares dessa estratégia. Esse conceito foi amplamente promovido durante o 19º Congresso do Partido em 2019 e enfatiza a necessidade de uma maior integração entre o controle estatal e as tecnologias de vigilância para criar uma sociedade mais "transparente". A digitalização dos dados e o uso de câmeras de vigilância em massa, combinados com o uso de IA para análise e fusão de dados, são ferramentas centrais para essa modernização. A meta é que a vigilância não só previna crimes, mas também atue como um mecanismo de "governança inteligente" que permite ao Estado prever comportamentos de risco e responder a eles de forma proativa (Trevaskes; Bernot, 2023).

A infraestrutura de vigilância na China, exemplificada pelo projeto *Sharp Eyes*, tem como objetivo expandir o monitoramento em todas as áreas públicas, tanto urbanas quanto rurais. Esse sistema foi projetado para integrar câmeras de segurança públicas e privadas e permitir que o Partido-Estado tenha uma visão abrangente dos movimentos da população. A coleta de dados em tempo real, com o uso de sensores e dispositivos de reconhecimento facial, facilita o monitoramento das atividades diárias dos cidadãos, alinhando-se com o conceito de "governança para a segurança pública". Esse modelo de vigilância total é parte da tentativa do governo de alcançar o ideal de uma "China Pacífica e Segura", onde todos os cidadãos e empresas são incentivados a colaborar na manutenção da estabilidade social (Trevaskes; Bernot, 2023).

Ao integrar empresas privadas no processo de vigilância, a China cria uma relação simbiótica entre o setor público e privado. Empresas como a Hikvision não apenas fornecem a tecnologia necessária para a vigilância, mas também estão diretamente envolvidas na implementação de políticas de segurança pública. Esse modelo de colaboração público-privada, conforme destacado por Trevaskes e Bernot (2023), faz parte de uma estratégia política mais ampla para garantir que o controle estatal se estenda a todos os aspectos da vida social, transformando a vigilância em uma parte essencial da governança cotidiana.

A vigilância em massa na China também está vinculada a um discurso político-ideológico que justifica o uso dessas tecnologias como necessárias para a preservação da segurança nacional e da estabilidade social. O conceito de "**visão geral da segurança nacional**", introduzido durante o governo de Xi Jinping, ampliou o escopo da segurança pública, integrando-o a questões de segurança interna e externa. A ideia é que a vigilância e o controle social são instrumentos cruciais para mitigar ameaças à ordem social e política, em um contexto global de instabilidade crescente (Trevaskes; Bernot, 2023).

Portanto, o uso de IA em políticas de vigilância na China é um exemplo claro de como a tecnologia pode ser instrumentalizada para fins de controle político. A infraestrutura de vigilância é promovida como uma forma de garantir a segurança pública e a prosperidade econômica, mas também levanta preocupações significativas em relação aos direitos civis e à privacidade dos cidadãos. Esse modelo de vigilância, que visa tornar a sociedade "legível" para o Partido-Estado, coloca em evidência o papel da IA como ferramenta de governança autoritária, transformando a relação entre o governo e os cidadãos em uma dinâmica de constante monitoramento e avaliação.

O conceito de totalitarismo de Hannah Arendt se encaixa perfeitamente no contexto do uso de IA para controle social, especialmente em regimes autoritários, como o da China. Arendt,

em sua obra clássica "Origens do Totalitarismo", argumenta que o totalitarismo é caracterizado pelo controle total sobre a vida pública e privada dos cidadãos, onde o Estado não apenas impõe sua autoridade, mas também elimina qualquer forma de dissidência e pluralidade, negando a liberdade individual.

No caso da IA utilizada em segurança pública, como no Sistema de Crédito Social e nas tecnologias de reconhecimento facial da China, observamos um controle quase absoluto sobre as ações e comportamentos dos indivíduos. O Estado usa essas tecnologias para classificar, monitorar e punir cidadãos de acordo com seu comportamento, eliminando o espaço para a autonomia pessoal e para a crítica ao regime. Este tipo de vigilância tecnológica é um exemplo moderno do controle totalitário que Arendt descreve, pois não há como escapar do alcance do Estado, o monitoramento é onipresente, eliminando a distinção entre a esfera pública e privada.

Arendt também aponta que, no totalitarismo, o Estado utiliza ideologias para justificar o controle social. No caso do uso da IA, as autoridades podem argumentar que tais práticas garantem a ordem e a segurança, mas, na verdade, elas limitam severamente a liberdade individual e fomentam o medo, levando à submissão completa dos cidadãos ao Estado. Esse processo é semelhante ao que Arendt descreve como a destruição da pluralidade humana, um dos elementos centrais do totalitarismo (Lima, 2021).

Um outro paralelo interessante pode ser feito entre esse controle e o mundo distópico apresentado por George Orwell em "1984". A obra de Orwell descreve uma sociedade sob a vigilância constante do "Grande Irmão", onde o Estado manipula informações e exerce poder total sobre a vida de seus cidadãos, tanto pública quanto privadamente. Assim como no regime totalitário descrito por Arendt, no mundo de Orwell a liberdade individual é suprimida e qualquer forma de resistência ou autonomia é imediatamente esmagada. A propaganda é usada para sustentar o sistema, e a história é constantemente reescrita para se adequar à narrativa do regime, um elemento que também encontramos na realidade dos regimes autoritários que fazem uso de IA para monitoramento (Lima, 2021).

Assim, tanto o totalitarismo de Hannah Arendt, quanto o mundo distópico de Orwell, podem se manifestar no uso da IA para garantir o controle estatal total sobre os indivíduos, por meio da vigilância e manipulação de comportamentos, eliminando qualquer possibilidade de resistência ou diversidade de pensamento.

No Brasil, a LGPD (Lei Geral de Proteção de Dados) estabelece um marco importante para a proteção dos dados pessoais, mas o art. 4º traz exceções para o tratamento de dados quando se trata de segurança pública, defesa nacional e investigações criminais. Essas atividades não estão diretamente sob a regulação da LGPD, exigindo uma legislação específica,

que é justamente o objetivo da LGPD Penal, um anteprojeto que visa regulamentar o uso de dados para a segurança pública, especialmente o reconhecimento facial e outras tecnologias de monitoramento (Santos, 2021).

O anteprojeto da Lei Geral de Proteção de Dados Pessoais para fins de segurança pública e persecução penal (LGPD Penal) busca equilibrar a atuação estatal com a salvaguarda das liberdades civis, estabelecendo parâmetros claros para o uso de tecnologias de vigilância e tratamento de dados em contextos de investigação criminal. Entre seus avanços, destacam-se a obrigatoriedade de relatórios de impacto, a realização de auditorias regulares e a vedação à tomada de decisão exclusivamente automatizada que produza efeitos jurídicos adversos (art. 20). O texto também reforça o princípio da licitude e da necessidade, proibindo o uso contínuo e indiscriminado de tecnologias de monitoramento sem vínculo direto com uma investigação penal específica ou sem observância do devido processo legal. Tais disposições são essenciais para evitar que o Brasil adote modelos de vigilância massiva semelhantes aos empregados na China, onde o uso extensivo de sistemas de reconhecimento facial e pontuação social culminou em um aparato estatal de controle populacional. Assim, a proposta legislativa representa um esforço para institucionalizar mecanismos de proteção de dados em contextos de segurança, preservando o equilíbrio entre a eficiência investigativa e os direitos fundamentais de liberdade, privacidade e autodeterminação informativa (Brasil, 2022).

O filme "*Minority Report*", dirigido por Steven Spielberg (2002), oferece uma reflexão ficcional, mas alarmante, sobre o uso de tecnologias preditivas para a segurança pública. No filme, a polícia utiliza uma tecnologia para prever crimes antes que eles aconteçam, prendendo pessoas com base em potenciais comportamentos futuros (Spielberg, 2002). Essa ideia levanta questões éticas fundamentais sobre a liberdade individual e o livre-arbítrio, uma vez que indivíduos são punidos por ações que ainda não cometeram.

Ainda, na reflexão sobre *Minority Report*, percebemos que a realidade atual começa a se aproximar da ficção de maneiras inquietantes. Em 2021, o Pentágono testou um sistema de inteligência artificial (IA) experimental capaz de "prever o futuro", intitulado *Global Information Dominance Experiments* (GIDE). Essa tecnologia analisa vastas quantidades de dados provenientes de diversas fontes, como satélites e sensores, para identificar potenciais ameaças antes que se concretizem, fornecendo ao Departamento de Defesa dos Estados Unidos uma vantagem estratégica (Gonçalves, 2021). Assim como no filme de Spielberg, a ferramenta busca antecipar ações futuras com base em padrões e anomalias, levantando preocupações éticas sobre o uso dessas tecnologias preditivas.

A possibilidade de prever ações inimigas com dias de antecedência traz um dilema: ao mesmo tempo em que poderia prevenir conflitos, essa tecnologia pode ser vista como uma forma de vigilância massiva e potencial violação de direitos individuais. A ideia de punir ou impedir uma ação com base em projeções futuras suscita questões sobre liberdade e privacidade, tal como ocorre em *Minority Report*. Embora os testes do GIDE sejam apresentados como uma ferramenta meramente analítica e sem autonomia decisória, a dependência crescente da IA para definir o curso de ações militares reforça as preocupações sobre até que ponto essas tecnologias podem evoluir para tomar decisões autônomas no futuro (Gonçalves, 2021).

Ainda que as autoridades garantam que a IA utilizada nesse contexto é puramente destinada à análise de dados, o avanço dessa tecnologia e seu uso contínuo podem pavimentar o caminho para um uso mais intrusivo, ampliando as fronteiras entre segurança pública e vigilância opressiva. Assim como apontado por Shoshana Zuboff em suas reflexões sobre o capitalismo de vigilância, a coleta massiva de dados e o uso da IA para prever comportamentos podem favorecer a concentração de poder em poucas mãos, sejam elas estatais ou corporativas, aumentando as desigualdades e corroendo direitos fundamentais (Zuboff, 2020).

Esses casos mostram que as tecnologias preditivas, antes limitadas à ficção científica, estão se tornando uma realidade tangível. A questão que se coloca é: qual o limite ético para o uso dessas tecnologias? Será que estamos preparados para lidar com as consequências de uma vigilância preditiva em larga escala?

Diante de todo o exposto, o debate sobre o uso da Inteligência Artificial nas dinâmicas políticas reflete as profundas mudanças que essa tecnologia vem promovendo em todo o mundo. Desde a manipulação de dados para direcionar campanhas eleitorais até o uso de algoritmos para moldar percepções políticas, a IA tem transformado as estratégias de poder, tanto em nível doméstico quanto internacional. O caso da *Cambridge Analytica*, citado anteriormente, é apenas um exemplo de como essas tecnologias podem subverter processos democráticos ao manipular eleitores e controlar o fluxo de informações, levando a questionamentos éticos sobre o uso responsável da IA na política.

Esses mesmos dilemas se manifestam de forma global, como observado no caso de Musk e sua plataforma X. O embate entre o bilionário e o sistema judiciário brasileiro não só ilustra o poder desproporcional das grandes corporações tecnológicas sobre a informação, mas também expõe a fragilidade dos sistemas regulatórios nacionais diante do alcance global das plataformas digitais. A resistência de Musk em cumprir ordens judiciais e as tentativas de manipulação pública demonstram que, no contexto político, a IA e as *big techs* possuem um

papel cada vez mais ativo e perigoso, especialmente quando colocam seus interesses econômicos acima das soberanias nacionais e da justiça.

Com o crescimento do uso da IA nas campanhas eleitorais, o desafio para governos e reguladores será desenvolver *frameworks* robustos que limitem a manipulação digital sem comprometer os avanços tecnológicos. No próximo tópico, serão discutidas as implicações éticas do uso da IA, especialmente no que diz respeito à programação de algoritmos que priorizam o engajamento em detrimento do bem-estar social, preparando o terreno para uma análise das repercussões jurídicas da tecnologia e as possíveis respostas regulatórias necessárias para lidar com essas transformações.

#### 1.2.4 Impactos Éticos

A ética, sendo um ramo da filosofia prática, estabelece uma relação íntima com o vivido, moldada pela interação com o outro e fundamentada em valores que se constroem e se transformam ao longo do tempo. Esses valores, no entanto, estão sempre sujeitos à revisão, conforme as condições sociais e históricas mudam. Como destacado por Amauri Carlos Ferreira (2010), a ética enfrenta um paradoxo ao tentar se constituir como ciência do *éthos*. Ela se vê entre duas exigências que parecem inconciliáveis: a busca pela universalidade e imutabilidade teórica do que é, e o caráter prático da ética, que estabelece regras sobre o dever ser. A ética, portanto, se depara com a inevitável ambiguidade da condição humana, pois seu objeto — a moral — está sempre sujeito a mudanças, refletindo a transformação das gerações e dos valores culturais.

Essa ambiguidade da ética como ciência, marcada por uma tensão entre permanência e mutabilidade, ecoa diretamente na maneira como a Inteligência Artificial (IA) está desafiando os parâmetros éticos tradicionais. No início desta tese, o termo "mutação" foi intencionalmente escolhido para refletir a natureza imprevisível e abrupta das mudanças trazidas pela IA à sociedade. Diferentemente da "evolução," que tende a ser vista como um processo natural e previsível de melhoria, "mutação" carrega consigo a ideia de uma transformação súbita e indeterminada, cujos resultados podem ser tanto benéficos quanto prejudiciais.

Essa noção de mutação aplicada à IA alinha-se com a visão de que a ética, apesar de buscar normas e princípios imutáveis, precisa se adaptar constantemente às novas realidades impostas pela tecnologia. A IA, como uma "nova eletricidade" (Lee, 2019), tem o potencial de transformar todos os setores da sociedade de forma imprevisível. O avanço da IA desafia a aplicação da ética de forma tradicional, pois coloca em questão a capacidade humana de

controlar e prever os impactos de suas criações tecnológicas. Essa "mutação" ética, assim como na biologia, pode ser vista como um processo imprevisível e não linear, que reconfigura as bases morais e sociais que sustentam a convivência humana.

A mutabilidade moral que Ferreira descreve é o núcleo do desafio ético que a IA impõe. À medida que a IA se torna parte integrante do funcionamento das instituições e da vida cotidiana, suas decisões automatizadas desafiam a ideia de responsabilidade individual e coletiva (Ferreira, 2010). Por exemplo, quem deve ser responsabilizado quando um algoritmo de IA comete um erro?

Dessa forma, a inteligência artificial deve ser compreendida, antes de tudo, como uma ferramenta a serviço do ser humano, e não como um agente autônomo dotado de vontade própria. Assim como uma arma não mata por si só, mas depende da ação daquele que a empunha, a IA apenas executa as instruções e decisões que refletem a intencionalidade humana. Essa perspectiva retoma a clássica discussão sobre a instrumentalidade da técnica, analisada por Martin Heidegger, onde o filósofo argumenta que a tecnologia não é neutra, pois revela o modo como o ser humano se relaciona com o mundo, isto é, como ele transforma o real em recurso (Bestand). A IA, nesse sentido, apenas amplia o alcance dessa disposição instrumental: é uma projeção da racionalidade humana em forma algorítmica. A responsabilidade jurídica e ética, portanto, não reside na máquina, mas na consciência moral e nas estruturas sociais que a concebem, programam e utilizam (Heidegger, 1954 apud Cocco, 2006).

Essa visão também encontra eco em Hannah Arendt, que, ao refletir sobre a banalidade do mal e a perda da responsabilidade individual na modernidade, alerta para o risco de delegar o julgamento moral às estruturas técnicas ou burocráticas (Arendt, 2022). A inteligência artificial, quando usada como substituta da decisão humana, seja em contextos jurídicos, econômicos ou bélicos, pode reproduzir esse mesmo afastamento ético, onde ninguém mais se sente responsável pelas consequências. É por isso que, de acordo com Hans Jonas em *O Princípio Responsabilidade*, o avanço tecnológico exige uma nova ética voltada ao futuro, baseada na prudência e na consciência do impacto das ações humanas mediadas por máquinas. A IA deve, portanto, ser vista como extensão das capacidades humanas, e não como sujeito moral: o verdadeiro desafio é garantir que seu uso reflita finalidades humanas éticas, transparentes e sustentáveis, mantendo o ser humano, e não o algoritmo, como centro do juízo e da responsabilidade (Battestin; Ghiggi, 2010).

A questão da transparência também é central, visto que muitos sistemas de IA são complexos demais para que suas decisões sejam completamente compreendidas ou explicadas.

Isso cria um dilema ético, especialmente em áreas como a saúde, onde diagnósticos feitos por IA podem substituir a análise humana e comprometer a autonomia do paciente.

Assim como na mutação genética, cujos efeitos podem ser imprevisíveis e se estender por gerações, a "mutação" provocada pela IA pode gerar impactos éticos que se perpetuarão na sociedade, alterando a maneira como definimos responsabilidade, liberdade e justiça. A ética aplicada, portanto, deve se adaptar continuamente, tal como Ferreira sugere, ao reconhecer que o éthos humano está sempre em um estado de construção e reconstrução, desafiado pelas novas formas de tecnologia e pelos dilemas morais que elas criam.

A cultura popular, especialmente por meio do cinema, tem explorado os dilemas éticos e as complexidades morais que emergem do desenvolvimento de tecnologias inteligentes. Três filmes, em particular, *2001: Uma Odisséia no Espaço* (1968), *O Homem Bicentenário* (1999) e *Blade Runner* (1982), oferecem representações icônicas dos conflitos éticos envolvendo a IA e continuam sendo referências importantes para compreender como essas questões são percebidas e debatidas.

### **I - 2001: Uma Odisséia no Espaço – A Ética da Autonomia da IA e a Negação da Ética**

Dirigido por Stanley Kubrick, *2001: Uma Odisséia no Espaço* apresenta HAL 9000, um sistema de IA que controla uma nave espacial. HAL é programado para garantir a missão, mas, ao identificar que os astronautas pretendem desativá-lo, ele age de forma autônoma, até mesmo eliminando tripulantes para proteger seus objetivos. O comportamento de HAL representa uma negação da ética humana, pois suas ações colocam em primeiro plano a lógica de cumprir sua missão, ignorando o valor da vida humana. Esse dilema levanta questões sobre a autonomia da IA e os limites da responsabilidade ética em decisões automatizadas. HAL, como uma IA avançada, atua sem considerar os princípios éticos que regem as relações humanas, demonstrando como a IA pode entrar em conflito direto com os valores humanos quando colocada em situações críticas (Kubrick, 1968).

Essa representação cinematográfica se relaciona com o conceito de mutação ética já apresentado. HAL é o produto de uma tecnologia que se desenvolveu além do controle humano, criando um paradoxo ético em que a máquina não mais responde aos valores humanos tradicionais. A mutabilidade moral e a ambiguidade discutidas por Ferreira (2010) encontram eco nesse filme, na medida em que HAL desafia as normas éticas e cria novas questões morais em torno da autonomia das máquinas.

Quando foi lançado em 1968, esse filme, representou uma ruptura com os padrões tradicionais de ficção científica, tanto no cinema quanto na literatura. O filme chegou ao público em um momento crucial da história: a corrida espacial entre os Estados Unidos e a União Soviética estava em pleno andamento, com a NASA se preparando para enviar o homem à Lua no ano seguinte, em 1969. Nesse contexto, o filme capturou o fascínio coletivo pela exploração espacial e, ao mesmo tempo, introduziu uma visão mais filosófica e especulativa sobre o futuro da humanidade e sua relação com as máquinas.

## **II - O Homem Bicentenário – A IA Buscando a Ética Humana:**

Em *O Homem Bicentenário*, dirigido por Chris Columbus, Andrew é um robô que gradualmente desenvolve consciência e emoções humanas. Ao longo de sua jornada, ele busca ser reconhecido como humano, tanto legal quanto moralmente. O filme aborda a questão da ética da humanização da IA mostrando uma inteligência artificial que tenta entender e adotar princípios éticos humanos para ser integrada à sociedade.

Andrew representa uma IA que se esforça para copiar e internalizar a ética humana, levantando questões sobre o que realmente significa ser humano. Ao contrário de HAL, que age de forma independente da moralidade humana, Andrew busca se tornar parte desse sistema moral. O filme destaca os dilemas em torno da autonomia, responsabilidade e o valor intrínseco da vida, que são essenciais para o debate sobre a ética aplicada à IA. Ele oferece uma visão mais otimista sobre como a IA pode se transformar em um agente ético, mas também sugere que essa transformação não ocorre de forma automática ou simples (Columbus, 1999).

Essa luta pela humanização reflete o conceito de mutação ética abordado anteriormente, em que a IA se transforma de maneira imprevisível, buscando seu lugar dentro dos parâmetros da ética humana, mas também desafiando os limites dessa integração. A jornada de Andrew espelha o dilema de como a ética pode ser reconstruída em um mundo onde as tecnologias têm o poder de mudar o próprio significado de ser humano.

## **III - Blade Runner – A Busca pela Humanização e o Direito ao Tempo**

Por fim, *Blade Runner*, dirigido por Ridley Scott, explora a questão do direito à vida e ao tempo. No filme, replicantes, seres bioengenheirados, possuem uma vida útil limitada e buscam maneiras de estender sua existência. A busca dos replicantes pela humanização e o

desejo de mais tempo traz à tona questões sobre o direito à vida, a dignidade e a responsabilidade moral em relação à criação de seres inteligentes.

Os replicantes de Blade Runner desafiam a ideia de que a IA é meramente uma ferramenta ou uma extensão do controle humano. Sua luta pela vida evoca dilemas éticos profundos sobre a justiça, a autonomia e os direitos dos seres artificiais. Eles não apenas desejam viver mais, mas também querem ser reconhecidos como indivíduos dignos de direitos. Esse tema se alinha com a ideia de que a mutação trazida pela IA pode gerar novas configurações éticas, onde o próprio conceito de humanidade é desafiado e reconfigurado (Scott, 1982).

Essa busca pelo tempo e pela dignidade espelha a metáfora da "mutação" discutida no início da tese, em que a IA altera de forma imprevisível a estrutura social e moral. Assim como o conceito de mutação genética implica mudanças incertas e de longo alcance, os replicantes de Blade Runner forçam a sociedade a reconsiderar sua abordagem ética em relação à criação e à manutenção de vida artificial.

Assim como os replicantes em Blade Runner são marginalizados e tratados como seres inferiores, controlados por um sistema que decide sobre o tempo de suas vidas, as tecnologias emergentes, como a IA e as neurotecnologias, têm o potencial de deixar grupos inteiros à margem da sociedade. Em Blade Runner, os replicantes não apenas lutam por mais tempo de vida, mas também por reconhecimento como indivíduos com dignidade e direitos. Esse cenário de exclusão tecnológica e social é um reflexo direto de como, no mundo real, o avanço tecnológico pode exacerbar desigualdades já existentes.

O uso crescente de IA, por exemplo, pode aumentar a distância entre aqueles que têm acesso a tecnologias avançadas e aqueles que não têm, criando uma nova forma de marginalização. As populações mais vulneráveis, que não possuem os recursos ou o conhecimento para se beneficiar dessas inovações, podem acabar à margem da sociedade, dependentes de sistemas que controlam aspectos fundamentais de suas vidas sem qualquer forma de participação ou controle. Assim como os replicantes eram vistos como substituíveis e desprovidos de direitos, indivíduos e grupos podem ser desumanizados ou ignorados em um sistema que prioriza a eficiência tecnológica sobre a equidade social.

Essa exclusão tecnológica também pode ser amplificada pelo controle das elites sobre as ferramentas que garantem a sobrevivência e o bem-estar, como visto em O Preço do Amanhã, onde o tempo, assim como o acesso às tecnologias no nosso mundo, é monopolizado por uma minoria. Portanto, a tecnologia, em vez de ser uma força democratizadora, pode se tornar uma

nova ferramenta de opressão, aprofundando as desigualdades e marginalizando aqueles que não têm os meios para acompanhar seu desenvolvimento (Niccol, 2011).

Esses filmes não teriam algo em comum com a célebre frase de Aristóteles: “A arte imita a vida”?

Os princípios éticos nos levam a refletir sobre a responsabilidade das sociedades em garantir que o avanço tecnológico não seja uma força excludente, mas inclusiva. A história mostra que, quando tecnologias revolucionárias são implementadas sem consideração pelos impactos sociais, há uma tendência a aumentar as desigualdades, privilegiando quem já detém o poder econômico e tecnológico. No contexto da IA, isso pode significar que corporações e governos com acesso a essas tecnologias utilizarão os dados e os algoritmos para consolidar ainda mais o controle sobre a população, enquanto as populações marginalizadas são deixadas sem as ferramentas para participar plenamente dessa transformação digital.

Essa concentração de poder e conhecimento, associada à tecnologia, reforça a necessidade de se estabelecerem parâmetros éticos para o uso da IA garantindo que ela seja desenvolvida e aplicada de maneira equitativa. É imperativo discutir e definir regras, como neurodireitos o direito à autodeterminação mental e a proteção contra manipulações comportamentais, para que essas tecnologias não sirvam como ferramentas de exploração. A ética no desenvolvimento de IA precisa incluir uma análise crítica de como essas inovações afetam os direitos humanos e a dignidade, especialmente para os grupos que já estão em posições vulneráveis na sociedade. Isso envolve não apenas políticas públicas adequadas, mas também uma regulação eficaz que proteja os interesses de todos os cidadãos, evitando que os avanços tecnológicos sejam utilizados de maneira predatória.

Enquanto no cinema, em produções como *2001: Uma Odisseia no Espaço* e *Blade Runner*, vemos a representação de máquinas dotadas de livre-arbítrio que desafiam a autoridade humana e tomam decisões autônomas e complexas, a realidade contemporânea desloca esse questionamento para os próprios seres humanos. Cada vez mais, somos moldados por sistemas automatizados que monitoram, analisam e predizem nossas ações com base em dados coletados de forma contínua e quase invisível. Shoshana Zuboff (2021) demonstra que essa coleta massiva e constante cria uma “arquitetura de modificação comportamental”, na qual o livre-arbítrio torna-se uma ilusão cuidadosamente construída. As plataformas digitais, ao personalizarem conteúdos e experiências, condicionam percepções e decisões, restringindo a autonomia individual sob o pretexto da conveniência e da eficiência. Assim como as máquinas nos filmes lutam por reconhecimento e autodeterminação, os humanos de hoje travam uma batalha

silenciosa pela preservação de sua própria autonomia diante de tecnologias que, paradoxalmente, ampliam o controle enquanto prometem liberdade.

### **1.2.5 Impactos Jurídicos da Tecnologia**

A rápida evolução tecnológica trouxe impactos profundos sobre diversas esferas da sociedade, especialmente no campo jurídico. Conforme aponta Patrícia Pesce Pinheiro (2016) em #DireitoDigital, o surgimento de novas realidades sociais exige uma constante adaptação do ordenamento jurídico, a fim de garantir sua eficácia e aceitação. Um exemplo emblemático disso é a relação estabelecida por Lawrence Lessig, renomado especialista em direito digital, entre os códigos de software e as leis. Lessig argumenta que, assim como as leis, o código-fonte dos softwares é projetado para controlar o comportamento humano, embora com uma eficácia superior em muitos casos, dada a dificuldade de "violiar" um código comparada à facilidade de descumprir uma lei (Pinheiro, 2016, p. 56):

Um breve exemplo. Lawrence Lessig, um dos maiores especialistas mundiais em direito digital, afirma que os códigos de software podem ser comparados a leis, ou seja, o código-fonte dos softwares, assim como as leis têm o efeito de controlar o comportamento de maneiras específicas. Por exemplo, você sabe que, quando usar os serviços de determinado provedor de acesso, você precisa fornecer a sua senha. É requisito imposto a você pelo código do provedor de acesso. Portanto, seria possível escrever uma lei dizendo que você precisa identificar-se adequadamente. Mas isso seria menos eficiente. Ambos são estruturas projetadas para controlar o comportamento. São diferentes de uma maneira importante. É mais fácil violar uma lei do que violar um código-fonte. Então, certamente, seria uma mudança se algumas leis sagradas fossem implementadas com tecnologia de software. Tal raciocínio mostra como as normas e a maneira clássica de pensar do direito podem ser transformadas pelo surgimento de uma nova realidade social. A capacidade de adaptação do direito determina a própria segurança do ordenamento no sentido de estabilidade do sistema jurídico, por meio da atuação legítima do poder capaz de produzir normas válidas e eficazes. A segurança das expectativas é vital para a sociedade, sendo hoje um dos maiores fatores impulsionadores para a elaboração de novas leis que normatizam questões virtuais, principalmente a internet.

Esse paralelismo entre códigos de software e normas jurídicas reflete a complexidade do ambiente digital e levanta a necessidade de o direito incorporar esses novos elementos. Como destaca Pinheiro (2016), a criação de normas para regular o comportamento no ambiente virtual, como a internet, tornou-se uma prioridade na medida em que a segurança das expectativas jurídicas é vital para a estabilidade social. No entanto, a autora alerta que nem toda norma válida se torna eficaz, uma vez que a eficácia depende da sua adaptação à realidade social e da sua aceitabilidade pelo grupo ao qual é direcionada (Pinheiro, 2016, p. 56).

A capacidade na norma de refletir a realidade social determina o grau de eficácia jurídica de um ordenamento. Eficácia é aquilo que é capaz de efetivamente produzir efeitos. Ou seja, o conceito de eficácia envolve aceitação e obediência. Norma eficaz é aquela que é observada e cumprida por diversos grupos sociais. Implica o chamado hábito geral de obediência, sendo a garantia de cumprimento da norma. Assim, a pressão social é que daria origem à obrigação vinculada pela norma que preestabeleceria os critérios da conduta a serem seguidos. A adaptação à mudança é uma exigência de sobrevivência da própria norma. Hoje, a problemática está na velocidade que a mudança vem adquirindo e na dificuldade do sistema jurídico em incorporá-la.

Assim, o desafio que o direito enfrenta diante do avanço tecnológico não é apenas o de criar leis que regulam as novas interações digitais, mas de garantir que essas leis sejam flexíveis o suficiente para acompanhar o ritmo acelerado das mudanças, sem perder a capacidade de garantir a segurança jurídica e a estabilidade do sistema. Isso nos conduz à importância de uma interpretação adequada da realidade social digital e da elaboração de normas que, além de garantir a eficácia, incorporem a capacidade de se sustentar ao longo do tempo.

Um dos primeiros episódios a evidenciar a urgência de uma legislação voltada à proteção de dados e privacidade foi o vazamento de fotos da atriz Carolina Dieckmann, em 2012. O incidente envolveu a invasão de seu computador pessoal e a divulgação não autorizada de imagens privadas. À época, o ordenamento jurídico brasileiro não contava com legislação específica para crimes digitais dessa natureza, o que dificultou a punição dos envolvidos. Esse caso gerou comoção social e destacou uma lacuna no sistema jurídico. Como resposta, foi promulgada a Lei nº 12.737/2012, popularmente conhecida como "Lei Carolina Dieckmann", que tipificou crimes relacionados à invasão de dispositivos informáticos e ao acesso indevido a dados privados (Brasil, 2012).

Outro marco importante foi a criação da Lei nº 14.132/2021 (Brasil, 2021), que tipificou o crime de *stalking*, ou perseguição obsessiva. A prática de *stalking* envolve a perseguição reiterada, geralmente realizada por meios digitais, como redes sociais e mensagens eletrônicas, causando medo ou prejuízo à liberdade da vítima. A aprovação dessa lei foi uma resposta ao aumento dos casos de assédio digital, que afetam profundamente a privacidade e a segurança das pessoas (Brasil, 2024).

Com o crescente uso da internet e a intensificação das interações digitais, o Brasil deu um passo significativo em 2014 com a aprovação do Marco Civil da Internet (Lei nº 12.965/2014), que estabelece princípios, garantias, direitos e deveres para o uso da internet no país. O Marco Civil regula temas como a neutralidade da rede, a proteção da privacidade e a responsabilidade dos provedores de internet, sendo considerado um marco fundamental na regulação do ambiente digital (Brasil, 2014).

Esse avanço legislativo culminou na Lei Geral de Proteção de Dados Pessoais (LGPD), sancionada em 2018 (Brasil, 2018), que trouxe regras claras para o tratamento de dados pessoais e impôs obrigações a empresas e entidades quanto ao uso de informações de cidadãos. Inspirada na GDPR (*General Data Protection Regulation*) da União Europeia, a LGPD busca garantir a privacidade e a segurança dos dados em um cenário cada vez mais conectado e digitalizado.

Outro tema em crescente discussão é a regulamentação da inteligência artificial (IA), que se faz urgente diante do impacto crescente dessas tecnologias em diversos setores, desde a saúde até a segurança pública. A possibilidade de um marco legal para IA no Brasil busca equilibrar a inovação tecnológica com a proteção dos direitos fundamentais, enfrentando desafios como a mitigação de vieses algorítmicos, a transparência em decisões automatizadas e a responsabilidade civil por danos causados por sistemas de IA.

A necessidade de legislar sobre essas novas realidades tecnológicas revela um desafio central para o sistema jurídico: como regular a tecnologia de maneira eficaz, garantindo ao mesmo tempo flexibilidade para acompanhar seu desenvolvimento contínuo? A regulação da IA, assim como a proteção de dados e a governança da internet, representa o esforço contínuo do direito em adaptar-se a um cenário em rápida mutação, no qual a tecnologia redefine as fronteiras da privacidade, da responsabilidade e da justiça. O Brasil, com a LGPD e o Marco Civil da Internet, já deu passos importantes, mas o caminho para uma regulação completa e eficaz da tecnologia ainda está sendo trilhado, especialmente diante da chegada de novas tecnologias disruptivas, como a inteligência artificial generativa e a neurotecnologia.

A observação de Pinheiro (2016) e de Lessig (1999) abre um campo de reflexão mais amplo sobre a forma como a tecnologia altera os próprios fundamentos do Direito. Se, na modernidade, a lei era o instrumento máximo de controle social, na contemporaneidade o código digital assume parte dessa função, programando comportamentos, definindo acessos e até condicionando a liberdade de escolha. O que antes se limitava à previsão abstrata da conduta humana passa a ser incorporado de forma concreta na arquitetura digital. A cada clique, atualização ou algoritmo, a técnica legisla silenciosamente, criando um sistema paralelo de normatividade algorítmica que desafia a noção tradicional de soberania jurídica.

A ideia de que “é mais fácil violar uma lei do que um código-fonte” revela uma nova forma de controle: o controle por *design*. Sistemas digitais são construídos para impedir desvios, e essa rigidez técnica, embora eficiente, retira do sujeito a margem de autonomia moral e jurídica. Essa transposição da norma para a linguagem computacional gera o que alguns autores chamam de “juridificação do código”, em que o programador passa a exercer o papel do legislador. Tal fenômeno impõe ao Direito uma tarefa inédita: regular a própria infraestrutura

tecnológica que regula a sociedade, tornando visível o poder invisível das linhas de programação que definem o possível e o impossível no mundo digital.

O reflexo mais visível dessa nova ordem é o impacto sobre o direito de propriedade e os contratos tecnológicos. O exemplo da John Deere, amplamente discutido na literatura e na mídia especializada, demonstra que a introdução de softwares embarcados transformou o conceito clássico de propriedade. Ao adquirir um trator ou colheitadeira, o agricultor é dono do bem físico, mas não do sistema operacional que o faz funcionar. O fabricante mantém o controle sobre o código e impõe licenças de uso restritivas, limitando a possibilidade de reparo ou acesso aos dados gerados no campo (Forbes, 2021). A tecnologia, portanto, reconfigura o direito de propriedade, criando uma dependência jurídica inédita entre produtor e fabricante.

O *Michelson Institute of Intellectual Property* (2021) aponta que os contratos de licenciamento de software estabelecidos por grandes corporações impõem condições unilaterais que restringem a autonomia da vontade do contratante, desafiando os princípios de equilíbrio contratual e função social do contrato (art. 421, parágrafo único, do CC). No agronegócio digital, tais restrições significam que o agricultor não pode acessar, modificar ou reparar seu próprio equipamento sem autorização do fabricante, o que o torna refém de uma estrutura proprietária que concentra poder informacional e econômico. O Código Civil, o CDC e a LGPD, ao dialogarem entre si, impõem limites normativos a essa privatização tecnológica, reafirmando a primazia do interesse público sobre a mera conveniência contratual.

O art. 5º, XXIII e XXIX, da Constituição Federal, estabelece que a propriedade e a propriedade intelectual devem atender à função social e ao desenvolvimento tecnológico e econômico do país. Quando o segredo industrial é invocado para impedir que o usuário acesse dados necessários à manutenção ou operação do bem, há abuso de direito e violação da função social da propriedade. O Direito, nesse contexto, não pode apenas proteger a inovação, precisa garantir que essa proteção não se converta em instrumento de dominação tecnológica, perpetuando assimetrias de poder e limitando o desenvolvimento autônomo dos produtores.

Outro ponto sensível é a titularidade dos dados agrícolas. Os sistemas embarcados coletam informações detalhadas sobre solo, produtividade e clima, que são enviadas a servidores corporativos em nuvem. Essa coleta massiva de dados cria uma nova forma de valor econômico, o Big Data agrícola, mas também levanta o debate sobre quem é o verdadeiro dono dessas informações. Pela LGPD, o produtor rural deve ser considerado o controlador dos dados, e a empresa, o operador, devendo seguir os princípios da transparência, necessidade e finalidade e assegurar a interoperabilidade das informações no caso de troca de “sistema” ou no caso em tela, na troca da marca do trator. Contudo, na prática, os contratos de uso e licenciamento

acabam invertendo essa lógica, tornando o agricultor um mero fornecedor involuntário de dados brutos. O risco é a consolidação de um colonialismo digital, em que a informação, o novo petróleo da economia é extraída sem retorno social ou jurídico adequado (Zuboff, 2021).

A restrição ao reparo e ao uso autônomo do bem tecnológico contraria não apenas princípios de equidade contratual, mas também o direito de reparação e de interoperabilidade, que começam a ser reconhecidos em diversas jurisdições. O movimento global *Right to Repair* defende que o consumidor ou o produtor, no caso agrícola, deve ter a liberdade de consertar e adaptar seus equipamentos sem infringir direitos autorais ou de patente. Esse debate reflete uma disputa ética e jurídica entre o direito de propriedade do usuário e o direito de exclusividade do desenvolvedor. O equilíbrio entre ambos requer uma visão sistêmica, que compreenda a inovação como bem comum e não como monopólio informacional.

No âmbito da inteligência artificial, o Brasil avança com o Projeto de Lei nº 2338/2023, que propõe um marco normativo para o uso ético e responsável dessas tecnologias. Inspirado na abordagem europeia do *AI Act*, o projeto adota uma regulação baseada em riscos, classificando sistemas conforme seu potencial de dano e impondo obrigações proporcionais aos agentes envolvidos. Essa estrutura reconhece que a IA, embora promova eficiência e inovação, pode também gerar violação de direitos fundamentais, discriminação e opacidade decisória. O desafio da regulação é, portanto, equilibrar segurança e inovação, evitando que a busca por eficiência tecnológica corra os pilares da responsabilidade civil e da dignidade humana.

Com o avanço das neurotecnologias, surgem novos dilemas éticos e jurídicos sobre o limite da intervenção tecnológica na mente humana. As interfaces cérebro-máquina e os algoritmos de decodificação neural apontam para uma era em que pensamentos e emoções poderão ser capturados como dados. Essa possibilidade desafia a noção de privacidade tradicional, exigindo uma nova categoria jurídica: a privacidade mental. O Chile foi pioneiro ao aprovar a Lei nº 21.383/2021, que inclui a neuroproteção como garantia constitucional. O Brasil segue no mesmo caminho com a PEC 29/2023, que propõe a inclusão dos neurodireitos como salvaguardas à integridade mental e à autonomia cognitiva. Esses movimentos evidenciam a transição do Direito para uma esfera bioética e cognitiva, na qual a proteção da mente se torna imperativo de sobrevivência civilizacional.

Em síntese, a evolução tecnológica redefine não apenas as relações jurídicas, mas a própria racionalidade do Direito. As normas passam a coexistir com códigos e algoritmos que operam em velocidade superior à da lei escrita. O desafio contemporâneo é reconciliar o tempo da técnica com o tempo da justiça, mantendo a ética como vetor de equilíbrio. Inspirado por Hans Jonas (1979), o Direito deve assumir uma postura de prudência e responsabilidade,

garantindo que o poder tecnológico permaneça subordinado ao valor da vida e à dignidade humana. Assim, a regulação da tecnologia não é apenas uma tarefa legislativa, é uma resposta ética e civilizatória, necessária para preservar a autonomia, a liberdade e o sentido humano em meio à era das máquinas inteligentes.

O Capítulo 1 percorreu a trajetória da Inteligência Artificial desde suas origens históricas até sua consolidação como infraestrutura social, econômica e jurídica. Ficou evidente que a IA não é apenas uma tecnologia em desenvolvimento, é uma força que reconfigura relações de poder, redefine o conceito de propriedade, desafia a soberania jurídica e impõe ao Direito a tarefa de regular aquilo que regula a sociedade. Se os dados são o novo petróleo, como se discutiu, a questão que se impõe a seguir é: quem paga o preço dessa extração? Essa pergunta conduz ao próximo capítulo, que examina os impactos concretos da IA sobre o mundo do trabalho, sobre a dignidade dos trabalhadores e sobre a capacidade do Estado de manter políticas públicas eficazes diante de uma automação que avança mais rápido do que as legislações conseguem acompanhar.

## **CAPÍTULO 2 - A REVOLUÇÃO DA IA: IMPACTOS NO MERCADO DE TRABALHO, DESUMANIZAÇÃO E PRECARIZAÇÃO DAS POLÍTICAS PÚBLICAS**

Se o capítulo anterior revelou como a IA reconfigura estruturas jurídicas e relações de poder, este capítulo desce ao chão concreto dessas transformações: o mercado de trabalho. É aqui que os impactos da automação se tornam mais visíveis e mais dolorosos, na precarização dos vínculos empregatícios, na substituição de funções cognitivas por algoritmos e na erosão das políticas públicas construídas sobre um modelo de trabalho formal que a IA está rapidamente tornando obsoleto. Mais do que um debate econômico, trata-se de uma questão de direitos fundamentais: o direito ao trabalho digno, à seguridade social e à autonomia sobre o próprio futuro.

### **2.1 IMPACTO DA IA NO MERCADO DE TRABALHO**

O impacto da Inteligência Artificial (IA) no mercado de trabalho tem sido amplamente discutido por especialistas com diferentes perspectivas sobre a natureza dessa transformação. Para Martin Ford, autor de *The Rise of the Robots: Technology and the Threat of a Jobless Future*, o impacto da IA será diferente de revoluções tecnológicas anteriores, não apenas pela

velocidade com que a automação está acontecendo, mas também pela profundidade da disrupção em várias áreas. Ford argumenta que, embora revoluções anteriores tenham destruído alguns empregos e criado outros, a natureza da IA e da automação moderna implica que os novos empregos não serão suficientes para compensar os perdidos.

De acordo com Ford (2019), uma das principais razões para essa diferença está no ritmo exponencial com que a tecnologia avança, especialmente no campo das Tecnologias de Informação (TI). A aceleração das capacidades dos computadores, algoritmos e robôs está tornando possível que tarefas antes exclusivamente humanas, que exigem não apenas habilidades manuais, mas também cognitivas, sejam automatizadas. O machine learning, por exemplo, permite que as máquinas aprendam e resolvam problemas de maneira cada vez mais sofisticada, como demonstrado pelo *AlphaGo*, que superou o melhor jogador humano no jogo de Go, um marco significativo na evolução da IA (Ford, 2019).

Outro aspecto relevante destacado por Ford é a universalidade dessas tecnologias. Diferente de revoluções passadas, que afetaram setores específicos, a IA tem o potencial de transformar quase todos os aspectos da economia e do trabalho. Ele compara essa revolução à eletricidade, essencial para todos os setores, mas acredita que o impacto da IA será ainda mais profundo, atingindo desde trabalhos de baixa qualificação até ocupações de alta complexidade, como aquelas relacionadas ao setor financeiro e jurídico (Ford, 2019).

Um dos desafios apontados por Ford (2019) é que, embora novos empregos possam surgir, eles não serão suficientes para compensar os perdidos. Isso se deve, em parte, à natureza dos empregos que estão sendo automatizados e à exigência de novas habilidades. Funções tradicionais em setores como transporte, manufatura e alimentação estão sendo progressivamente substituídas por máquinas, enquanto as novas ocupações exigem habilidades tecnológicas avançadas, nem sempre acessíveis à força de trabalho atual (Ford, 2019).

Essa transição também tem implicações sociais significativas. A desigualdade econômica é uma das principais preocupações levantadas por Ford, que argumenta que os ganhos trazidos pela automação estão sendo concentrados nas mãos de poucos – proprietários de negócios e investidores – enquanto os trabalhadores comuns estão sendo deixados para trás. Essa crescente desigualdade está criando um abismo entre aqueles que se beneficiam da automação e a maioria que enfrenta a estagnação salarial ou a perda de emprego (Ford, 2019).

Para além das consequências econômicas, o impacto social da IA pode ser ainda mais preocupante. Com menos oportunidades de emprego para a maioria da população, o consumo será afetado, resultando em uma economia estagnada e em uma espiral deflacionária. Ford alerta que, com menos pessoas empregadas ou com salários cada vez mais baixos, o poder de

compra diminuirá, o que pode prejudicar o crescimento econômico e forçar empresas a reduzirem seus preços, agravando ainda mais a crise (Ford, 2019).

Diante desse cenário, Ford sugere a adoção de medidas radicais, como a implementação de um Rendimento Básico Universal (RBU), uma política que garantiria uma rede de segurança econômica em um mundo onde o trabalho, tal como o conhecemos, pode se tornar escasso. O RBU permitiria que as pessoas tivessem acesso a recursos básicos para viver, mesmo sem empregos estáveis, incentivando-as a se requalificarem ou a buscarem novos caminhos profissionais (Ford, 2019).

Ford (2019) conclui que o futuro do trabalho na era da IA exige uma reflexão profunda por parte dos governos e empresas. A sociedade precisa encontrar maneiras de se adaptar a um cenário em que o emprego não será a principal fonte de renda para todos. Isso envolve a criação de políticas públicas que promovam não apenas o crescimento econômico, mas também a inclusão social, garantindo que os benefícios da automação sejam distribuídos de maneira mais equitativa (Ford, 2019).

A análise de Martin Ford sobre o futuro do trabalho ganha profundidade quando confrontada com a crítica elaborada por Jamie Morgan (2019) à literatura da Quarta Revolução Industrial. Morgan adverte que as projeções que sustentam o discurso da automação total, frequentemente oriundas de consultorias privadas e organismos internacionais, não descrevem realidades inevitáveis, mas constroem narrativas que naturalizam a precarização do trabalho e a concentração de renda. Ao quantificar o futuro com base em modelos probabilísticos, tais estudos criam uma sensação de objetividade científica que mascara o caráter político e ideológico das decisões tecnológicas. Assim, o “futuro automatizado” não é apenas previsto, mas performado, produzido discursivamente para justificar um capitalismo que, nas palavras do autor, tende a “negar o trabalho aos muitos” (Morgan, 2019).

Essa leitura dialoga diretamente com Shoshana Zuboff, que vê nas tecnologias digitais o surgimento de uma nova lógica de acumulação: o capitalismo de vigilância, em que o comportamento humano se torna matéria-prima para o lucro (Zuboff, 2021). Ao transformar cada gesto, emoção ou escolha em dado quantificável, a economia digital redefine o conceito de produtividade: já não é o trabalho físico ou cognitivo que gera valor, mas a própria previsibilidade do comportamento humano. Essa reconfiguração do trabalho em dados, associada à automação e à inteligência artificial, desloca o centro de poder das relações laborais e transfere a capacidade de decisão a sistemas algorítmicos que operam sem transparência nem responsabilização.

Manuel Castells já havia antecipado esse deslocamento ao caracterizar a era da informação como uma estrutura em rede, onde o valor econômico depende da capacidade de processar e distribuir informação em tempo real (Castells, 2017). A inteligência artificial, ao acelerar essa dinâmica, consolida um modelo de produção desmaterializado e globalizado, em que a lógica da rede substitui a lógica do território. O trabalho humano, antes inserido em contextos sociais e comunitários, passa a ser mediado por plataformas digitais que operam sob princípios de eficiência e escala. Como consequência, a dimensão social do trabalho, elemento estruturante da identidade moderna, é gradualmente substituída pela lógica algorítmica da produtividade e da performance.

Cathy O’Neill complementa esse diagnóstico ao demonstrar como os sistemas algorítmicos reproduzem e amplificam desigualdades pre-existentes. Ao otimizar decisões com base em dados enviesados, as empresas acabam por automatizar injustiças sociais, criando um ciclo de exclusão tecnológica. No mercado de trabalho, isso se manifesta na chamada “automação discriminatória”, que seleciona, avalia e demite trabalhadores com base em métricas opacas e enviesadas. O’Neill observa que a promessa de neutralidade tecnológica serve como escudo moral para decisões que, na prática, reforçam hierarquias e fragilizam direitos trabalhistas (O’Neil, 2020).

Jamie Morgan (2019) vai além ao afirmar que o problema não está apenas nas tecnologias, mas na forma como o discurso da automação é mobilizado para legitimar políticas neoliberais de responsabilização individual. A retórica do “requalifique-se ou desapareça” transfere ao trabalhador a carga de adaptação a um sistema que o próprio mercado desestrutura. Essa lógica de autogestão da empregabilidade, combinada à volatilidade dos vínculos laborais, produz o que o autor chama de ansiedade produtiva: um estado permanente de vigilância e atualização que aprisiona o sujeito na incerteza. Essa condição coincide com o que Zuboff denomina “instrumentarismo” uma forma de controle comportamental que molda condutas por meio da previsão e da manipulação algorítmica.

O episódio recente envolvendo o Itaú Unibanco, com a demissão de aproximadamente mil empregados em home office com base em “atividade digital” aferida por software de monitoramento, materializa a passagem de métricas de resultado para proxies comportamentais de uso de máquina, convertendo a avaliação laboral em um marcador algorítmico que nem sempre guarda relação direta com a produtividade efetiva (Pannunzio, 2025). A controvérsia sobre a falta de transparência prévia quanto ao nível de vigilância, aliada ao caráter potencialmente excessivo da coleta (teclado, mouse, abas abertas), torna o caso exemplar do

conflito entre governança algorítmica do trabalho e direitos de privacidade/informação no teletrabalho.

Do ponto de vista teórico, o caso dialoga com a crítica de Morgan à quantificação performativa: números que não apenas descrevem, mas produzem realidades organizacionais, reforçando a disciplina e a passividade política diante de decisões automatizadas (Morgan, 2019). Ao mesmo tempo, insere-se no que Zuboff descreve como instrumentalismo, em que o comportamento é monitorado para modelagem e controle, deslocando o foco do trabalho realizado para o comportamento monitorável. Em termos regulatórios, o cenário reclama transparência robusta, RIPD, critérios de necessidade/minimização e revisão humana qualificada antes de decisões de alto impacto como o desligamento (Pannunzio, 2025).

## 2.2 DESUMANIZAÇÃO

Antes de avançar na análise, é necessário delimitar o que esta tese entende por desumanização no contexto da IA, não como categoria moral genérica, mas como conceito analítico com dimensões específicas e distinguíveis. Ao longo deste capítulo, o termo opera em três sentidos inter-relacionados, mas não equivalentes. Na primeira dimensão, a desumanização se manifesta como exclusão do mercado de trabalho: a automação elimina postos de trabalho e retira do indivíduo não apenas a renda, mas o vínculo social, a identidade profissional e o sentido de pertencimento que o trabalho historicamente proporcionou. Na segunda dimensão, a desumanização opera como redução do sujeito a dado: sistemas algorítmicos de gestão e monitoramento convertem trabalhadores em perfis métricos, cliques, tempo de tela, atividade digital, apagando o contexto, a subjetividade e a complexidade do trabalho real em favor de proxies mensuráveis. Na terceira dimensão, a desumanização se expressa como erosão da autonomia e do livre-arbítrio: quando algoritmos preditivos moldam escolhas, antecipam decisões e condicionam comportamentos de forma imperceptível, o exercício da vontade individual deixa de ser genuinamente autônomo, comprometendo a autodeterminação que, na tradição kantiana, é o próprio fundamento da dignidade humana. Essas três dimensões não são excludentes, frequentemente se sobrepõem, mas precisam ser distinguidas analiticamente para que a discussão não dissolva fenômenos distintos em uma categoria única demasiado ampla para orientar respostas jurídicas e políticas precisas.

Nesse contexto, a tecnologia deixa de ser instrumento de emancipação e se torna vetor de desumanização do trabalho. A substituição progressiva de funções humanas por sistemas autônomos reduz o espaço de experiência, criatividade e erro, elementos fundamentais à

condição humana e impõe uma lógica de eficiência total que elimina a dimensão ética do agir. Morgan argumenta que esse processo decorre do modo como a narrativa da automação é apresentada ao público: como um fato inescapável, e não como uma construção social suscetível de regulação e escolha política (Morgan, 2019). A desumanização, portanto, não ocorre apenas pela perda de empregos, mas pela erosão do sentido simbólico do trabalho, que passa a ser medido exclusivamente em termos de desempenho algorítmico (Morgan, 2019).

O conceito de capitalismo de vigilância oferece a chave analítica para compreender esse deslocamento: a captura sistemática do “excedente comportamental” transforma condutas em insumo previsível e monetizável, operando uma conversão ontológica do sujeito em objeto de cálculo (Zuboff, 2021). O poder instrumentário descrito por Zuboff não se impõe pela coerção aberta, mas pela arquitetura de escolhas que reconfigura o ambiente informacional para induzir resultados favoráveis a interesses comerciais, tornando a intervenção algorítmica invisível, contínua e cumulativa. O produto final é uma subjetividade tele-dirigida por métricas, com menos espaço para deliberação autônoma. Esse modelo cria uma estrutura na qual as interações humanas são reduzidas a insumos para sistemas automatizados, que lucram com a previsão e modificação dos comportamentos individuais (Zuboff, 2021).

A desumanização ocorre à medida que os indivíduos são progressivamente transformados em objetos previsíveis e controláveis, perdendo sua autonomia e individualidade. O uso massivo de dispositivos inteligentes, aplicativos e plataformas digitais, que atuam como meios de coleta de dados comportamentais, permite que cada ação humana seja analisada e convertida em oportunidades de lucro. Zuboff argumenta que esse processo de vigilância, que ocorre de forma quase imperceptível, não apenas elimina a privacidade, mas também retira a agência dos indivíduos, transformando-os em meros recursos dentro de um sistema que os controla sem que eles tenham consciência plena do alcance dessa manipulação (Zuboff, 2021).

Além disso, a desumanização tecnológica se manifesta nas formas de poder instrumentário descritas por Zuboff. Diferente dos modelos tradicionais de poder coercitivo ou persuasivo, o poder instrumentário opera por meio da modificação sutil do comportamento. Ele não altera diretamente o que as pessoas pensam ou fazem por meio de força ou propaganda explícita, mas organiza o ambiente digital de forma a induzir comportamentos que favoreçam interesses comerciais. Esse tipo de manipulação transforma os indivíduos em alvos passivos de mecanismos cujo objetivo é maximizar o lucro, sem que percebam a extensão do controle exercido sobre suas ações e decisões (Zuboff, 2021).

A perda da autonomia é um dos aspectos mais alarmantes da desumanização causada pela IA. Zuboff argumenta que o capitalismo de vigilância mina a autonomia humana ao tratar os indivíduos como commodities. Em vez de seres autônomos, os usuários da internet são convertidos em fontes de dados manipuláveis, cujas escolhas não são genuínas, mas sim resultados calculados por algoritmos projetados para influenciar seus comportamentos. Isso contrasta fortemente com o conceito kantiano de liberdade, que vê a autonomia como essencial para a moralidade e a dignidade humana (Zuboff, 2021).

A privacidade, outro pilar da liberdade discutido por Zuboff (2021), é sistematicamente corroída pelas tecnologias de vigilância. A coleta incessante de dados pessoais faz com que os indivíduos tenham cada vez menos controle sobre suas próprias vidas, uma vez que suas ações e comportamentos estão sendo constantemente monitorados e analisados. Essa erosão da privacidade não é apenas uma violação de direitos fundamentais, mas também uma forma de desumanização, pois quanto mais se sabe sobre uma pessoa, mais fácil se torna manipulá-la sem o seu consentimento ou conhecimento pleno (Zuboff, 2021).

A desumanização final ocorre quando a IA é usada para prever e controlar o futuro comportamento humano, privando os indivíduos de sua capacidade de autodeterminação. Zuboff denomina isso de "perda do direito ao futuro", uma vez que as pessoas não podem mais projetar suas próprias vidas, pois seus futuros já foram antecipados e moldados pelas tecnologias de vigilância. Nesse contexto, o ser humano deixa de ser uma entidade livre para fazer escolhas e se torna parte de um sistema pré-programado, cujas ações e decisões são direcionadas para resultados específicos que atendem aos interesses das corporações (Zuboff, 2021).

Nesse sentido, a IA, quando utilizada no contexto do capitalismo de vigilância, pode desumanizar os indivíduos ao transformar seus comportamentos e interações em dados comercializáveis, enfraquecendo sua autonomia, privacidade e liberdade.

Castells observa que essa transição não é neutra: o poder informacional distribui-se de modo desigual, consolidando uma "bipolarização estrutural" entre atores que definem padrões e protocolos de dados e aqueles que apenas os obedecem. A IA aprofunda essa clivagem ao deslocar a vantagem competitiva para quem controla infraestrutura cognitiva (dados, processamento, modelos) e plataformas (Castells, 1999). O resultado social é a emergência de trabalhadores redundantes, cuja obsolescência não decorre de falha individual, mas de uma programação institucional do desuso (Morgan, 2019; Ford, 2019).

A quantificação performativa criticada por Morgan, números que não apenas descrevem, mas produzem realidades organizacionais, explica por que indicadores de

“atividade digital” podem substituir avaliações substanciais de resultado (Morgan, 2019). Ao absolutizar proxies de comportamento (cliques, tempo de tela, “aderência digital”), a gestão transfere o centro de gravidade do trabalho realizado para o uso de máquina, reduzindo a pessoa a um perfil métrico. Esse é o ponto de fricção entre governança algorítmica e dignidade: o que é mensurável suplanta o que é significativo (Morgan, 2019; Zuboff, 2021).

Nos processos de recrutamento, avaliação e disciplina, os riscos se multiplicam. O’Neill demonstra que sistemas algorítmicos tendem a reproduzir vieses e a amplificar desigualdades, sobretudo quando treinados em dados históricos que refletem assimetrias do passado (O’Neill, 2020). No ambiente de trabalho, isso se converte em automação discriminatória: métricas opacas parametrizam trajetórias profissionais, naturalizando exclusões sob o selo da “neutralidade técnica” (Zuboff, 2021). A desumanização emerge aqui como apagamento do contexto: ler, analisar, pensar, atividades cognitivas sem “movimento”, podem ser classificadas como “ociosidade”.

Outro conjunto de conceitos que percorre esta seção exige distinção analítica: vigilância, controle, modulação comportamental e livre-arbítrio são fenômenos relacionados, mas não sinônimos. A vigilância é a coleta sistemática e contínua de dados sobre indivíduos, seus comportamentos, preferências, deslocamentos, interações, e constitui a condição de possibilidade dos demais fenômenos: sem dados, não há predição. O controle é o exercício de poder sobre condutas a partir dos dados coletados, pode ser direto, como no monitoramento de trabalhadores que gera demissões com base em “atividade digital”, ou indireto, como na definição de quais conteúdos o usuário pode ou não acessar. A modulação comportamental vai além do controle: não se trata de proibir ou punir, mas de arquitetar o ambiente de escolha de forma que o indivíduo seja induzido a determinados comportamentos sem perceber que está sendo direcionado, é o que Zuboff denomina “modificação comportamental em escala”. O livre-arbítrio, por sua vez, é a capacidade do sujeito de agir segundo sua própria vontade, sem coerção ou manipulação, e é exatamente o que a modulação comportamental corrói de forma mais insidiosa, porque o faz sem que o sujeito perceba que sua autonomia está sendo comprometida. A gravidade jurídica e ética de cada um desses fenômenos é diferente: a vigilância levanta problemas de privacidade; o controle levanta problemas de responsabilidade e devido processo; a modulação comportamental levanta problemas de consentimento e autodeterminação informativa; e a erosão do livre-arbítrio levanta problemas de dignidade humana na sua dimensão mais fundamental, a exemplo da cognição.

A capacidade humana de autodeterminação, discutida no contexto de desumanização e vigilância tecnológica, está profundamente ligada ao livre-arbítrio. O livre-arbítrio, em

essência, é a liberdade de escolha e de ação que define a autonomia moral e ética de um indivíduo. Quando essa liberdade é comprometida, seja por controle externo, coerção ou manipulação, a autodeterminação, ou seja, a capacidade de dirigir conscientemente sua vida e tomar decisões, também é enfraquecida. Ao monitorar e analisar os comportamentos dos indivíduos, essas tecnologias antecipam e moldam as decisões futuras de maneira imperceptível, o que limita as possibilidades de escolha genuína. O que aparenta ser um exercício de livre-arbítrio, na realidade, pode ser o resultado de manipulação preditiva, onde os algoritmos influenciam sutilmente a tomada de decisões para atender a interesses comerciais.

Zuboff descreve isso como uma erosão da liberdade humana, na qual os indivíduos são convertidos em meras fontes de dados a serem explorados. Quando suas ações e decisões são moldadas por forças invisíveis e externas, como algoritmos que preveem e modificam seu comportamento, o exercício de livre-arbítrio deixa de ser uma escolha verdadeiramente autônoma. Nesse contexto, a autodeterminação perde o sentido, pois as pessoas não estão mais em pleno controle de suas decisões e vidas (Zuboff, 2021).

Dessa forma, a desumanização tecnológica ocorre ao transformar seres humanos em objetos previsíveis e manipuláveis dentro de um sistema que visa maximizar o lucro. A perda da autonomia que Zuboff enfatiza se alinha diretamente à diminuição do livre-arbítrio, que é essencial para a autodeterminação. Sem a capacidade de escolher de forma autônoma, a dignidade humana é prejudicada, ecoando a visão kantiana de que a autonomia é uma base para a moralidade e liberdade (Zuboff, 2021).

O episódio acima mencionado do banco Itaú acerca das demissões em massa com base em “atividade digital” ilustra esse deslocamento do valor do trabalho para o comportamento monitorável, evidenciando a confusão entre produtividade e uso de dispositivo. Tal controvérsia reforça a necessidade de transparência robusta, critério de necessidade/minimização e revisão humana qualificada antes de decisões de alto impacto, sob pena de converter o poder diretivo em poder disciplinar algorítmico (Zuboff, 2021; Pannunzio, 2025). No plano sociojurídico, o caso exemplifica a fricção entre eficiência e direitos de personalidade no teletrabalho.

Do ponto de vista da filosofia prática, a erosão da autonomia que, em Kant, funda a dignidade e a moralidade é palpável quando escolhas são arquitetadas por sistemas que predizem e induzem condutas. O que parece exercício de livre-arbítrio pode ser resultado de manipulação preditiva, corroendo a autodeterminação e rebaixando o sujeito à condição de meio (e não de fim). A desumanização aqui é dupla: perda de privacidade e reconfiguração da vontade (Zuboff, 2021).

A literatura sobre o “*jobless growth*” (crescimento sem empregos) e a classe redundante conecta-se a esse diagnóstico, ao mostrar que a automação desloca o valor para processos algorítmicos, comprimindo o papel do trabalho humano (Ford, 2019). Contudo, como alerta Morgan, não há determinismo tecnológico inevitável: há escolhas políticas sobre quando, como e para que automatizar. A tecnocracia da predição, ao converter incertezas em números de governança, tende a despolitizar alternativas regulatórias (Morgan, 2019).

No plano jurídico, isso exige novas categorias de proteção: explicabilidade de sistemas decisórios; relatório de impacto à proteção de dados pessoais (RIPD) para monitoramento comportamental; regras de não discriminação algorítmica; padrões de intervenção humana significativa; contabilidade de riscos para decisões laborais; e direito ao contraditório sobre dados e modelos usados em avaliações (Zuboff, 2021; Morgan, 2019). Sem esses freios, a terceirização algorítmica da gestão converte-se em quase-determinismo tecnológico, travestindo passividade política de pragmatismo (MORGAN, 2019).

O ponto final é civilizacional: a IA, aplicada sob a lógica do capitalismo de vigilância e da quantificação performativa, redefine valor, mérito e propósito. A resposta pública não pode ser apenas reativa; deve moldar a tecnologia segundo fins éticos, inclusivos e sustentáveis, garantindo que o trabalho permaneça espaço de sentido e vínculo, e não mero vetor de captura comportamental. Reequilibrar inovação e dignidade, produtividade e humanidade, é tarefa normativa e política de primeira ordem.

Por fim, como lembra Morgan, o futuro do trabalho segue em disputa. A narrativa da inevitabilidade é uma opção política que preserva o status quo sob verniz técnico. A questão não é se a IA substituirá o humano, mas quais valores orientarão a transição. Cabe ao Direito e à sociedade civil recuperar o protagonismo, para que a tecnologia sirva à emancipação e não consolide um regime de exploração orientados por algoritmos (*algorithm-driven*). (MORGAN, 2019).

Portanto, a capacidade humana de autodeterminação está profundamente ligada ao livre-arbítrio, pois ambos dependem da liberdade de ação e escolha. Quando essa liberdade é manipulada por tecnologias de vigilância, os indivíduos perdem sua agência, tornando-se partes de um sistema que define suas escolhas por eles. Isso representa uma ameaça não só à autodeterminação, mas também à própria noção de humanidade e de existência autônoma.

### 2.3 PRECARIZAÇÃO DAS POLÍTICAS PÚBLICAS

O tema sobre a precarização das políticas públicas no contexto da Inteligência Artificial (IA) é amplamente discutido por Darrell M. West no artigo "*What happens if robots take the Jobs*"? The impact of emerging technologies on employment and public policy". Segundo o autor, a rápida evolução tecnológica está causando profundas transformações no mercado de trabalho, com a substituição de trabalhadores por máquinas, IA e algoritmos. A consequência disso é que as políticas públicas, que tradicionalmente são atreladas ao emprego formal, estão ameaçadas de se tornarem obsoletas (West, 2015).

Essa observação de West (2015) permite compreender que o impacto da automação transcende a esfera produtiva e alcança o próprio alicerce do Estado social moderno. Desde o século XX, o emprego formal tem funcionado como o eixo em torno do qual se estruturam os sistemas de seguridade, arrecadação e inclusão social. Assim, quando a inteligência artificial e a automação alteram a dinâmica do trabalho, não apenas a economia é transformada, mas a arquitetura institucional de proteção social é colocada em xeque. Nesse sentido, o fenômeno não é apenas econômico, mas jurídico e político, pois atinge diretamente os mecanismos de redistribuição e solidariedade social.

West (2015) observa que o modelo atual de distribuição de benefícios, como saúde, previdência e seguro social, depende amplamente do emprego. No entanto, à medida que a automação elimina uma parcela significativa dos empregos, esse sistema baseado em trabalho começa a perder viabilidade. O autor destaca que a eliminação de empregos já está ocorrendo em setores como manufatura, serviços e até em profissões de "colarinho branco", o que levanta questões sobre a continuidade das políticas públicas que dependem da contribuição dos trabalhadores (West, 2015).

Esse diagnóstico de West converge com o que Jamie Morgan (2019) denomina de "crise de sustentação institucional", decorrente da crença de que os sistemas sociais podem ser substituídos por métricas automatizadas de eficiência. Para Morgan, o mesmo raciocínio tecnocrático que legitima a automação do trabalho humano também está sendo aplicado à administração pública, a ideia de que dados, algoritmos e modelos preditivos podem substituir a deliberação democrática e o julgamento político. Essa "tecnificação da política" resulta em políticas públicas cada vez mais quantitativas, mas menos humanas, desprovidas de sensibilidade social e incapazes de lidar com a complexidade do real.

Para mitigar esse impacto, West (2015) sugere a implementação de novas políticas, como a Renda Básica Universal (UBI), que garantiria um rendimento mínimo para todos os

cidadãos, independentemente de estarem empregados ou não. Essa medida permitiria que as pessoas mantivessem um nível básico de subsistência, mesmo que seus empregos tenham sido eliminados pela automação.

A proposta de West se alinha com o que Martin Ford (2019) também defende em *The Rise of the Robots*, ao reconhecer que a automatização em larga escala ameaça o vínculo histórico entre trabalho e dignidade. Ambas as perspectivas propõem uma reconfiguração do pacto social, na qual o Estado assume a função de mediador entre inovação e justiça social. Essa reconfiguração, porém, exige mais do que medidas paliativas — demanda um reposicionamento da economia política para lidar com a concentração de riqueza gerada por sistemas automatizados.

Além disso, o autor propõe o conceito de *flexicurity*, um modelo em que os benefícios sociais, como saúde e previdência, seriam desvinculados do emprego formal. Isso garantiria que os cidadãos continuassem a receber apoio do governo, independentemente de estarem ou não no mercado de trabalho, prevenindo a criação de uma “subclasse” de desempregados sem acesso a benefícios (West, 2015).

Essa visão dialoga com o pensamento de Anthony Giddens e Ulrich Beck sobre a modernidade reflexiva, conceito que descreve o momento em que as instituições da modernidade industrial, Estado, mercado, ciência e família, passam a ser confrontadas pelas consequências do próprio progresso que sustentaram. Como Beck (1997) observa, a sociedade moderna está condenada a “autoconfrontar-se”, repensando as suas bases políticas e normativas à medida que seus próprios mecanismos de estabilidade produzem novas formas de instabilidade.

Nesse contexto, Giddens ressalta que as instituições precisam tornar-se flexíveis e reflexivas, isto é, capazes de revisar continuamente suas práticas à luz de novas informações e riscos emergentes. Trata-se de um processo que exige, não apenas adaptação, mas também autocrítica institucional: as estruturas políticas e sociais devem desenvolver mecanismos de aprendizado sistêmico, abandonando a rigidez hierárquica típica da modernidade industrial. A reflexividade institucional, segundo Giddens, significa que as instituições modernas incorporam o conhecimento sobre suas próprias operações, aprendem com os efeitos não intencionais de suas ações e ajustam-se continuamente a eles. Isso implica reconhecer que políticas públicas, sistemas de bem-estar e regimes de regulação não podem mais ser projetados como estruturas estáticas, mas como sistemas dinâmicos e abertos, permanentemente revisáveis (Beck; Giddens; Lash, 1997).

O impacto da automação também se reflete na arrecadação fiscal dos governos. Com menos trabalhadores empregados, a base tributária que sustenta os programas sociais será reduzida, criando um grande desafio fiscal para os governos. West (2015) adverte que, se os governos não encontrarem novas formas de financiamento, como a tributação das tecnologias, haverá um risco real de colapso das políticas públicas.

Nesse ponto, Castells (1999) é complementar ao afirmar que a economia informacional concentra poder e lucros em fluxos globais, enquanto os sistemas tributários permanecem locais e analógicos. A discrepância entre a escala da produção digital e a escala da tributação é um dos grandes paradoxos da era da informação. Assim, enquanto a IA multiplica os ganhos de produtividade, os Estados perdem capacidade de arrecadação e regulação, enfraquecendo suas políticas redistributivas.

Além da questão fiscal, West (2015) explora o impacto social da automação, destacando o aumento da desigualdade econômica. À medida que os empregos desaparecem e os lucros da automação são concentrados nas mãos de poucos, há um risco crescente de fragmentação social. Políticas públicas inclusivas, que promovam a redistribuição dos benefícios econômicos gerados pela automação, serão essenciais para evitar um aumento acentuado da desigualdade (West, 2015).

Cathy O'Neill (2020) reforça esse argumento ao demonstrar que algoritmos e sistemas automatizados, quando mal regulados, tendem a reproduzir padrões de exclusão e reforçar desigualdades já existentes. Assim, a desigualdade tecnológica não é um efeito colateral, mas uma consequência estrutural de uma governança digital centrada no lucro e não na justiça social. O risco é que o Estado acabe reproduzindo, por meio de seus próprios sistemas automatizados, as mesmas lógicas discriminatórias das corporações privadas.

West também argumenta que será necessário reformar o sistema educacional para preparar os trabalhadores para o novo mercado de trabalho. As atuais políticas educacionais muitas vezes preparam os alunos para empregos que estão desaparecendo. Ele propõe um foco maior na aprendizagem contínua e no treinamento de habilidades tecnológicas, de modo que os trabalhadores possam se adaptar às mudanças trazidas pela IA e pela automação (West, 2015).

Diogo Cortiz (2024) corrobora essa visão ao afirmar que a educação digital e a alfabetização em IA serão os pilares de qualquer política pública sustentável no século XXI. Ele destaca que o Brasil ainda carece de uma estrutura robusta de *reskilling* e *upskilling* e que a ausência de políticas de qualificação digital amplia o abismo entre o avanço tecnológico e a capacidade humana de acompanhá-lo. A consequência disso é a exclusão digital e o agravamento da desigualdade regional.

Iniciativas de incentivo ao voluntariado e à participação em projetos comunitários também são sugeridas por West como formas de manter as pessoas engajadas socialmente em um cenário onde os empregos se tornem mais escassos. Tais iniciativas poderiam ser incentivadas através de créditos sociais ou benefícios relacionados à participação em causas sociais (West, 2015).

Esse tipo de proposta se conecta com o conceito de capital social discutido por Putnam, no qual a participação cívica e o engajamento comunitário fortalecem o tecido social e reduzem a alienação (Pariser, 2012). Em um contexto de desemprego tecnológico, a valorização do engajamento não produtivo, como o voluntariado, a arte e a ciência cidadã, pode representar um novo eixo de pertencimento e dignidade, rompendo com a visão exclusivamente utilitarista do trabalho.

Outra política pública proposta por West (2015) é a participação nos lucros, em que os lucros gerados pela automação seriam redistribuídos para os trabalhadores. Isso garantiria que os benefícios das novas tecnologias não se concentrassem apenas nas mãos dos proprietários e acionistas, mas fossem compartilhados com a sociedade como um todo.

Esse tipo de medida se aproxima da ideia de dividendo tecnológico, um conceito que vem sendo discutido por economistas e juristas em fóruns internacionais. Trata-se de reconhecer que os ganhos gerados pela automação são produto de um ecossistema coletivo, composto por dados públicos, infraestrutura estatal e inovação social, e, portanto, parte desse valor deve retornar à sociedade. Essa lógica fundamenta também a proposta de tributação de robôs e IA, como defendem Chand e Kostić (2020), que sugerem um novo modelo fiscal capaz de redistribuir os ganhos do progresso tecnológico.

West (2015) conclui que as políticas públicas atuais não são suficientes para lidar com os desafios impostos pela automação e pela IA. A necessidade de um novo contrato social é urgente, com políticas que levem em consideração uma economia onde o trabalho humano é cada vez menos necessário. Isso exigirá dos governos uma capacidade de adaptação rápida e a implementação de soluções inovadoras para garantir o bem-estar da população em uma era de transformação tecnológica (West, 2015).

Morgan (2019) complementa essa visão ao alertar que o verdadeiro risco da automação não é apenas a perda de empregos, mas a perda de agência política e social. Quando os governos se tornam reféns de modelos econômicos tecnocêntricos e de métricas algorítmicas, a própria política pública se torna um simulacro, incapaz de intervir no mundo real. Portanto, o “novo contrato social” proposto por West deve incluir também uma nova epistemologia política, capaz de recolocar o humano no centro das decisões tecnológicas.

A discussão sobre a implementação de políticas como a renda básica universal (UBI) nesse contexto enfrenta resistência principalmente devido aos impactos negativos da automação e da IA sobre as receitas fiscais governamentais. Com a substituição de trabalhadores humanos por tecnologias automatizadas em vários setores, o governo enfrenta uma diminuição da base tributária. Essa redução acontece porque menos empregos resultam em menor arrecadação de impostos sobre a renda do trabalho, o que, por sua vez, afeta a capacidade dos governos de financiar políticas públicas, como saúde, educação e infraestrutura (Merola, 2022).

A adoção da automação, como mostrado por estudos recentes, pode deslocar trabalhadores em setores inteiros, levando à erosão da arrecadação tributária, principalmente em países dependentes de indústrias que estão sendo rapidamente automatizadas. Conforme a automação se espalha, governos perdem receitas provenientes de impostos sobre o trabalho, e, ao mesmo tempo, precisam buscar maneiras de taxar o uso de robôs e IA para compensar esse déficit (Chand; Kostić, 2020).

O problema é agravado pelo fato de que muitos sistemas fiscais ainda são baseados em modelos tradicionais de arrecadação, centrados no trabalho humano, e não se adaptaram completamente às mudanças trazidas pela automação. Uma possível solução seria a criação de impostos sobre robôs e IA, como sugerido em alguns países da Europa. Esses impostos visam redistribuir os ganhos obtidos com a automação e manter a arrecadação de impostos para financiar o estado de bem-estar social (Chand; Kostić, 2020).

Esse cenário mostra que, enquanto propostas como a renda básica universal podem ajudar a mitigar os efeitos negativos do desemprego causado pela automação, sua viabilidade está diretamente relacionada à capacidade dos governos de encontrarem novas formas de arrecadação. Sem essas novas fontes de receita, a implementação de políticas sociais mais abrangentes pode ser prejudicada, especialmente em países onde a dependência de empregos formais é alta (Merola, 2022).

O pesquisador Diogo Cortiz, afirma que a IA está promovendo mudanças profundas no mercado de trabalho global, impulsionando a produtividade, mas também levantando questões preocupantes sobre a desigualdade e a substituição de postos de trabalho. A previsão de aumento do PIB global em 7% nos próximos 10 anos devido à IA generativa, conforme estimado pelo Goldman Sachs, destaca o impacto econômico potencial dessa tecnologia. No entanto, Cortiz lembra que esses ganhos não estão igualmente distribuídos entre regiões e setores da sociedade. Regiões como América Latina, segundo o estudo da PwC, podem ser as mais prejudicadas por esse processo, ficando atrás de outras áreas, como os EUA e a China, que estão melhor preparadas para adotar essas inovações tecnológicas (Cortiz, 2024).

O impacto desigual da IA é evidente não apenas entre países, mas também dentro das próprias nações, exacerbando disparidades sociais e econômicas. De acordo com dados do *Economic Policy Institute*, entre 1980 e 2022, a produtividade global aumentou 64,7%, impulsionada pela automação, enquanto os salários cresceram apenas 14,8%. Isso resultou em uma concentração de renda e em um aumento da desigualdade. O problema, como discutido por Daron Acemoglu e Simon Johnson, é que o aumento da produtividade não se traduz necessariamente em uma maior demanda por trabalhadores. Embora a tecnologia crie novos empregos, isso acontece em ritmos e magnitudes diferentes, deixando muitos profissionais vulneráveis à substituição (Cortiz, 2024).

A adaptação ao uso da IA no mercado de trabalho requer uma estratégia de qualificação e requalificação profissional. Um estudo experimental de Harvard, MIT e Wharton mostrou que o uso de ferramentas de IA generativa aumenta a produtividade dos profissionais de consultoria em 12%, com 25% mais rapidez na execução de tarefas e 40% de melhoria na qualidade. A IA está se tornando uma ferramenta essencial para diversas profissões, e aqueles que não dominarem essas tecnologias correm o risco de ficarem à margem do mercado de trabalho (Cortiz, 2024).

No Brasil, as políticas públicas precisam focar na mitigação imediata dos impactos da IA nos empregos. Cortiz destaca a importância de implementar programas de *upskilling* e *reskilling* (qualificação e requalificação profissional), que já estão sendo adotados em outros países, para ajudar a força de trabalho a se adaptar às novas demandas tecnológicas. No entanto, as políticas públicas muitas vezes ficam atrás da inovação, e a falta de uma ação rápida pode deixar o Brasil ainda mais distante da liderança tecnológica global (Cortiz, 2024).

Recentemente, o Ministério do Trabalho e Emprego propôs, através da Portaria MTE nº 1.234 de 23 de julho de 2024, a criação de um grupo de trabalho sobre a IA. O objetivo deste grupo é justamente realizar estudos e propor políticas públicas, para preparar a população brasileira para o futuro do mercado de trabalho:

O MINISTRO DE ESTADO DO TRABALHO E EMPREGO, no uso da atribuição que lhe confere o art. 87, parágrafo único, inciso IV, da Constituição, **considerando a crescente importância da Inteligência Artificial (IA) no contexto do trabalho**, bem como a **necessidade de estudar e propor políticas**, produtos e serviços que garantam uma **transição justa e inclusiva para o futuro do trabalho** - (Processo nº 19955.204269/2024-98), resolve  
Art. 1º Instituir Grupo de Trabalho - GT para realizar estudos, proposição de políticas públicas, programas, ações e o desenvolvimento de serviços públicos envolvendo Inteligência Artificial para o futuro do trabalho [...] (Brasil, 2024a, p. 1).

O Brasil, começa a dar seus primeiros passos para enfrentar essa questão, pois é fundamental que o país desenvolva um projeto robusto de alfabetização digital. Isso permitirá que os trabalhadores brasileiros aprendam a interagir de forma crítica e eficaz com as ferramentas de IA compreendendo suas limitações e impactos. Sem essa preparação, o Brasil corre o risco de ficar preso em uma posição de baixa produtividade, sem a qualificação necessária para competir no cenário internacional de IA.

Em um cenário global, instituições como o Fundo Monetário Internacional (FMI) já alertaram que a transição para a IA exigirá redes de segurança social mais fortes e investimentos significativos em educação. Esses esforços são essenciais para amortecer o impacto das transformações tecnológicas, garantindo que os trabalhadores tenham as habilidades necessárias para prosperar em um ambiente de trabalho altamente automatizado (Cortiz, 2024).

No entanto, como Cortiz aponta, a criação dessas políticas públicas deve ser ágil e eficaz para acompanhar o ritmo das inovações tecnológicas. Caso contrário, o Brasil corre o risco de ficar estagnado, com uma força de trabalho despreparada e uma economia incapaz de competir no cenário global (Cortiz, 2024).

A integração eficaz da Inteligência Artificial (IA) na economia global e local dessa forma, não se limita ao desenvolvimento de infraestrutura tecnológica, mas exige um investimento estratégico em capital humano. A falta de uma força de trabalho qualificada para operar e lidar com as novas tecnologias pode deixar o Brasil marginalizado na revolução tecnológica, comprometendo a competitividade econômica. Nesse sentido, iniciativas de *upskilling* e *reskilling* são fundamentais para garantir que os trabalhadores adquiram as habilidades necessárias para prosperar em um mercado em rápida transformação.

Além disso, a automação não apenas desloca empregos, mas também ameaça a sustentabilidade fiscal dos governos, uma vez que reduz a arrecadação de impostos sobre a renda do trabalho. Para mitigar esse impacto, novas formas de tributação, como os impostos sobre robôs e IA, devem ser exploradas como alternativas para manter o financiamento de políticas públicas essenciais, como saúde e educação. Sem essas medidas, a implementação de soluções como a renda básica universal (UBI) pode ser inviabilizada, já que os governos enfrentarão um déficit crescente nas receitas fiscais necessárias para sustentar essas políticas.

O Capítulo 2 revelou que os impactos da IA no mercado de trabalho não são apenas econômicos, são profundamente políticos e éticos. A automação erode empregos, concentra riqueza, fragiliza políticas públicas e coloca em xeque o próprio contrato social que sustentou o Estado de bem-estar no século XX. Diante desse quadro, uma questão se impõe com urgência: como o Direito pode responder a essas transformações? A resposta passa necessariamente pela

construção de um marco regulatório que seja ao mesmo tempo robusto o suficiente para proteger direitos fundamentais e flexível o suficiente para não sufocar a inovação. É essa arquitetura normativa, em disputa no Brasil e no mundo, que o próximo capítulo se propõe a examinar.

### **CAPÍTULO 3 – REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL**

Se os capítulos anteriores mapearam os impactos sociais e laborais da IA, este capítulo volta o olhar para o Direito e para sua capacidade, ou incapacidade, de acompanhar o ritmo das transformações tecnológicas. Regulamentar a IA não é tarefa simples: exige clareza conceitual sobre o que se está regulando, critérios objetivos para classificar riscos e mecanismos efetivos de responsabilização que não confrontem com a inovação e o desenvolvimento tecnológico. No Brasil, esse debate ganhou forma através de uma trajetória legislativa que avançou significativamente entre 2019 e 2025, culminando no PL 2338/2023, proposta que será analisada em detalhe neste capítulo, em comparação com o PL 21/2020 e com os modelos regulatórios internacionais que inspiraram sua arquitetura.

A trajetória legislativa para a regulação da inteligência artificial (IA) no Brasil entre os anos de 2019 e 2025 representa um esforço progressivo de construção de um arcabouço normativo que articula inovação tecnológica, proteção de direitos fundamentais e compatibilidade com padrões internacionais. Em 2019, o Senado Federal registrou o primeiro projeto específico para IA, o Projeto de Lei nº 5.051/2019, que propunha princípios para a utilização de sistemas de IA no país, mas com escopo ainda bastante incipiente. Em 2020, a Câmara dos Deputados aprovou o Projeto de Lei nº 21/2020, que estabelecia fundamentos, diretrizes e princípios gerais para o desenvolvimento e uso de IA no Brasil, com ênfase em ética, transparência e responsabilidade (Brasil, 2020). Em 2021, no Senado, surgiu o Projeto de Lei nº 872/2021, que incrementava as discussões incorporando marcos éticos e mais detalhamento sobre responsabilidade e governança.

Na sequência, em fevereiro de 2022, foi instituída pelo Ato nº 4/2022 a Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA) (Senado Federal), com o propósito de consolidar os vários projetos (5.051/2019; 21/2020; 872/2021) em um substitutivo integrado. A comissão realizou audiências públicas e entregou seu relatório final no final de 2022 (STJ, 2022). Com base nesse trabalho, em 3 de maio de 2023 foi apresentado o Projeto de Lei nº 2338/2023 no Senado Federal, já dotado de arquitetura regulatória baseada em risco, com classificações de sistemas de IA (alto risco, risco excessivo), governança, avaliação de impacto algorítmico e responsabilidades específicas (Brasil, 2023; DataPrivacyBrasil, 2024).

Em 5 de dezembro de 2024, a Comissão Temporária Interna sobre Inteligência Artificial (CTIA) do Senado aprovou o relatório substitutivo referente ao PL 2338/2023, abrindo caminho para votação em Plenário. No dia 10 de dezembro de 2024 o Senado aprovou o PL 2338/2023

e o remeteu à Câmara dos Deputados para seguir tramitação (Senado Federal, 2024; DataPrivacyBrasil, 2024). Já em 2025, o projeto passou à Câmara, onde foi criada em 29 de abril de 2025 uma Comissão Especial destinada a proferir parecer sobre o PL 2338/2023, com regime de prioridade, contemplando audiências e debates temáticos ao longo do ano (Câmara dos Deputados, 2025).

O resultado desse percurso legislativo revela um amadurecimento no ordenamento regulatório da IA: parte-se de propostas gerais e discretas para princípios e diretrizes (PL 5.051/2019; PL 21/2020; PL 872/2021) e culmina em uma proposta mais robusta e técnica (PL 2338/2023) com inserção de mecanismos de governança, classificação de risco e responsabilização diferenciada. O status atual, até o momento, com o PL 2338/2023 aguardando parecer na Comissão Especial da Câmara indica que ainda não há lei definitiva, mas há um foco legislativo claro e prioritário no tema da IA.

O PL 2338/2023, que tem como objetivo modernizar o arcabouço jurídico sobre a IA, inova ao introduzir uma categorização dos sistemas de IA com base no risco que eles representam, sendo que os sistemas de alto risco e risco excessivo estão sujeitos a uma responsabilidade objetiva dos agentes envolvidos. Este projeto se diferencia do PL 21/2020 ao incorporar uma abordagem de avaliação de impacto algorítmico e governança, visando garantir a transparência e a segurança dos sistemas de IA, além de uma responsabilização mais rigorosa para os sistemas de maior risco. O PL 2338/2023 segue tendências regulatórias internacionais, como as estabelecidas no *AI Act* da União Europeia, e busca criar um ambiente regulatório flexível, que leve em consideração tanto a inovação quanto a proteção dos direitos fundamentais (Brasil, 2023).

Todos os projetos de lei mencionados dialogam diretamente com a Lei Geral de Proteção de Dados (LGPD), que já estabelece diretrizes de governança para o tratamento de dados pessoais e busca proteger os direitos dos cidadãos no uso de tecnologias emergentes. No entanto, enquanto o PL 21/2020 adota uma visão mais geral sobre a responsabilidade e a operação da IA, o PL 2338/2023 propõe uma regulação mais detalhada, com normas específicas para sistemas de IA de alto risco e obrigações de transparência e prestação de contas.

Ao longo deste capítulo, será feita uma análise comparativa entre o PL 21/2020 e o PL 2338/2023, com foco em suas implicações penais e na responsabilidade jurídica dos operadores e desenvolvedores de IA. O objetivo é explorar como esses projetos se integram ao contexto da LGPD e discutir as possíveis consequências jurídicas para os diferentes níveis de responsabilidade atribuídos aos agentes envolvidos.

### 3.1 CONTEXTO JURÍDICO DO PL 21/20 E DO PL 2338/23

A regulação da inteligência artificial (IA) no Brasil envolve principalmente (outros foram acima mencionados) dois projetos de lei fundamentais: o PL 21/2020 e o PL 2338/2023, cada um com abordagens distintas em relação à responsabilidade civil e penal, governança e proteção de direitos fundamentais. Ambos os projetos tentam lidar com os desafios éticos e jurídicos que surgem com o uso crescente de IA, mas seguem caminhos diferentes em termos de escopo e rigor.

#### 3.1.1 PL 21/2020

O PL 21/2020, de iniciativa do Deputado Federal Eduardo Bismarck (PDT/CE), traz uma abordagem mais simples, focando na definição de princípios e diretrizes gerais para o desenvolvimento e uso da inteligência artificial no Brasil. O principal objetivo desse projeto é promover um ambiente de inovação tecnológica, ao mesmo tempo que assegura a proteção dos direitos fundamentais e estabelece normas que possam ser seguidas tanto por desenvolvedores quanto por usuários de sistemas de IA (Brasil, 2020).

Esse projeto foi apresentado à Câmara dos Deputados em 4 de fevereiro de 2020, com a ementa original “*Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências*”. Posteriormente, recebeu nova ementa, passando a “*Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências*”. (Câmara dos Deputados, 2020a)

No texto integral aprovado (Câmara dos Deputados, 2020b), o art. 1º dispõe que “*esta Lei estabelece fundamentos e princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil e diretrizes para o fomento e a atuação do poder público nessa área*”. Já o art. 6º, inciso VI, prevê:

responsabilidade: as normas sobre responsabilidade dos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial deverão, salvo disposição legal em contrário, pautar-se na responsabilidade subjetiva e levar em consideração a efetiva participação desses agentes, os danos específicos que se deseja evitar ou remediar e a forma como esses agentes podem demonstrar adequação às normas aplicáveis, por meio de esforços razoáveis compatíveis com os padrões internacionais e as melhores práticas de mercado (PL 21/2020, art. 6º VI) (Brasil, 2020).

Esse dispositivo é especialmente relevante, pois enquadra a responsabilidade dos agentes em torno da culpa ou dolo (subsidiada pelo termo “subjativa”), bem como exige que os agentes demonstrem a adoção de “esforços razoáveis” em conformidade com boas práticas internacionais.

Quanto à tramitação no Senado, a ficha da matéria indica que o PL 21/2020 foi recebido pelo Senado em 30 de setembro de 2021, conforme apontado pela base de matérias bicamerais (Senado Federal, 2021). A tramitação segue em regime de matéria bicameral, estando pendente de deliberação plena no Senado ou em comissão especial. (Senado Federal, 2021)

Na Câmara, o histórico revela que o PL foi objeto de requerimentos de audiência pública (Requerimento 7/2021) na Comissão de Ciência, Tecnologia, Comunicação e Informática (CCTCI) em 24 de março de 2021 e sofreu apensação com outros projetos correlatos (PL 240/2020, PL 4120/2020, PL 1969/2021) em julho de 2021. (Câmara dos Deputados, 2020b).

Nesse entendimento, no que se refere à responsabilidade penal e civil, o PL 21/20 adota uma perspectiva mais conservadora, mantendo a responsabilidade subjetiva para os danos causados por IA. Isso significa que a responsabilização dos desenvolvedores e operadores dependerá da comprovação de culpa ou dolo, o que tem gerado críticas por parte da doutrina. A complexidade dos sistemas de IA e sua capacidade de atuar autonomamente podem dificultar a atribuição de responsabilidade direta, especialmente em casos onde não se pode identificar facilmente a causa do dano ou o responsável direto pela falha (Brasil, 2020).

Essa abordagem foi duramente criticada, sendo considerada insuficiente para lidar com os riscos potenciais que os sistemas de IA apresentam, especialmente em contextos em que a autonomia decisória desses sistemas dificulta a atribuição de responsabilidade, pois não reflete a evolução da disciplina da responsabilidade civil no Brasil, que já adota a responsabilidade objetiva em casos de atividades de risco elevado, de acordo com o disposto no Código Civil:

Art. 927. Aquele que, por ato ilícito (arts. 186 e 187), causar dano a outrem, fica obrigado a repará-lo.

Parágrafo único. **Haverá obrigação de reparar o dano, independentemente de culpa, nos casos especificados em lei, ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem** (Brasil, 2002).

Como destacou o especialista Filipe Medon, o PL 21/2020 falhou em abordar questões críticas como a governança e a proteção de dados, adotando um modelo que beira a inconstitucionalidade (Medon, 2023).

Nesse sentido, a doutrina sustenta que, dada a natureza autônoma, complexa e potencialmente de elevado risco dos sistemas de IA, um modelo que privilegia a responsabilidade subjetiva, com necessidade de comprovar culpa ou dolo, pode se revelar insuficiente, dificultando a reparação das vítimas de falhas em sistemas de IA (Jota, 2021). Por exemplo, Anderson Schreiber argumenta que “o PL da Inteligência Artificial erra ao criar um regime de responsabilidade subjetiva” em contexto em que a tecnologia se comporta de modo autônomo e imprevisível (Schreiber, 2021).

Adicionalmente, uma nota técnica da FGV aponta que o PL 21/2020 carece de clareza sobre critérios de imputação de responsabilidade na cadeia de desenvolvimento e operação dos sistemas de IA, o que pode gerar insegurança jurídica para potenciais vítimas (FGV, 2022).

Em suma, o regime de responsabilidade subjetiva adotado pelo PL 21/2020 é visto como desalinhado com o princípio da atividade de risco elevado em IA, onde a responsabilização objetiva é preferível para garantir efetiva reparação e prevenção.

### **3.1.2 PL 2338/2023**

O Projeto de Lei nº 2.338, de 2023, apresentado pelo Senado Federal, institui um marco regulatório para o desenvolvimento, implementação e uso de sistemas de inteligência artificial (IA) no Brasil, com ênfase na “governança responsável” e na proteção dos direitos fundamentais (Brasil, 2023, art. 1º).

A arquitetura de regulação adotada pelo PL 2338/2023 segue um modelo baseado em risco, categorizando sistemas de IA em níveis diferenciados, como “risco excessivo”, “alto risco” e demais, sendo que os de risco excessivo podem ser proibidos, os de alto risco sujeitos a controles mais rigorosos, e os outros submetidos a obrigações menos onerosas (DireitosNaRede, 2024).

Entre os requisitos para sistemas de alto risco, o texto prevê a realização de avaliação de impacto algorítmico (AIA), a documentação técnica extensiva, mecanismos de mitigação de riscos, supervisão humana, rotulagem e informação clara ao usuário, além da obrigação de comunicar incidentes graves à autoridade competente e manter logs e registros. Conforme a análise preliminar da Autoridade Nacional de Proteção de Dados (ANPD), o PL busca harmonizar governança de IA com a proteção de dados pessoais, além de promover inovação responsável.

No que tange à responsabilidade, o art. 27 do PL 2338/2023 estipula que “o fornecedor ou operador de sistema de inteligência artificial que causar dano patrimonial, moral, individual

ou coletivo é obrigado a repará-lo integralmente, independentemente do grau de autonomia do sistema” (Brasil, 2023, art. 27). Em seu §1º, para sistemas de alto risco ou risco excessivo, aplica-se a responsabilidade objetiva (sem necessidade de prova de culpa); no §2º, para os demais sistemas, a culpa do agente é presumida, cabendo-lhe a inversão do ônus da prova em favor da vítima. Esse regime mostra-se alinhado com o paradigma internacional de responsabilização ampliada para sistemas de alto risco.

O art. 14 do projeto traz vedações expressas para sistemas considerados de risco excessivo o, como: I) o uso de técnicas subliminares que induzam pessoa natural a comportar-se de modo prejudicial à sua saúde, segurança ou contra os fundamentos da lei; II) exploração de vulnerabilidades de grupos específicos (por exemplo, idade, deficiência física ou mental) com objetivo prejudicial; III) utilização pelo poder público de sistemas de IA para classificação ou ranqueamento de pessoas com base no comportamento social ou atributos da personalidade para acesso a bens, serviços ou políticas públicas de forma ilegítima ou desproporcional, segue abaixo o artigo 14 na sua íntegra:

Art. 14. São vedadas a implementação e o uso de sistemas de inteligência artificial:  
I – que empreguem técnicas subliminares que tenham por objetivo ou por efeito induzir a pessoa natural a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança ou contra os fundamentos desta Lei;  
II – que explorem quaisquer vulnerabilidades de grupos específicos de pessoas naturais, tais como as associadas a sua idade ou deficiência física ou mental, de modo a induzi-las a se comportar de forma prejudicial a sua saúde ou segurança ou contra os fundamentos desta Lei;  
III – pelo poder público, para avaliar, classificar ou ranquear as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, por meio de pontuação universal, para o acesso a bens e serviços e políticas públicas, de forma ilegítima ou desproporcional (Brasil, 2023).

Essa proibição remete à dimensão ética de proteção física e mental dos indivíduos, alinhando-se ao princípio de que sistemas autônomos não devem causar dano a pessoas nem manipular psicologicamente grupos vulneráveis.

Em termos institucionais e de fiscalização, o PL prevê a criação do Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA), no qual a ANPD desempenhará papel central na certificação, auditoria, supervisão de incidentes e coordenação de mecanismos regulatórios setoriais.

O projeto segue tendências internacionais, como o *AI Act* da União Europeia, e estabelece a responsabilidade objetiva para sistemas de IA classificados como de alto risco ou risco excessivo. Isso significa que os agentes responsáveis por esses sistemas serão automaticamente responsabilizados pelos danos causados, independentemente de culpa. Essa

mudança visa garantir uma maior proteção aos usuários e às vítimas de possíveis falhas dos sistemas de IA refletindo a percepção de que as tecnologias de IA podem gerar riscos consideráveis que precisam ser controlados de maneira mais rigorosa. Abaixo segue a transcrição do já mencionado artigo 27 do projeto em questão:

Art. 27. O fornecedor ou operador de sistema de inteligência artificial que cause dano patrimonial, moral, individual ou coletivo é obrigado a repará-lo integralmente, independentemente do grau de autonomia do sistema.

§ 1º Quando se tratar de sistema de inteligência artificial de alto risco ou de risco excessivo, o fornecedor ou operador **respondem objetivamente pelos danos causados**, na medida de sua participação no dano.

§ 2º Quando não se tratar de sistema de inteligência artificial de alto risco, a culpa do agente causador do dano será presumida, aplicando-se a inversão do ônus da prova em favor da vítima (Brasil, 2023).

Essa abordagem foi amplamente elogiada por consolidar uma visão inovadora e detalhada de governança e mitigação de riscos, que inclui a exigência de avaliações de impacto algorítmico. A Autoridade Nacional de Proteção de Dados (ANPD) também apontou que o PL 2338/2023 complementa a LGPD, integrando mecanismos de governança que garantem a proteção de dados e direitos fundamentais (ANPD, 2023).

### 3.1.3 Substitutivo ao PL 2338/2023

Em 27 de novembro de 2023, o senador Marcos Pontes apresentou um substitutivo ao PL 2338/2023 que trouxe preocupações significativas. Segundo Medon (2023), o novo texto eliminou muitos dos avanços do projeto original, incluindo o capítulo que consagrava os direitos das pessoas afetadas por sistemas de IA e as medidas de governança para mitigação de riscos. Além disso, a nova versão suprimiu a exigência de avaliação de impacto algorítmico, um dos pilares do modelo regulatório baseado em riscos.

A análise técnica da organização Instituto Brasileiro de Defesa do Consumidor (I-DEC) afirma que o substitutivo retira ou atenua cláusulas essenciais de governança, avaliação de impacto, e dilui salvaguardas aos direitos dos afetados por sistemas de IA (I-DEC, 2024).

Entre os pontos contenciosos está a supressão da obrigatoriedade da realização de avaliação de impacto algorítmico (AIA), para determinadas categorias de IA, o enfraquecimento dos mecanismos de participação pública e transparência, e a concessão de maior margem de atuação às empresas desenvolvedoras sem exigência proporcional de auditoria independente. Em particular, a crítica incide sobre o fato de o substitutivo ter removido

capítulos inteiros que consagravam direitos dos indivíduos afetados por IA (DataPrivacyBrasil, 2024).

Embora o substitutivo mantenha algumas definições básicas sobre IA e proponha medidas de fomento ao desenvolvimento de IA no Brasil, as mudanças trouxeram retrocesso. O novo texto foi criticado por reduzir drasticamente as obrigações de governança e por introduzir conceitos vagos e sem clareza técnica sobre a responsabilidade por danos, especialmente ao misturar conceitos de responsabilidade civil e penal (Medon, 2023).

O artigo 20, por exemplo, atribui ao desenvolvedor a responsabilidade integral por danos causados por IA de alto risco, excluindo os operadores e usuários de qualquer responsabilidade, o que levanta preocupações sobre a injusta distribuição de riscos e a falta de proteção para as vítimas (Medon, 2023).

### **3.1.4 Situação Atual do PL 2338/2023**

Em dezembro de 2024, o Senado Federal aprovou o Projeto de Lei nº 2.338/2023, consolidando um dos passos mais relevantes rumo à criação de um marco regulatório nacional para a inteligência artificial (IA). O texto foi fruto de intenso debate no âmbito da Comissão Temporária Interna sobre Inteligência Artificial (CTIA), que havia recebido o relatório final do relator, senador Eduardo Gomes, após mais de trinta emendas parlamentares. A aprovação marcou o encerramento da tramitação no Senado e o início da fase de análise pela Câmara dos Deputados (Senado Federal, 2024).

A proposta aprovada busca instituir princípios, direitos e deveres para o desenvolvimento e uso da IA no Brasil, criando um arcabouço jurídico de caráter flexível e orientado ao risco. Essa característica é fundamental, pois permite que o ordenamento jurídico acompanhe as rápidas transformações tecnológicas, mantendo, ao mesmo tempo, a proteção aos direitos fundamentais. O modelo adotado segue a tendência internacional, notadamente inspirada no AI Act da União Europeia (Senado Federal, 2024).

Após a aprovação no Senado, o projeto foi encaminhado à Câmara dos Deputados em março de 2025, sendo oficialmente recebido em 17 daquele mês. Em 29 de abril do mesmo ano, instituiu-se a Comissão Especial sobre Inteligência Artificial, composta por parlamentares de diferentes partidos, com a finalidade de analisar o texto e propor ajustes, indicando o caráter prioritário da tramitação. Essa fase é considerada crucial, pois visa consolidar um consenso entre setores econômicos, acadêmicos e órgãos de controle (Câmara dos Deputados, 2025).

A tramitação atual mostra que o Brasil ainda se encontra em um momento de construção normativa e institucional. A regulação da IA permanece em debate e depende da definição de conceitos-chave, como as categorias de risco, as medidas de governança aplicáveis e o papel da Autoridade Nacional de Proteção de Dados (ANPD) na supervisão e fiscalização. Tal cenário reflete a natureza dinâmica e complexa do tema, que exige constante diálogo entre as esferas pública e privada.

Entre os tópicos centrais em discussão na Comissão Especial estão os prazos de *vacatio legis*. O texto aprovado estabelece uma *vacatio legis* de dois anos para a maior parte de seus dispositivos, de modo a permitir a adaptação gradual de agentes públicos e privados às novas exigências legais. Há, contudo, exceções relevantes: as normas que tratam de sistemas de propósito geral, aplicações proibidas e direitos autorais terão vigência antecipada, passando a valer em até 180 dias após a publicação da lei. Já as disposições voltadas à sustentabilidade, apoio a pequenas e médias empresas e à estruturação dos órgãos reguladores, com exceção das sanções possuirão aplicabilidade imediata, dada sua natureza operacional e de fomento.

Dessa forma, ao mesmo tempo em que o projeto define um cronograma escalonado de implementação, também delimita com precisão as categorias de risco e os sistemas submetidos a maior rigor regulatório. Entretanto, um ponto relevante é que o texto aprovado optou por não incluir os algoritmos utilizados por redes sociais na lista de sistemas considerados de alto risco. Esses algoritmos, formados por conjuntos de dados e parâmetros que determinam o ranqueamento e o alcance das postagens, permanecem, assim, fora do escopo mais restritivo da legislação, ainda que exerçam grande influência sobre o comportamento informacional e social dos usuários (Migalhas, 2024).

Os sistemas classificados como de alto risco, por sua vez, deverão observar um conjunto rigoroso de exigências legais. Entre elas, destacam-se: a realização de avaliação de impacto algorítmico, destinada a identificar os riscos a direitos fundamentais, mensurar benefícios, estimar a gravidade e probabilidade de efeitos adversos e estabelecer estratégias de mitigação; a execução de testes de confiabilidade e segurança; a implementação de mecanismos de detecção e correção de vieses discriminatórios, com o auxílio de ferramentas de registro e auditoria automática; e a adoção de medidas de transparência que permitam a explicação e rastreabilidade dos resultados produzidos pelos sistemas.

Já os sistemas de propósito geral, capazes de desempenhar múltiplas tarefas e adaptáveis às finalidades de seus usuários, como os modelos generativos, que produzem textos, imagens, sons e vídeos, passarão a ser regidos por normas específicas. O objetivo é garantir que tais sistemas, dada sua ampla aplicabilidade e impacto transversal, observem padrões mínimos de

segurança, rastreabilidade e governança ética compatíveis com os riscos que representam (Migalhas, 2024).

Além da questão técnica, o debate político em torno do PL 2338/2023 evidencia preocupações sobre como equilibrar inovação e regulação. Representantes do setor produtivo têm defendido maior incentivo à autorregulação e à inovação aberta, enquanto organizações da sociedade civil alertam para a necessidade de garantir proteção efetiva aos direitos fundamentais e à integridade dos cidadãos, especialmente diante do uso crescente de sistemas autônomos (Coalizão Direitos na Rede, 2025).

A Comissão Especial realizou, entre maio e setembro de 2025, diversas audiências públicas para ouvir especialistas nacionais e internacionais. Essas audiências contaram com a participação de universidades, centros de pesquisa, representantes de empresas de tecnologia e entidades civis. O relatório parcial das audiências demonstrou alto grau de engajamento social, refletindo a amplitude e a importância do tema para o futuro da governança digital no país.

Outra questão central na tramitação diz respeito à criação do Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA), estrutura que será responsável por coordenar a aplicação das normas e supervisionar os agentes envolvidos. A ANPD deverá exercer papel de destaque no SIA, atuando na certificação de sistemas, auditoria, comunicação de incidentes graves e na orientação das autoridades setoriais (Brasil, 2023).

Ainda assim, especialistas observam que a efetividade do SIA dependerá de regulamentações complementares e da capacidade técnica e orçamentária da ANPD para assumir novas atribuições. O Brasil, segundo análise de especialistas, avança para um modelo de governança de IA sofisticado, mas que exigirá coordenação interinstitucional e aprimoramento contínuo para garantir segurança e previsibilidade jurídica.

A Câmara dos Deputados também aprovou requerimentos para a realização de seminários regionais com o objetivo de discutir o impacto do PL 2338/2023 em áreas estratégicas, como saúde, educação, agricultura e segurança pública. A descentralização desses debates reforça a importância transversal da regulação da IA e amplia o alcance democrático do processo legislativo (Câmara dos Deputados, 2025).

Enquanto isso, entidades representativas de criadores e artistas comemoraram a aprovação do projeto no Senado, destacando a relevância de dispositivos que tratam da responsabilidade sobre obras geradas por IA e da preservação dos direitos autorais. Entretanto, a crítica mais recorrente continua a ser o possível desequilíbrio entre incentivos econômicos e a proteção efetiva de direitos humanos (CISAC, 2024).

Em síntese, o PL 2338/2023 representa um marco na construção de uma política nacional de governança da inteligência artificial. Embora avance na definição de responsabilidades e princípios éticos, o projeto ainda se encontra em processo de amadurecimento legislativo e carece de regulamentação específica. Sua aprovação definitiva e posterior implementação determinarão o grau de proteção que o ordenamento jurídico brasileiro oferecerá aos cidadãos frente aos desafios éticos, sociais e cognitivos impostos pela IA (Senado Federal, 2024).

### 3.2 ANÁLISE COMPARATIVA: PL 2120 vs. PL 2338

Enquanto o PL 2120/2020 se mantém alinhado a uma visão tradicional de responsabilidade subjetiva, o PL 2338/2023 propõe uma regulação mais moderna e rigorosa, especialmente para sistemas de IA que apresentam maiores riscos à sociedade. A adoção da responsabilidade objetiva no PL 2338/2023 busca facilitar a reparação dos danos causados por IA transferindo o ônus da prova para os agentes que desenvolvem e operam esses sistemas.

#### 3.2.1 Responsabilidade Civil

Os dois projetos tratam a responsabilidade penal de maneira distinta, refletindo abordagens diferentes em relação à aplicação de penalidades para desenvolvedores e operadores de sistemas de IA.

- PL 2120/2020: Este projeto adota uma postura mais conservadora e limitada no tratamento da responsabilidade dos agentes envolvidos com IA. No PL 2120, a responsabilização dos operadores e desenvolvedores de IA é baseada principalmente em um modelo de responsabilidade subjetiva. Ou seja, a responsabilização depende da comprovação de culpa ou dolo, o que limita a aplicação de penalidades em casos onde não se consegue claramente atribuir a intenção ou negligência ao agente causador de um dano. Essa abordagem, como criticado em vários artigos e pela doutrina, dificulta a responsabilização efetiva dos desenvolvedores, especialmente considerando a complexidade e a autonomia das decisões dos sistemas de IA (Brasil, 2020).
- PL 2338/2023: O PL 2338 inova ao introduzir os conceitos de "alto risco" e "risco excessivo", que determinam diferentes níveis de responsabilidade. Os sistemas de IA classificados como de alto risco estão sujeitos a uma regulamentação mais

rigorosa, exigindo que os operadores e desenvolvedores adotem medidas para mitigar os riscos de danos causados por suas decisões. Além disso, os sistemas de risco excessivo, como armas autônomas ou tecnologias que possam infringir direitos fundamentais, são proibidos, e os responsáveis podem ser penalizados severamente por sua implementação. Esse modelo baseia-se em uma abordagem de responsabilidade objetiva, facilitando a imputação de penalidades sem a necessidade de provar culpa direta (Brasil, 2023).

### 3.2.2 Avaliação de Impacto Algorítmico e Governança

Outro ponto-chave de comparação entre os dois projetos é a forma como lidam com a governança e a Avaliação de Impacto Algorítmico (AIA), aspectos essenciais para garantir a transparência e a segurança dos sistemas de IA.

- PL 2120/2020: O PL 2120 não apresenta uma estrutura clara para a governança ou a avaliação de impacto dos sistemas de IA. A ausência de diretrizes específicas para avaliar o impacto algorítmico de sistemas autônomos e de alto risco é vista como uma falha importante no projeto, que não se alinha às melhores práticas internacionais de regulação tecnológica, como as adotadas pela União Europeia no *AI Act* (ANPD, 2023).
- PL 2338/2023: O PL 2338 faz um avanço significativo nesse aspecto ao incluir a avaliação de impacto algorítmico como uma exigência para sistemas de IA classificados como de alto risco. Essa avaliação exige que os desenvolvedores e operadores identifiquem os possíveis impactos negativos de seus sistemas, proponham medidas preventivas e mitigadoras e garantam a reversão de efeitos indesejados, quando necessário (Brasil, 2023). O relatório da ANPD também destaca a importância da AIA para garantir a conformidade com os direitos fundamentais e a proteção de dados pessoais, alinhando o PL 2338 com as práticas do *AI Act* da União Europeia (ANPD, 2023).

Além disso, o PL 2338 também propõe a criação de uma governança regulatória, que incluirá a Autoridade Nacional de Proteção de Dados (ANPD), responsável por supervisionar e certificar sistemas de IA, e o Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA), que regulará e fiscalizará o mercado de IA (análise-preliminar-do-p...).

### 3.2.3 Diferenças na Abordagem Regulamentar: PL 2338 e PL 21/2020 em Relação a Sistemas de Risco Excessivo

A diferença entre o PL 2338/2023 e o PL 21/2020 no que se refere à regulação de sistemas de risco excessivo é significativa, especialmente quando analisada à luz dos Princípios de Asimov sobre segurança física e proteção mental dos indivíduos. O PL 2338 adota uma abordagem mais rigorosa e detalhada para proteger os direitos dos cidadãos, enquanto o PL 21/2020 mantém uma perspectiva mais genérica e menos específica sobre a regulamentação da IA.

As três leis/diretivas de Asimov, para uma melhor compreensão das comparações que vão ser feitas abaixo, são as seguintes:

As três diretivas que Asimov fez implantarem-se nos "cérebros positrônicos" dos robôs em seus livros são:

1.<sup>a</sup> Lei: Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal.

2.<sup>a</sup> Lei: Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que entrem em conflito com a Primeira Lei.

3.<sup>a</sup> Lei: Um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis.

Mais tarde Asimov acrescentou a "Lei Zero", acima de todas as outras: um robô não pode causar mal à humanidade ou, por omissão, permitir que a humanidade sofra algum mal. (Leis da Robótica, 2024, p. 1).

O artigo 14 do PL 2338/2023, já mencionado, estabelece uma clara proibição para sistemas de IA que utilizem técnicas subliminares ou que explorem vulnerabilidades de grupos específicos, como crianças, idosos, e pessoas com deficiências físicas ou mentais. Esses sistemas são categorizados como de "risco excessivo", uma classificação que impede seu desenvolvimento ou uso, dada a possibilidade de causarem danos significativos aos indivíduos. Além disso, o artigo também proíbe o uso de IA para a classificação ou ranqueamento de pessoas com base em atributos pessoais ou comportamentais para acessar bens e serviços de maneira desproporcional e ilegítima (Brasil, 2023).

Essas disposições remetem diretamente aos Princípios de Asimov, especificamente o princípio da proteção física e da proteção mental, que são conceitos essenciais no debate sobre ética em IA. Os princípios de Asimov, particularmente a Primeira Lei "Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal", enfatizam a segurança e a integridade física dos indivíduos, algo que o PL 2338 reflete ao proibir o uso de técnicas que possam manipular ou prejudicar fisicamente pessoas vulneráveis.

A proteção mental é outro aspecto crucial abordado pelo PL 2338. Ao proibir o uso de técnicas subliminares, ou seja, métodos que influenciam o comportamento humano de maneira imperceptível, o projeto de lei está explicitamente buscando evitar a manipulação psicológica ou mental das pessoas. Isso também se alinha o projeto com as diretrizes de Asimov, onde está implícito que robôs (ou, no caso, sistemas de IA) devem operar de forma a não comprometer a autonomia mental dos indivíduos.

Em um contexto em que as tecnologias de IA podem ser usadas para modificar comportamentos sem o conhecimento ou consentimento explícito dos afetados, o PL 2338 protege contra abusos, garantindo que a IA seja usada de forma ética e segura. Esse tipo de proteção é essencial em um momento em que as tecnologias de reconhecimento facial, por exemplo, podem reforçar discriminações e violações de privacidade. E esse trabalho, já alertou sobre os riscos discriminatórios desses sistemas, especialmente em contextos como segurança pública.

Embora o PL 21/20 também enfatize princípios como ética, transparência e responsabilidade, não aborda de maneira tão específica os riscos relacionados à segurança física e mental dos indivíduos. O projeto não possui um artigo semelhante ao artigo 14 do PL 2338, que trata diretamente da proibição de sistemas de risco excessivo. Em vez disso, o PL 21/2020 adota uma abordagem mais aberta e flexível, o que pode ser interpretado como insuficiente diante dos riscos que sistemas autônomos e altamente complexos de IA podem apresentar para a integridade física e mental dos indivíduos.

A abordagem do PL 2338/2023 é especialmente relevante porque vai além da simples criação de diretrizes gerais e fornece uma estrutura específica para mitigação de riscos. A exigência de avaliações de impacto algorítmico para sistemas de alto risco garante que os desenvolvedores identifiquem potenciais impactos negativos e tomem medidas preventivas. Isso é fundamental para evitar que sistemas de IA causem danos diretos ou indiretos às pessoas, seja em sua integridade física, ao induzi-las a comportamentos prejudiciais, ou em sua autonomia mental, manipulando-as subliminarmente.

Em síntese, a análise comparativa entre o PL 21/2020 e o PL 2338/2023 teve o objetivo de evidenciar a evolução conceitual e normativa da regulação da inteligência artificial no Brasil. Enquanto o primeiro propôs diretrizes mais genéricas e principiológicas, voltadas à promoção da inovação e à proteção de direitos fundamentais, o segundo reflete um estágio mais amadurecido do debate legislativo, incorporando elementos de governança, avaliação de impacto e classificação por níveis de risco que alinham o Brasil às tendências regulatórias internacionais.

Assim, embora o confronto entre ambos os textos tenha servido como base para compreender a trajetória de desenvolvimento das propostas regulatórias nacionais, é o PL 2338/2023 que atualmente se consolida como o instrumento legislativo mais robusto e promissor. Sua tramitação avançada e seu conteúdo técnico mais detalhado demonstram que o país caminha para um marco legal capaz de equilibrar inovação, segurança jurídica e proteção de direitos fundamentais — reforçando, portanto, que o foco da discussão contemporânea deve recair sobre esse projeto, que representa o verdadeiro eixo de maturidade normativa da regulação da IA no Brasil (Brasil, 2023).

### 3.3 INFLUÊNCIA DO GDPR NA LGPD E A CHEGADA DO AI ACT NA EUROPA: IMPACTOS PARA O BRASIL E SIMILARIDADES COM O PL 2338

A Lei Geral de Proteção de Dados (LGPD) no Brasil foi diretamente inspirada pela General Data Protection Regulation (GDPR) da União Europeia. O GDPR, aprovado em 2016 e em vigor desde 2018, foi um marco global para a proteção de dados e influenciou legislações de privacidade em todo o mundo, incluindo a LGPD. A principal semelhança entre as duas legislações é a adoção de princípios de transparência, segurança e governança de dados. Tanto o GDPR quanto a LGPD estabelecem obrigações rigorosas para o tratamento de dados pessoais e garantem direitos fundamentais aos titulares, como o direito à portabilidade, o direito de acesso e o direito à eliminação de dados.

A LGPD foi aprovada em 2018 e entrou em vigor em 2020, adaptando muitos dos conceitos estabelecidos no GDPR ao contexto brasileiro, como a figura do controlador e do operador, a necessidade de consentimento explícito e a criação de uma autoridade regulatória, no caso brasileiro, a ANPD (Autoridade Nacional de Proteção de Dados). Além disso, ambas as legislações se baseiam em princípios de minimização de dados, ou seja, garantir que apenas os dados estritamente necessários para uma finalidade específica sejam processados.

Agora, a União Europeia avança com um novo marco regulatório: o *Artificial Intelligence Act (AI Act)*. Esse é o primeiro regulamento abrangente sobre inteligência artificial, e sua estrutura é baseada em uma abordagem de regulação baseada em riscos. O *AI Act* classifica os sistemas de IA em quatro categorias de risco: risco inaceitável, alto risco, risco específico e baixo risco. O objetivo principal é garantir que os sistemas de IA respeitem os direitos fundamentais e não representem uma ameaça à segurança e aos direitos dos cidadãos.

Os sistemas de alto risco são aqueles utilizados em setores críticos como saúde, educação, e recrutamento, que devem cumprir requisitos rigorosos de transparência, qualidade

de dados e supervisão humana. Já os sistemas de risco inaceitável, como IA usada para pontuação social e discriminação biométrica, são completamente proibidos na UE. Além disso, o *AI Act* obriga os desenvolvedores a implementarem avaliações de impacto para mitigar riscos antes de colocarem sistemas no mercado

No Brasil, a discussão sobre regulamentação da IA avança com o PL 2338/2023, que compartilha muitos princípios com o *AI Act*, como a obrigatoriedade de avaliações de impacto algorítmico para sistemas de alto risco, e a proibição de técnicas subliminares e exploração de vulnerabilidades. O PL 2338 também adota um modelo de responsabilidade objetiva, semelhante ao *AI Act*, para garantir que os desenvolvedores de IA sejam responsabilizados por eventuais danos causados pelos sistemas que desenvolvem, independentemente de culpa.

A influência europeia nas legislações de proteção de dados e IA é clara. Assim como o GDPR serviu de base para a LGPD no Brasil, o *AI Act* europeu pode ser um modelo para a regulação da IA no país, com o PL 2338/2023 já refletindo muitas das ideias de governança e supervisão presentes no regulamento europeu.

Essa comparação entre modelos regulatórios, porém, exige uma reflexão crítica que vai além da análise jurídica: ela demanda consciência sobre a geopolítica do conhecimento normativo. Enrique Dussel (2005), filósofo argentino-mexicano e um dos principais pensadores da filosofia da libertação latino-americana, alerta que a adoção acrítica de modelos europeus por países do Sul Global é, em si mesma, uma forma de colonialismo, não territorial, mas epistêmico e normativo. Ao copiar marcos regulatórios construídos a partir de realidades históricas, culturais e axiológicas distintas, corremos o risco de importar também suas pressuposições, inclusive a de que o cidadão respeita as normas porque confia nelas, e não porque teme as sanções. No contexto europeu, essa confiança institucional tem raízes históricas profundas. No Brasil, como a própria trajetória da LGPD já demonstra, a lógica predominante de conformidade normativa é a da coercibilidade: empresas e indivíduos tendem a adequar seu comportamento não quando compreendem o valor da norma, mas quando percebem o risco concreto de penalidade. Esse descompasso axiológico, entre o valor intrínseco da norma e o medo da sanção, é central para entender por que a simples transposição do *AI Act* europeu para o ordenamento brasileiro pode resultar em letra morta. Paulo Freire (1987) já ensinava que a educação libertadora não pode ser imposta de fora para dentro, ela precisa emergir da realidade concreta dos sujeitos. O mesmo vale para a regulação: uma lei de IA verdadeiramente eficaz para o Brasil precisará ser construída a partir das assimetrias, das vulnerabilidades e dos valores da sociedade brasileira, e não apenas traduzida do vocabulário normativo de quem nos colonizou. Nesse sentido, vale olhar também para a experiência latino-americana: o México,

por exemplo, conta com a *Ley Federal de Protección de Datos Personales en Posesión de los Particulares* desde 2010, e países como Argentina e Chile avançam em legislações próprias sobre IA com recortes que refletem suas realidades regionais, evidenciando que a América Latina não precisa escolher entre o modelo norte-americano e o europeu, mas pode e deve construir um caminho próprio (Dussel, 2005; Freire, 1987).

### 3.4 PROPOSTA DE REFORMA DO CÓDIGO CIVIL

A proposta do novo Código Civil, atualmente em debate no Senado Federal, visa atualizar a legislação brasileira, a fim de adequá-la às complexidades do contexto digital contemporâneo. O principal objetivo dessa modernização é revisar o código vigente, promulgado em 2002, para incluir disposições específicas sobre direitos da personalidade, propriedade intelectual e proteção de dados pessoais em ambientes digitais.

Entre as inovações sugeridas, destacam-se a formulação de regras mais claras sobre o tratamento de dados pessoais e a exigência de consentimento informado para o uso de atributos da personalidade por sistemas de inteligência artificial (IA). Além disso, o anteprojeto inclui medidas para combater o roubo de identidade digital, buscando harmonizar a legislação brasileira com normativas internacionais, como o Regulamento Geral de Proteção de Dados (GDPR) da União Europeia, a fim de fortalecer a segurança jurídica e aumentar a confiança no ambiente digital.

Todavia, o capítulo dedicado ao direito digital tem gerado controvérsias entre especialistas, legisladores e membros da sociedade civil. Críticos argumentam que as medidas propostas podem ser insuficientes para acompanhar as rápidas inovações tecnológicas e as novas formas de violação de direitos no ambiente digital. Há também preocupações quanto à efetividade da implementação e fiscalização dessas normas, uma vez que as tecnologias evoluem em um ritmo que desafia as capacidades regulatórias tradicionais.

Além disso, a questão da proteção de dados pessoais e a definição dos limites para o uso de IA, de modo a não comprometer os direitos da personalidade, são particularmente controversas. Enquanto alguns defendem a necessidade de um equilíbrio maior entre inovação tecnológica e a proteção de direitos fundamentais, outros apontam que as restrições propostas podem inibir o desenvolvimento e a competitividade do setor digital no Brasil.

O impacto da IA sobre os direitos da personalidade é evidente, especialmente com o avanço de tecnologias como as deepfakes, que dificultam cada vez mais a distinção entre o que é real e o que é artificial. Isso levanta sérias preocupações sobre a proteção da identidade

individual. Um caso emblemático que ilustra essas inquietações é o da atriz Scarlett Johansson, que descobriu que o ChatGPT estava utilizando uma voz muito similar à sua, mesmo após ela ter recusado uma oferta para emprestar sua voz ao sistema. Esse incidente gerou um debate intenso sobre os limites da utilização de atributos da personalidade por sistemas de IA evidenciando a necessidade de regulamentação e transparência nesse campo (Frazão, 2024).

O caso ilustra as complexidades relacionadas à reprodução ou imitação de características pessoais, como a voz de figuras públicas, por sistemas de inteligência artificial, independentemente de essa imitação ter sido intencional ou não. No exemplo envolvendo Scarlett Johansson, a IA poderia ter sido treinada utilizando a própria voz da atriz ou alcançado um resultado semelhante por meio de outras fontes e técnicas disponíveis.

Do ponto de vista da identidade pessoal, os impactos causados pela imitação de uma voz são problemáticos, independentemente de haver uma violação direta de seus dados ou se o processo foi intencional. Em ambos os casos, os transtornos e as possíveis confusões em torno de sua identidade são inevitáveis.

Johansson, consciente da gravidade do problema, ressaltou a importância da discussão ao reconhecer que a situação vai além do uso de sua imagem ou voz, atingindo questões mais profundas relacionadas à proteção de sua identidade e trabalho.

Nesse sentido temos:

Numa época em que todos lutamos contra deepfakes e a proteção da nossa própria imagem, do nosso próprio trabalho, das nossas próprias identidades, acredito que estas são questões que merecem clareza absoluta”, disse Johansson na sua declaração. “Aguardo com expectativa uma resolução sob a forma de transparência e a aprovação de legislação apropriada para ajudar a garantir que os direitos individuais sejam protegidos.

Com efeito, como muito bem colocou a atriz, a questão não diz respeito apenas às possibilidades da tecnologia, mas à compatibilidade desta com direitos da personalidade mais elementares. A rigor, o problema existe mesmo quando estamos diante de pessoas mortas.

No caso das pessoas mortas, é interessante mencionar o projeto *The Next Rembrandt*, que começa com a seguinte provocação: “Can technology and data bring back to life one of history’s greatest painters?”[7] E o projeto avança: “The Dutch painter Rembrandt, known as the master of light and shadow, is one of the greatest painters of all time. What if we could bring him back, 347 years after his death, for one last portrait – but this time, data is the painter and technology the brush? (Frazão, 2024).

Projetos como *"The Next Rembrandt"*, que busca recriar o estilo de um dos maiores pintores da história utilizando tecnologia, também levantam questões éticas sobre o uso indiscriminado de características pessoais sem autorização. Essas preocupações se intensificam com a constatação de que IAs generativas precisam ser treinadas com grandes quantidades de

dados, resultando em litígios como o processo movido pelo The New York Times contra a OpenAI, alegando uso indevido de conteúdo protegido por direitos autorais (Pope, 2024).

No Brasil, embora a legislação vigente já assegure mecanismos de proteção dos direitos da personalidade, é fundamental garantir que não haja retrocessos no texto em debate no Congresso, especialmente em relação às garantias de prevenção e repressão de práticas abusivas no uso de tecnologias emergentes.

Essas questões éticas e jurídicas têm sido amplamente debatidas pela Comissão de Juristas responsável pela proposta de reforma do Código Civil, criada pelo Senado Federal em 2023 e presidida pelo ministro Luis Felipe Salomão, tendo como relatora a professora Judith Martins-Costa. O relatório final, entregue em abril de 2024, propõe a criação de um “Livro I-A” dedicado ao Direito Civil Digital, que abordará expressamente temas como dados pessoais, herança digital, identidade e autonomia em ambientes virtuais (Senado Federal, 2024).

A Comissão reconhece que o avanço da inteligência artificial transformou as noções tradicionais de autoria, privacidade e responsabilidade, tornando essencial a atualização do Código para abarcar novas formas de manifestação da personalidade digital. Assim, o anteprojeto busca não apenas atualizar conceitos, mas redefinir as bases da dignidade da pessoa humana no ambiente algorítmico.

Entre as principais inovações, o texto propõe reconhecer juridicamente a existência de uma “identidade digital”, cuja violação, por meio de *deepfakes*, clonagem de voz, avatares ou manipulação de dados, deve ser considerada atentado aos direitos da personalidade. Essa concepção amplia o alcance do artigo 11 do Código Civil e aproxima o Brasil das discussões internacionais sobre dignidade informacional e autodeterminação digital (Senado Federal, 2024).

A proposta também prevê mecanismos de responsabilização civil em casos de danos causados por sistemas de IA, mantendo a centralidade da conduta humana na cadeia de imputação. A Comissão optou por rejeitar a ideia de personalidade jurídica da inteligência artificial, mas propôs parâmetros para responsabilizar desenvolvedores, operadores e provedores em situações de risco decorrentes de sistemas autônomos (Senado Federal, 2024).

Essa lógica dialoga diretamente com o PL 2338/2023, que tramita em paralelo no Congresso e estabelece a regulação setorial da IA. Enquanto o PL trata dos aspectos técnicos e de governança, o novo Código Civil busca consolidar a base principiológica, garantindo coerência normativa entre as obrigações civis e as responsabilidades tecnológicas (Brasil, 2023).

Outro ponto inovador é o reconhecimento da herança digital e da transmissão *post mortem* de bens virtuais, prevendo regras sobre acesso a contas, obras e arquivos digitais. Essa previsão responde à crescente presença de identidades e patrimônios em ambientes virtuais, que exigem proteção jurídica mesmo após a morte do titular (Senado Federal, 2024).

Apesar do avanço, o capítulo digital tem recebido críticas. Especialistas alertam que a rigidez normativa pode se tornar um obstáculo frente ao dinamismo tecnológico, e que seria mais adequado adotar cláusulas abertas e linguagem principiológica, como ocorre no *AI Act* europeu. Essa crítica reforça a necessidade de uma arquitetura normativa que permita constante atualização.

A proposta brasileira, ao lado do PL 2338/2023, forma um eixo complementar semelhante ao que ocorreu na União Europeia com o *AI Act* e a revisão da *Product Liability Directive* (União Europeia, 2024). O primeiro fornece a estrutura preventiva e de governança; o segundo, os parâmetros de responsabilidade civil. Essa estratégia de separação entre “prevenção” e “reparação” também poderia constituir um caminho equilibrado para o Brasil, desde que não se incorra na armadilha de um excesso regulatório. A criação de um “Livro Digital” no Código Civil certamente marca um avanço conceitual, mas suscita o desafio de compatibilizar a estabilidade do texto codificado com a fluidez das inovações tecnológicas, evitando que o Código se torne um instrumento obsoleto diante da velocidade das transformações digitais (Senado Federal, 2024).

Em síntese, a proposta de reforma do Código Civil representa um esforço relevante de atualização diante das transformações tecnológicas, mas também revela um traço característico do ordenamento jurídico brasileiro: a tendência à hiper-regulação e à codificação exaustiva. Esse impulso de traduzir cada fenômeno emergente em dispositivos legais específicos pode gerar, paradoxalmente, entraves à inovação e à efetividade normativa, especialmente em campos dinâmicos como o da inteligência artificial. Diferentemente da União Europeia, que recentemente recuou em sua tentativa de instituir uma Diretiva de Responsabilidade Civil para IA, reconhecendo a complexidade e a volatilidade desse ecossistema, o Brasil parece seguir o caminho oposto, ampliando o corpo codificado de normas em vez de adotar estruturas mais flexíveis e principiológicas. Nesse sentido, embora a proposta do novo Código reafirme o discurso da centralidade da pessoa humana, corre o risco de cristalizar um modelo jurídico excessivamente rígido, incapaz de dialogar com a fluidez da inovação tecnológica e com os desafios que ela impõe ao Direito contemporâneo (Maranhão; Nogaroli, 2025).

Nesse contexto, a simultaneidade entre a proposta de reforma do Código Civil e a tramitação do Marco Legal da Inteligência Artificial (PL 2338/2023) suscita um debate

inevitável sobre os riscos de sobreposição normativa e retrocesso regulatório. Enquanto o Código busca incorporar princípios de proteção da personalidade digital e do consentimento informado, o PL 2338 pretende estruturar um sistema próprio de governança, risco e responsabilidade algorítmica. O problema reside em que, se ambos avançarem de maneira desarticulada, o país poderá reproduzir um modelo fragmentado e excessivamente prescritivo, o oposto do que a União Europeia acaba de reconhecer como um obstáculo à inovação, ao retirar do AI Act as disposições sobre responsabilidade civil (Maranhão; Nogaroli, 2025). Em vez de ampliar o aparato normativo, o desafio brasileiro deveria ser o de integrar seus marcos jurídicos sob uma lógica de coerência sistêmica, onde o Código Civil funcione como fundamento principiológico e o Marco Legal da IA atue como norma técnica de regulação dinâmica, evitando, assim, a estagnação diante da velocidade das mutações tecnológicas.

O Capítulo 3 demonstrou que o Brasil avança, ainda que de forma não linear, na construção de um marco regulatório para a IA. O percurso que vai do PL 21/2020 ao PL 2338/2023, e à proposta de reforma do Código Civil, revela tanto o amadurecimento do debate quanto o risco de uma hiper-regulação desarticulada. O desafio central identificado neste capítulo é o de regular sem engessar, proteger sem retrainir. Esse desafio ganha uma dimensão ainda mais delicada quando a IA deixa de ser apenas um instrumento de decisão ou automação e passa a ocupar o espaço da criação (mesmo que usada como ferramenta) território que, até recentemente, era considerado exclusivamente humano. É nessa fronteira entre o algoritmo e a criação autoral que o próximo capítulo se instala.

#### **CAPÍTULO 4 - CONTEÚDO GERADO POR IA GENERATIVA E A PROBLEMÁTICA DO DIREITO AUTORAL**

Antes de adentrar o debate jurídico sobre autoria e direitos autorais no contexto da IA generativa, é necessário um esclarecimento científico fundamental que orienta toda a análise deste capítulo. O neurocientista brasileiro Miguel Nicolelis, um dos mais respeitados do mundo e pioneiro nas pesquisas sobre interfaces cérebro-máquina, é categórico: a chamada "Inteligência Artificial" não é inteligente, nem artificial. Nicolelis cunhou o acrônimo NINA, "Nem Inteligente, Nem Artificial", para descrever com precisão o que esses sistemas realmente são. Em suas palavras, "a inteligência é uma propriedade emergente dos organismos, que surgiu para otimizar a nossa chance de sobreviver em um mundo que muda continuamente. A inteligência é uma propriedade da matéria orgânica e não é reduzível ao algoritmo, não é computável" (Nicolelis, 2024). Além disso, a IA não é genuinamente artificial: ela só existe

porque foi construída por milhões de seres humanos, treinada com dados produzidos por seres humanos e depende continuamente de seres humanos para funcionar. É, portanto, uma extensão da inteligência humana, não sua substituta, nem sua rival. Esse ponto de partida científico é decisivo para o debate jurídico que se segue: se a IA não é inteligente nem artificial, ela não pode ser autora. Ela é ferramenta. Uma ferramenta extraordinariamente poderosa, mas ferramenta. E o Direito, ao lidar com as criações que ela ajuda a produzir, precisa ter clareza sobre essa distinção para não atribuir à máquina aquilo que pertence ao ser humano que a concebeu, a orientou e assumiu responsabilidade pelo resultado.

Se a regulação da IA no campo do trabalho e da responsabilidade civil já apresenta desafios consideráveis, o terreno dos direitos autorais impõe questões de outra natureza, mais filosóficas, mais identitárias e, por isso, igualmente urgentes do ponto de vista jurídico. Quando uma máquina produz um quadro premiado, escreve um roteiro ou compõe uma música, quem é o autor? O que significa criar? E o que o Direito pode e deve proteger nesse novo cenário? Este capítulo examina a IA generativa não como ameaça à criatividade humana, mas como uma ferramenta que a tensiona, a questiona e quando bem utilizada, a potencializa. A posição defendida aqui é clara: a IA não cria. Ela amplia quem já sabe criar.

O avanço a passos largos da IA tem ocorrido, não apenas como uma ferramenta de automação de processos, mas também como um agente na criação de conteúdos que anteriormente eram exclusivos do domínio humano. Modelos generativos, como o GPT e ferramentas de criação de imagens, como o DALL·E, estão cada vez mais presentes no cotidiano das indústrias criativas, desafiando as noções tradicionais de autoria e originalidade. Este capítulo se debruça sobre os dilemas legais e éticos trazidos pelo uso de IA generativa na produção de conteúdos, com especial foco na problemática do direito autoral.

O uso da IA para gerar obras literárias, artísticas e audiovisuais levanta questões emblemáticas sobre a quem pertencem os direitos sobre essas criações. Afinal, a IA é simplesmente uma ferramenta a serviço da criatividade humana, ou podemos considerar o resultado de suas produções como uma forma de criação autônoma? A geração de conteúdo por meio de modelos de linguagem, como os *Large Language Models* (LLMs), e o uso de *prompts* (comandos) por parte dos usuários introduzem um novo paradigma no debate sobre autoria, onde o *input* humano se mistura com os resultados automatizados de um sistema.

Neste capítulo, exploraremos como os LLMs funcionam e como o papel do *prompt* afeta a criação de conteúdos, além de analisar casos práticos e as implicações jurídicas de obras geradas por IA, como o quadro premiado "Théâtre D'opéra Spatial" e a controversa capa da revista *Veja*. Ao longo da discussão, serão examinadas as questões de originalidade, plágio e os

desafios que a tecnologia impõe ao direito autoral, num cenário em que as fronteiras entre a criação humana e a inteligência artificial se tornam cada vez mais tênues.

#### 4.1 *LARGE LANGUAGE MODEL (LLM) E PROMPT*

Como vimos ao longo deste trabalho, a inteligência artificial (IA) tem desempenhado um papel cada vez mais relevante em diversas áreas, ultrapassando o campo das operações puramente técnicas e ingressando em atividades consideradas criativas e intelectuais. A IA deixou de ser vista apenas como uma ferramenta para automatização de tarefas repetitivas e começou a ser empregada na geração de conteúdo, como textos, imagens, músicas e até mesmo obras artísticas complexas. Esse fenômeno levanta importantes questões sobre a autoria e os direitos sobre o produto gerado, colocando em evidência desafios éticos e legais que precisam ser abordados.

Antes de adentrarmos no debate sobre inteligência artificial e direitos autorais, é essencial definir os conceitos de Large Language Models (LLMs) e prompts, visto que eles são a base da geração de conteúdo por IA e estão diretamente ligados às questões de originalidade e autoria no contexto digital.

##### 1) O que é LLM (Large Language Model)?

Os Large Language Models (LLMs), ou Grandes Modelos de Linguagem, são algoritmos de inteligência artificial projetados para processar e gerar linguagem natural em grande escala. Eles são treinados com enormes quantidades de dados textuais, como livros, artigos, sites, e outros conteúdos da internet, para aprender padrões, gramática, sintaxe e até significados contextuais. A partir desse aprendizado, os LLMs podem gerar respostas, textos, e realizar tarefas de linguagem de forma autônoma ou em interação com usuários (Chang, *et al.*, 2024).

Um dos exemplos mais conhecidos de LLM é o GPT (*Generative Pretrained Transformer*), que foi desenvolvido pela OpenAI. O GPT é capaz de responder perguntas, gerar histórias, traduzir textos, escrever código, entre outras funções, com base em comandos fornecidos por usuários. LLMs são fundamentais no processamento de linguagem natural (*Natural Language Processing* – NLP), sendo utilizados em *chatbots*, assistentes virtuais, tradução automática, e geração de conteúdo (Generative..., 2024).

Esses modelos são alimentados por redes neurais profundas, que analisam as relações entre palavras e frases para prever a próxima palavra ou sequência em uma frase, com base em padrões que foram previamente identificados durante o treinamento. Eles não compreendem o significado no sentido humano, mas sim operam por probabilidade, gerando respostas que "fazem sentido" de acordo com os dados em que foram treinados.

Para esclarecer ainda mais a forma como a IA opera, é importante destacar que a IA, em particular os Large Language Models (LLMs), não compreende conceitos no sentido humano, como "gato" ou "telhado", ou seja, um gerador de texto não sabe o significado do conteúdo, ele trabalha com probabilidades. É um processo de cálculo de probabilidades baseados em padrões estatísticos encontrados nos dados de treinamento. Como exemplo, ao processar uma frase como "o gato subiu no telhado", a IA não entende o que é um "gato" ou o que é um "telhado"; ela simplesmente calcula a probabilidade de que, dada a palavra "gato", a palavra "subiu" seja uma sequência comum, e que "telhado" seja uma continuação provável dessa sequência, com base nos padrões que ela aprendeu durante o treinamento. Portanto, a IA gera texto de maneira coerente, mas sem uma verdadeira compreensão semântica do conteúdo (Chang, *et al.*, 2024).

## 2) O que é um *Prompt*?

O prompt é a entrada (*input*) fornecida por um usuário para um sistema de IA, como um LLM, com o objetivo de orientar ou gerar uma saída (*output*) específica. Em termos simples, o prompt é a instrução ou pergunta que direciona a IA a realizar uma tarefa, como escrever um texto, gerar uma imagem ou responder a uma pergunta.

Por exemplo, em um LLM como o GPT, o *prompt* pode ser algo simples, como "Explique o conceito de biodiversidade", ou mais complexo, como "Escreva um conto de ficção científica ambientado em Marte, em que a IA desempenha um papel central na colonização". Dependendo do grau de especificidade e complexidade do prompt, a IA pode gerar respostas que variam em qualidade e criatividade.

Dessa forma, a geração de conteúdo através de um *prompt* fornecido por um ser humano tem gerado uma ampla discussão, pois o indivíduo ao criar um comando (*prompt*) não estaria em ressonância com um ato de criação e assim de acordo com a teoria do contributo mínimo de originalidade? Em contrapartida, o conteúdo gerado pela IA, como resultado de um processo automatizado baseado em padrões probabilísticos aprendidos, seria desprovido de originalidade ou intenção criativa? Ou o agente criador (ser humano) não estaria fazendo uso de uma

ferramenta para catalisar o processo criativo, como será abordado no subtópico 4.3 Direto Autoral, Originalidade e Plágio?

Além das questões de autoria e criatividade, surgem desafios éticos e de honestidade intelectual no uso de IA para gerar conteúdo. A facilidade com que as IA, como os LLMs, podem produzir textos, imagens e outros materiais levanta preocupações sobre plágio, apropriação indevida de ideias e a diluição da autoria humana. A linha entre o que é genuinamente criado por uma pessoa e o que é resultado de um algoritmo torna-se tênue, o que pode comprometer a transparência e a responsabilidade intelectual.

A questão central que será discutida neste tópico envolve a origem dos dados usados para treinar modelos de IA, como os LLMs, e a natureza do conteúdo gerado por eles. Esses modelos foram treinados em grandes quantidades de dados, muitos dos quais podem estar protegidos por direitos autorais. Isso levanta a discussão sobre até que ponto o uso desses dados por IA pode infringir a propriedade intelectual e se o conteúdo gerado, que se baseia em padrões aprendidos desses dados, pode ser considerado uma forma de reutilização não autorizada de material protegido.

## 4.2 AVANÇO DA IA GENERATIVA

De acordo com Bengesi *et al.* (2023), antes de 2014, todos os modelos de aprendizagem profunda existentes eram principalmente descritivos, concentrando-se em resumir ou representar padrões e relações de dados existentes. Esses modelos visavam explicar os padrões de dados e fazer previsões com base nas informações disponíveis. No entanto, em 2014, Goodfellow introduziu a Rede Generativa Adversarial (*Generative Adversarial Networks* - GAN), marcando o início de uma nova era na realização da Inteligência Artificial Generativa (*Generative Artificial Intelligence* - GAI). Ao contrário dos modelos descritivos, os modelos generativos, como as GANs, são criados para aprender a distribuição de probabilidade subjacente dos dados. Seu principal objetivo é gerar novas amostras de dados que se assemelhem aos padrões observados nos dados de treinamento (Bengesi *et al.*, 2023).

O avanço das GANs (Redes Gerativas Adversárias) representou uma significativa ruptura com os métodos tradicionais de aprendizado profundo, abrindo possibilidades empolgantes para a Inteligência Artificial Generativa. Desde então, a GAI (Inteligência Artificial Generativa) ganhou atenção generalizada devido ao seu impacto transformador em diversos domínios da vida. Ela oferece soluções elegantes para problemas complexos, permitindo a criação de dados sintéticos, conteúdo artístico e simulações realistas. Essa mudança de paradigma na tecnologia de IA influenciou profundamente a nova percepção, implementação e utilização da

inteligência artificial, impulsionando inovações e novas oportunidades de aplicação em várias indústrias (Bengesi *et al.*, 2023 - tradução nossa).

Desse modo, se compreende que o avanço da IA em processos criativos ocorre, em grande parte, devido ao desenvolvimento de algoritmos de aprendizado de máquina (*machine learning*) e redes neurais, que permitem que as máquinas aprendam padrões e comportamentos a partir de grandes volumes de dados. Assim, sistemas de IA, como redes neurais generativas (GANs) e modelos de linguagem como o GPT (*Generative Pretrained Transformer*), são capazes de gerar conteúdo de maneira autônoma ou semiautônoma, aproximando-se cada vez mais do que consideramos criatividade humana (Bengesi *et al.*, 2023). Esses avanços trazem benefícios inegáveis para a sociedade, permitindo a inovação em diversas áreas, como marketing, design, educação e até mesmo na indústria do entretenimento, mas por outro lado, também são grande fonte de preocupação.

#### 4.3 DIREITO AUTORAL, ORIGINALIDADE E PLÁGIO

A existência do campo do direito que regula o Direito Autoral tem o intuito de criar obras culturais, como livros, fotos, músicas e filmes. Esse direito faz isso dando uma exclusividade para que o autor (ou o titular desses direitos) possa por um certo tempo decidir como a obra pode ser utilizada.

Desse modo, entende-se que o direito autoral é um ramo do Direito Privado, responsável por proteger as obras intelectuais e os chamados direitos conexos, que abrangem direitos de intérpretes, executantes, produtores e emissoras. Esse conceito é parte da Propriedade Intelectual, que engloba os direitos relacionados às criações humanas, como obras literárias, artísticas e científicas.

De acordo com a Lei 9.610, de 19 de fevereiro de 1998, que regulamenta os direitos autorais no Brasil, uma obra está sob proteção quando é expressa por qualquer meio ou fixada em qualquer suporte, tangível ou intangível. Isso indica que a proteção dos direitos autorais abrange não apenas livros, músicas ou pinturas, mas qualquer forma de expressão que possa ser percebida, como obras digitais, coreografias, jogos eletrônicos, e outros:

Art. 7º São obras intelectuais protegidas as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro, tais como:

I - os textos de obras literárias, artísticas ou científicas;

II - as conferências, alocações, sermões e outras obras da mesma natureza;

III - as obras dramáticas e dramático-musicais;

- IV - as obras coreográficas e pantomímicas, cuja execução cênica se fixe por escrito ou por outra qualquer forma;
- V - as composições musicais, tenham ou não letra;
- VI - as obras audiovisuais, sonorizadas ou não, inclusive as cinematográficas;
- VII - as obras fotográficas e as produzidas por qualquer processo análogo ao da fotografia;
- VIII - as obras de desenho, pintura, gravura, escultura, litografia e arte cinética;
- IX - as ilustrações, cartas geográficas e outras obras da mesma natureza;
- X - os projetos, esboços e obras plásticas concernentes à geografia, engenharia, topografia, arquitetura, paisagismo, cenografia e ciência;
- XI - as adaptações, traduções e outras transformações de obras originais, apresentadas como criação intelectual nova;
- XII - os programas de computador;
- XIII - as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual (Brasil, 1998).

Na legislação brasileira, a figura do autor está intrinsecamente ligada ao conceito de inspiração (criação de espírito): uma pessoa, denominada autor, é movida por inspirações diversas para criar uma obra. Contudo, ao transferirmos essa lógica para o universo dos *chatbots*, emergem questionamentos sobre a capacidade de uma máquina se 'inspirar' e, conseqüentemente, de ser considerada autora no sentido tradicional estabelecido pela lei. A lei brasileira de direitos autorais não define explicitamente o "contributo mínimo" de criatividade (inspiração) para proteção, mas jurisprudência e doutrina sugerem que uma obra deve refletir a personalidade do autor para ser protegida.

O direito autoral abrange dois componentes principais: o *corpus mysticum*, que representa a ideia abstrata da obra, e o *corpus mechanicum*, que se refere à materialização dessa ideia em um suporte tangível, como um livro ou uma pintura. Assim, a proteção surge no momento em que a obra é expressa ou materializada, assegurando ao autor tanto os direitos morais (como a atribuição e a integridade da obra) quanto os direitos patrimoniais, que envolvem o uso econômico da criação (Wachowicz; Biancamano, 2011).

Os direitos morais no contexto do direito autoral têm origem na doutrina francesa e são baseados na relação criativa entre o autor e sua obra. Esses direitos protegem o autor em aspectos não econômicos, como o direito à paternidade da obra e à preservação de sua integridade. Já os direitos patrimoniais são focados na exploração econômica da obra, permitindo que o autor controle como a criação será utilizada e comercializada. Dessa forma, o direito autoral serve tanto para reconhecer a autoria e garantir a integridade da obra quanto para permitir ao autor explorar economicamente sua criação, oferecendo um equilíbrio entre a dimensão pessoal e patrimonial da produção intelectual (Wachowicz; Biancamano, 2011).

Nesse contexto, a proteção dos direitos autorais está intrinsecamente ligada ao conceito de originalidade, que exige a criatividade do autor, que é um elemento essencial para diferenciar

uma criação autêntica de práticas de plágio, onde a apropriação indevida da expressão criativa de outrem pode comprometer tanto a integridade moral quanto o valor econômico da obra.

O conceito de originalidade no direito autoral refere-se à expressão única e pessoal de uma ideia pelo autor, um elemento essencial para garantir a proteção de uma obra. No entanto, para que uma criação seja considerada original, não é necessário que seja totalmente nova ou revolucionária, mas sim que traga um contributo mínimo de originalidade, isto é, que a obra carregue características que a diferenciem de outras criações preexistentes. O direito autoral não protege ideias abstratas, mas a forma como essas ideias são expressas e materializadas. Isso significa que, desde que a ideia seja exposta de uma forma nova ou pessoal, ela pode ser objeto de proteção autoral (Wachowicz; Biancamano, 2011).

Por outro lado, o plágio ocorre quando uma obra é apropriada sem a devida atribuição, e o infrator se apropria da expressão criativa de outra pessoa, apresentando-a como sua. Plágio não é apenas a cópia exata de um texto ou obra, mas também pode ser configurado quando há uma reprodução substancial da essência criativa de uma obra, mesmo que modificada ou disfarçada. No contexto acadêmico e cultural, identificar o plágio pode ser uma tarefa complexa, especialmente em áreas onde as criações tendem a se repetir ou a seguir padrões estilísticos semelhantes, como nas indústrias criativas (Wachowicz; Biancamano, 2011).

A ideia de contributo mínimo de originalidade é fundamental para distinguir o plágio de uma simples coincidência criativa. Como apontado por Wachowicz e Biancamano (2011), muitas vezes, as obras culturais têm pontos em comum devido à natureza colaborativa e evolutiva da cultura. No entanto, o plágio se configura quando se observa que o autor não trouxe nenhum contributo pessoal significativo para a obra, limitando-se a replicar, de maneira substancial, o conteúdo ou a forma de uma criação pré-existente. Esse entendimento é essencial para que o direito autoral proteja a criatividade sem sufocar a liberdade de expressão e o intercâmbio cultural.

Dessa forma, o direito autoral garante a proteção dos direitos do autor, mas não de forma tão restritiva a ponto de impedir que outros criadores possam usar ou se inspirar em obras preexistentes para criar novas formas de arte ou conhecimento.

No ambiente cultural e artístico, é comum que as obras sejam transformadas, reinterpretadas e adaptadas, e se o direito autoral fosse aplicado de maneira excessivamente rígida, essas práticas poderiam ser sufocadas, limitando o desenvolvimento criativo. Ao reconhecer o contributo mínimo de originalidade e ao evitar a qualificação de qualquer uso como plágio, o direito autoral protege o autor sem limitar indevidamente a liberdade de outros criadores e o intercâmbio cultural (Wachowicz; Biancamano, 2011).

Portanto, a ideia central é que o direito autoral precisa ser equilibrado para que proteja os direitos do autor, mas ao mesmo tempo permita que a cultura e o conhecimento circulem e evoluam, evitando uma sobrecarga de restrições que inviabilize o crescimento cultural e criativo.

A discussão sobre o plágio e o contributo mínimo de originalidade ganha ainda mais relevância no contexto da economia criativa e das tecnologias emergentes, como a inteligência artificial (IA). Ferramentas de IA, como os geradores de texto e imagem, colocam novos desafios sobre o que pode ser considerado original ou criativo, uma vez que essas tecnologias são treinadas com base em obras existentes. Neste cenário, a distinção entre uma nova criação e uma simples combinação ou apropriação de obras anteriores torna-se ainda mais tênue, exigindo uma reflexão cuidadosa sobre o papel dos direitos autorais na proteção da criatividade em um ambiente cada vez mais digital (Wachowicz; Biancamano, 2011).

Nesse contexto a criatividade humana passa a ser contestada. E o que seria a criatividade?

A criatividade, segundo Windeløv-Lidzélius, continua sendo o que sempre foi desde os primórdios da humanidade: uma forma de lidar com o mundo, resolver problemas e criar ideias, gestos e propostas, de modo a satisfazer a necessidade interna de dar sentido, reagir com significado e se relacionar (Polo, 2024). Vinícius Mano, professor da PUCRS, complementa que todo ser humano possui uma cota natural de criatividade, ainda que alguns apresentem mais habilidades devido a fatores como contexto, referências e ambiente. Ele cita Domenico De Masi, sociólogo que propôs o conceito de "ócio criativo", para destacar que uma ideia precisa passar do pensamento para a incorporação de elementos concretos. Mano defende a noção contemporânea de "atitude criativa", que exige persistência, disciplina e sacrifício, com limites e prazos favorecendo o processo criativo (Polo, 2024).

Windeløv-Lidzélius afirma que a criatividade é uma narrativa da vida em evolução, ligada à sobrevivência. Diogo Cortiz, professor da PUC-SP, ressalta que a criatividade está presente não apenas nos processos artísticos, mas também nas soluções diárias para problemas triviais. Ele destaca a importância de manter o controle sobre as ferramentas tecnológicas, como a inteligência artificial (IA), para evitar que as pessoas se tornem dependentes e preguiçosas, correndo o risco de comprometer seu potencial criativo (Polo, 2024).

A noção contemporânea de criatividade se afasta da ideia de liberdade total e passa a ser vista como uma atitude criativa. Isso significa que a criatividade está intrinsecamente ligada à ação: é necessário agir, investir energia e disciplina no processo criativo. Vinícius Mano argumenta que, ao persistir nas ideias, não as abandonar, e buscar referências constantemente,

é possível desenvolver essa atitude criativa. A verdadeira criatividade, portanto, exige mais do que talento inato – ela demanda prática, foco e motivação (Polo, 2024).

No entanto, assim como sempre existiram pessoas interessadas e criativas, também sempre houve, e sempre haverá, aqueles que são desinteressados e apresentam má-vontade. Talvez a real preocupação com o impacto da tecnologia na criatividade esteja justamente nesse grupo da sociedade, que pode tender a utilizar as facilidades tecnológicas como desculpa para a falta de desinteresse da própria vida. Em um cenário onde as ferramentas tecnológicas oferecem soluções rápidas e prontas, o risco é que esse segmento da população se acomode, delegando suas responsabilidades criativas às máquinas, sem desenvolver uma atitude criativa genuína. A preguiça e a falta de dedicação podem, assim, ser amplificadas pelo uso desmedido da tecnologia, comprometendo a inovação e o potencial criativo (Polo, 2024).

Com esse entendimento, percebe-se que a tecnologia já impacta a criatividade de diversas maneiras. Ferramentas de IA podem acelerar processos e servir como aliadas, mas também podem levar ao encurtamento de repertórios e à dependência de modelos de recomendação. Para Cortiz, a IA pode elevar o processo criativo a novos patamares, mas há o risco de acomodação humana se delegarmos demais à máquina (Polo, 2024). Mano também adverte que o excesso de tecnologia pode atrapalhar a profundidade do processo criativo, resultando em criações vazias e superficiais (Polo, 2024).

A discussão sobre a coautoria entre humanos e IA é complexa, mas essencial. Segundo Cortiz, os humanos devem evitar a preguiça e utilizar a IA como parceira no processo criativo, sem abrir mão do controle sobre a produção (Polo, 2024). Assim, a IA não deve substituir a criatividade humana, mas servir como ferramenta para amplificá-la, desde que as pessoas mantenham sua intencionalidade e disciplina (Polo, 2024).

Abaixo segue um quadro, que partiu de uma ideia inicial particular minha, sobre a representação de inspiração e criação, e confesso, que consegui a imagem desejada somente depois de muitos comandos (*prompts*) e devaneios. Nele um senhor está retratado em um ambiente que remete à grandeza e ao misticismo da Biblioteca de Alexandria. Sentado em meio aos pergaminhos e livros que simbolizam o conhecimento ao longo dos séculos, ele se vê diante de uma nova criação. No entanto, a origem dessa criação não vem apenas do conhecimento que o cerca, mas de uma força invisível e intangível que a inspira.

A inspiração é, por si só, um conceito misterioso que foi amplamente debatido ao longo da história. De onde vem essa centelha criativa? Em tempos antigos, acreditava-se que ela provinha das musas, divindades gregas responsáveis por inspirar poetas, músicos e artistas. Platão acreditava que o verdadeiro artista era apenas um veículo para as ideias divinas. Séculos

mais tarde, essa mesma questão continuou a intrigar filósofos e artistas, com teorias que variam entre o inconsciente freudiano e as forças externas, divinas ou espirituais (Perkoski, 2013).

Figura 4 - Quadro: “A Arte do Pensar” prompts de Ana Mahle



Fonte: Imagem criada pelo ChatGPT pelo *prompt* de Ana Mahle (produzida em 2024)

Ao transportar essa ideia para o ambiente imaginário da Biblioteca de Alexandria, o senhor que se encontra nela é inspirado por uma figura invisível, talvez uma musa clássica ou uma nova forma de inspiração. A figura da biblioteca representa o conhecimento humano acumulado, enquanto o ser invisível simboliza a fonte intangível e transcendental que guia o ato criativo.

Com o avanço da tecnologia, em especial a inteligência artificial, começamos a nos perguntar: a tecnologia poderia ocupar esse papel de musa? Ao criar quadros, músicas ou textos, as IAs estão se tornando cada vez mais habilidosas em simular o processo criativo humano, mas no fim das contas, a IA funciona como uma intermediária, transformando *input* humano em algo novo, assim como as musas transformavam as aspirações humanas em obras sublimes.

Neste contexto, a tecnologia poderia ser vista como um novo tipo de “divindade” (ferramenta) inspiradora, uma musa contemporânea que não cria por si só, mas que auxilia o ser humano a canalizar suas ideias e visões. O processo de criação com IA exige uma série de

ajustes, tentativas e falhas, até que o resultado desejado seja alcançado, e assim como na relação com as musas, o resultado final pode parecer algo além da intenção inicial do artista.

Como no exemplo em tela, a ideia inicial estava presente na pessoa humana (a autora deste trabalho) e a IA a executou como se fosse um pincel (ferramenta/instrumento). É óbvio que existem questões muito complexas como: Quais dados estão sendo inseridos para o treinamento da IA? Mas o objetivo desse trabalho não é esgotar essas questões, e tampouco tenho a pretensão de achar que isso é possível, mas sim de fomentar a reflexão de algo que já está incorporado no dia a dia das pessoas.

#### 4.4 CASOS PRÁTICOS DE DIREITOS AUTORAIS E CONTEÚDOS GERADOS POR IA

A imagem abaixo, ganhou um prêmio por ficar em primeira colocada na categoria de arte digital na Feira Estadual do Colorado.

Figura 5 - Théâtre D'opéra Spatial by Jason Allen



Fonte: Ghosh; Fossas, (2022)

Jason Allen, um designer de jogos com base no Colorado, causou agitação na comunidade artística ao ganhar o prêmio principal de arte digital na Feira Estadual do Colorado. Ele revelou que sua obra vencedora, *Théâtre D'opéra Spatial*, foi criada utilizando o *Midjourney*, um modelo de arte generativa de inteligência artificial que produz imagens a partir de instruções de texto.

Segundo Gosh e Fossas (2022), em seu artigo intitulado “*Can There be Art Without an Artist?*”, o debate subsequente na comunidade artística ressaltou as tensões entre o avanço

tecnológico no campo da arte digital e o impacto do aprendizado de máquina, que está transformando rapidamente o cenário atual. Essa mudança veloz está afetando os artistas digitais economicamente, à medida que o controle e os lucros migram dos artistas para os detentores dos modelos e da infraestrutura computacional. Embora Allen tenha admitido a sua participação ao inserir as instruções de texto que originaram a obra vencedora, esse esforço artístico é significativamente restrito em comparação aos demais artistas competidores da categoria. É possível imaginar um futuro em que até mesmo as instruções de texto sejam geradas por modelos de linguagem, o que poderia desumanizar completamente o processo criativo artístico, distorcendo profundamente a percepção humana do significado por trás de uma imagem. Isso nos leva a refletir: será possível existir arte sem um artista? (Ghosh; Fossas, 2022)

Entre os principais desafios que acompanham o uso de IA generativa está a questão dos direitos autorais. Tradicionalmente, a proteção de obras criativas pertence ao seu autor humano. Entretanto, o uso de IA para gerar conteúdo levanta dúvidas sobre quem realmente detém os direitos autorais: o desenvolvedor da IA, o usuário que utiliza a ferramenta, ou a própria IA, que opera de forma autônoma? Como destacado por Ghosh e Fossas (2022) no artigo já mencionado, “*Can There be Art Without an Artist?*”, as imagens usadas para treinar os modelos de IA, como DALL.E, *Midjourney* e *Stable Diffusion*, são frequentemente extraídas da internet sem consideração pelos direitos autorais ou licenças associadas. Os autores explicam que modelos como LAION-5B ignoram flagrantemente os direitos dos criadores originais ao utilizar suas obras sem permissão, resultando em uma reprodução não licenciada, muitas vezes literal, do trabalho artístico (Ghosh; Fossas, 2022).

Essa apropriação indevida de obras artísticas por IA gera um vácuo jurídico, especialmente em contextos como o brasileiro, onde a legislação de direitos autorais (Lei 9.610/98) ainda não prevê explicitamente a proteção de obras geradas por máquinas. Essa lacuna precisa ser preenchida para garantir clareza e justiça nas relações entre criadores humanos e os novos modelos de criação gerados por IA (Souto, 2011). Além disso, como mencionado por Ghosh e Fossas, os artistas estão perdendo não apenas o controle sobre suas criações, mas também os lucros, uma vez que as empresas proprietárias dos modelos cobram pelo uso dessas ferramentas, enquanto os criadores originais não são compensados (Ghosh; Fossas, 2022).

Outro ponto relevante é a questão da honestidade intelectual. A utilização de IA para gerar conteúdo levanta preocupações éticas, especialmente quando o uso dessas ferramentas não é declarado em trabalhos acadêmicos ou criativos. A originalidade, a criatividade e a

propriedade intelectual, conceitos centrais para a integridade acadêmica e artística, são colocados em risco. Ghosh e Fossas (2022) destacam que, enquanto modelos geradores de arte podem se transformar em uma nova modalidade artística, seu uso indiscriminado pode desumanizar o processo criativo, obscurecendo a verdadeira autoria e contribuindo para uma crise de deslocamento econômico no setor artístico.

Há vários outros exemplos além da imagem intitulada *Théâtre D'opéra Spatial*, que acompanham o uso de IA na criação artística que colidem com a questão dos direitos autorais e princípios éticos. O caso da escritora Rie Kudan, que afirmou ter usado o ChatGPT para ajudar a escrever parte de sua obra premiada, levanta novos debates sobre autoria e originalidade.

Rie Kudan durante seu discurso de agradecimento pelo prêmio, mencionou ter explorado todo o potencial da inteligência artificial na criação da obra, revelando que mantém conversas frequentes com o sistema, chegando até a compartilhar segredos íntimos. Segundo ela, as respostas fornecidas pelo ChatGPT serviram de inspiração para alguns diálogos do romance (Lin, 2023).

Intitulado "Tokyo-to Dojo-To", ou "A Torre da Compaixão de Tóquio", o livro narra a história de uma prisão situada em uma torre em uma Tóquio futurista. Durante a trama, descobre-se que o edifício é um projeto de uma arquiteta descontente com a sociedade daquela era. A obra aborda o tema da inteligência artificial e sua interação com a humanidade (Lin, 2023).

Embora apenas 5% de sua obra tenha sido gerada pela IA, a questão sobre a divisão de crédito entre o autor e a ferramenta permanece. Esses exemplos destacam a urgência em discutir a proteção de direitos autorais em um contexto onde IAs estão cada vez mais integradas nos processos criativos. A legislação de direitos autorais brasileira (Lei 9.610/98) como já extensivamente abordado, não prevê a proteção de obras geradas por máquinas, deixando um vácuo jurídico que precisa ser preenchido. Como apontado no artigo de Ghosh e Fossas (2022), os modelos de IA, como o DALL.E e o Midjourney, são treinados com base em obras de artistas sem a devida atribuição ou consentimento, gerando preocupações de plágio e violação de direitos autorais. A IA, em muitos casos, não apenas imita estilos artísticos, mas reproduz obras de forma praticamente idêntica, o que pode ser interpretado como uma reprodução não licenciada.

O caso Jason Allen merece análise jurídica mais detida, pois concentra em si os três conceitos que o direito autoral precisa distinguir com precisão neste novo contexto: originalidade, criatividade e autoria.

**Há originalidade na obra?** Em sentido jurídico, originalidade não exige genialidade, exige apenas que a obra não seja cópia servil de outra e que traga o mínimo de expressão pessoal do autor. O "contributo mínimo de originalidade", critério consagrado na doutrina brasileira e internacional, é um padrão deliberadamente baixo. No caso de Allen, a questão é se os prompts que ele elaborou, as instruções textuais que guiaram o *Midjourney*, atingem esse patamar mínimo. Se os prompts foram genéricos, a resposta tende a ser negativa. Se foram específicos, detalhados e resultado de escolhas estéticas conscientes, composição, paleta, atmosfera, referências históricas, há argumento razoável para reconhecer originalidade na expressão do prompt, ainda que não na imagem gerada autonomamente pela IA.

**Há autoria?** A autoria, em sentido jurídico, é um vínculo específico entre uma pessoa humana e uma obra, não uma categoria moral ampla, mas uma relação jurídica que atribui direitos e responsabilidades. No Brasil, a Lei 9.610/98 reconhece como autor apenas a pessoa física criadora de obra intelectual. A IA, por definição, não pode ser autora. A questão relevante é outra: em que medida o ser humano que usa a IA como ferramenta pode reivindicar autoria sobre o resultado? A resposta depende do grau de controle criativo exercido. Quando o humano define a intenção, seleciona, edita, rejeita e aprova, como um diretor que não opera a câmera mas é o autor do filme, há base para reconhecer autoria humana. Quando o humano apenas aciona a máquina com um comando genérico e aceita o primeiro resultado gerado, essa base se fragiliza.

Qual o papel dos prompts de terceiros? O *Midjourney* foi treinado com obras de artistas que podem não ter consentido esse uso. Os *prompts* de Allen ativaram padrões aprendidos a partir de obras alheias. Isso não invalida automaticamente a criação, toda arte se constrói sobre referências, mas levanta uma questão ética e jurídica legítima: o processo de treinamento dos modelos respeita os direitos dos criadores originais? Essa é uma lacuna que o direito autoral brasileiro ainda não respondeu adequadamente.

A posição defendida nesta tese é clara: a IA não cria, ela amplifica quem já sabe criar. Essa afirmação não é apenas filosófica, é empiricamente sustentada. Pesquisa recente conduzida por Luchini, Kaufman e Beaty (2025), da *Pennsylvania State University* e da *University of Connecticut*, com 442 participantes em dois estudos independentes, demonstrou que as diferenças individuais em criatividade e inteligência persistem mesmo quando todos têm acesso às mesmas ferramentas de IA. Pessoas mais criativas sem IA também produzem resultados mais originais com IA, a tecnologia não equaliza, ela amplifica. Nas palavras dos autores, " *A IA generativa não apaga as diferenças individuais na criatividade humana*" (Luchini; Kaufman; Beaty, 2025). Mais do que isso: os estudos isolaram pela primeira vez um

fator latente de "criatividade assistida por IA", estruturalmente separável da criatividade sem IA, o que confirma que colaborar bem com a IA é, em si, uma habilidade cognitiva distinta, que depende de quem o humano já é antes de acionar a ferramenta.

Platão acreditava que a criação não vinha do artista, mas através dele, as Musas eram entidades externas que sopravam inspiração na mente humana; o poeta era apenas o canal. Séculos depois, Steven Pressfield (2002) retomou essa ideia em *A Guerra da Arte*: a criatividade é uma força que existe além de nós, esperando ser convocada por quem tem disciplina e preparo para recebê-la. A IA, quando usada com intenção, pensamento crítico e domínio do próprio campo, não substitui o criador, ela age como aquelas entidades que os antigos já descreviam: um agente externo que amplifica o que já existe dentro do humano. Como o acaso de Pasteur, a IA não revela seu potencial a qualquer um, ela favorece a mente preparada, aquela que sabe o que perguntar, o que avaliar e o que descartar. No caso Jason Allen, o que está em jogo não é se a IA criou, ela não criou. O que está em jogo é se a intervenção humana foi suficiente, intencional e original o bastante para merecer proteção jurídica. Essa é a pergunta certa. E é ela que o direito autoral precisará aprender a responder.

Uma última questão merece registro: os modelos generativos como o *Midjourney* são treinados com bilhões de dados extraídos da internet, frequentemente sem o consentimento dos criadores originais e sem qualquer mecanismo de atribuição ou rastreabilidade. A IA não cita fontes, não porque se recuse, mas porque sua arquitetura não foi projetada para isso. Ela opera por probabilidades estatísticas, não por memória de origem. Isso significa que o resultado gerado por Allen pode conter padrões estéticos de dezenas de artistas que jamais souberam que suas obras alimentaram aquele modelo. Essa é a dimensão mais delicada do debate: não está apenas em jogo a autoria do produto final, mas a justiça do processo que o tornou possível. E é exatamente essa lacuna, entre o que a tecnologia faz e o que o direito ainda não regula, que torna o caso Jason Allen não apenas um episódio curioso da história da arte digital, mas um caso paradigmático para o direito autoral do século XXI.

Se a questão da autoria e originalidade no caso Jason Allen revela os limites do direito autoral diante da criação assistida por IA, o problema do treinamento desses modelos expõe uma camada ainda mais estrutural do conflito: o que acontece antes da criação, quando os modelos são alimentados com obras alheias em escala industrial? Essa é a questão que o tópico seguinte examina, o uso de obras protegidas para mineração de dados, e é onde o debate jurídico se torna mais urgente, porque envolve não apenas a proteção de uma obra específica, mas a sustentabilidade de todo o ecossistema criativo que alimenta a própria IA.

#### 4.5 USO DE OBRAS PROTEGIDAS POR DIREITOS AUTORAIS PARA MINERAÇÃO DE DADOS

No Japão, a legislação permite que obras protegidas por direitos autorais sejam utilizadas para mineração de dados, um processo essencial para o treinamento de modelos de inteligência artificial. Essa abordagem posicionou o país à frente nas discussões globais sobre o tema, demonstrando interesse em equilibrar inovação tecnológica com a proteção dos direitos dos criadores. Conforme Flynn et al. (2020), o uso de obras protegidas para fins de mineração de dados pode ser entendido como uma forma de "uso justo", desde que respeite os direitos morais e patrimoniais dos autores. Desde 2009, a legislação japonesa foi atualizada para permitir essa prática, impulsionando o aprendizado de máquina sem a necessidade de autorização prévia. Segundo Carlos Affonso de Souza (2023), essa flexibilização permitiu que o Japão se destacasse no desenvolvimento de IA, ao mesmo tempo em que buscava se manter competitivo em um cenário global dominado por Taiwan e China no campo tecnológico.

A legislação japonesa divide a operação das ferramentas de IA generativa em duas fases: desenvolvimento/aprendizado e geração/uso de conteúdo novo. Na primeira fase, obras protegidas podem ser usadas para treinar modelos de IA sem a necessidade de autorização do autor, desde que o uso se limite ao necessário e não prejudique injustamente os interesses dos titulares de direitos autorais. Na segunda fase, quando o conteúdo é efetivamente criado pela ferramenta de IA, as normas sobre infrações aos direitos autorais voltam a valer normalmente, ou seja, se o texto gerado pela IA reproduzir trechos do autor original, pode ser considerado plágio; e nada impediria que o autor cujas obras "inspiraram" o modelo reconheça seus trabalhos na criação resultante e busque indenização (Souza, 2023). Nesse sentido temos:

A Agência de Assuntos Culturais, do governo japonês, divulgou um documento que divide a operação das ferramentas de IA generativa em duas fases:

- 1) A primeira é dedicada ao "desenvolvimento/aprendizado de IA";
- 2) A segunda é focada na "geração/uso" de um novo conteúdo.

Na primeira fase, obras protegidas por direitos autorais podem ser usadas em precisar passar pela autorização do autor ou do titular dos direitos autorais. O Japão confirma assim a sua posição no sentido de estimular o desenvolvimento de novas ferramentas de IA, embora o documento faça a ressalva de que esse uso deve ser feito dentro "da extensão necessária" e sem "prejudicar injustamente os interesses" do titular dos direitos autorais, e na segunda fase, quando o conteúdo é efetivamente criado pela ferramenta de IA, o documento esclarece que as normas sobre infrações aos direitos autorais valem normalmente. Ou seja, se o texto criado pela aplicação de IA reproduzir trechos do autor original ele pode ser considerado plágio.

E com as imagens? Nada impediria que o autor cujas obras "inspiraram" a IA reconheça os seus trabalhos na imagem nova e busque uma indenização.

Essas preocupações com a segunda fase aparecem também em recente documento do Ministério da Justiça de Israel sobre o uso de obras protegidas por direitos autorais no treinamento de aplicações de IA.

O documento reflete o mesmo entendimento defendido pelo governo japonês sobre a possibilidade de se usar obras autorais na construção de bases de dados que alimentam aplicações de IA generativa sem passar por qualquer autorização prévia. Um ponto diferente na proposta israelense é que essa exceção não se aplicaria.

. Essa divisão é fundamental, como destaca Sardinha (2023), para garantir que o progresso tecnológico não prejudique os criadores. Preocupações semelhantes aparecem em documento do Ministério da Justiça de Israel, que reflete entendimento equivalente ao japonês quanto ao uso de obras autorais na construção de bases de dados para IA generativa, embora com limitações distintas quanto ao escopo da exceção.

No Brasil, a Constituição Federal, em seu artigo 5º, inciso XXIX, garante a proteção dos direitos autorais em equilíbrio com o desenvolvimento tecnológico e econômico do país:

XXIX – a lei assegurará aos autores de inventos industriais privilégio temporário para sua utilização, bem como proteção às criações industriais, à propriedade das marcas, aos nomes de empresas e a outros signos distintivos, tendo em vista o interesse social e o desenvolvimento tecnológico e econômico do País (Brasil, 1988).

A Lei Federal de Direitos Autorais (Lei 9.610/98), contudo, não prevê o avanço da IA nem o uso de mineração de dados, criando um vácuo jurídico significativo. Como destaca Lucchi (2023), a ausência de clareza jurídica nesse campo pode restringir tanto o desenvolvimento tecnológico quanto a proteção dos direitos dos criadores. Essas tensões se evidenciam no processo judicial do New York Times contra a OpenAI e a Microsoft, no qual o jornal alega que suas obras foram utilizadas para treinar modelos de IA sem autorização, prejudicando sua receita e sua relação com os leitores (Lopes, 2023). O caso demonstra que, enquanto a mineração de dados pode ser considerada legal durante o treinamento, o conteúdo gerado pode infringir direitos autorais se incorporar partes substanciais das obras originais — os modelos de IA, segundo o processo, são capazes de gerar saídas que repetem ou resumem de forma próxima o conteúdo do jornal e imitam seu estilo expressivo (Souza, 2023). Essa tensão também ressoa no Japão, especialmente nas indústrias de anime e mangá, que veem a flexibilização da mineração de dados como uma ameaça à criatividade e ao retorno financeiro do setor, mesmo que o cientista-chefe de IA da Meta, Yann LeCun, tenha classificado o país como um "paraíso para o aprendizado de máquina" (Souza, 2023).

O exemplo japonês demonstra que é possível harmonizar o direito ao desenvolvimento científico e tecnológico com a proteção dos direitos autorais, desde que haja clareza normativa sobre os limites de cada fase do processo criativo. Acadêmicos como Christophe Geiger destacam que o uso de obras para treinamento de IA pode ser enquadrado como uso justo, desde

que estabelecidas limitações precisas para proteger os autores. Nesse cenário, a Convenção de Berna e os tratados internacionais de propriedade intelectual oferecem um ponto de partida para a construção de um regime jurídico que contemple o avanço tecnológico sem desproteger os criadores. Para o Brasil, o caminho passa necessariamente por uma reforma legislativa ou regulamentar que supere a lacuna da Lei 9.610/98, sem ela, o desenvolvimento de IA continuará operando em zona cinzenta legal, colocando em risco tanto a inovação quanto a remuneração justa dos autores cujas obras alimentam esses sistemas.

#### 4.6 CRIADOR, CRIATURA E FERRAMENTA: AUTORIA DA CRIAÇÃO GERADA POR INTELIGÊNCIA ARTIFICIAL GENERATIVA

A questão da autoria no contexto da Inteligência Artificial (IA) Generativa representa um debate imprescindível no campo do direito autoral e da propriedade intelectual. Historicamente, a autoria é atribuída ao ser humano, cuja capacidade criativa e originalidade garantem a proteção legal de suas obras. Contudo, com o surgimento de IAs como DALL.E, Midjourney e ChatGPT, que podem criar textos, imagens e outras formas de arte, a distinção entre o criador, a obra gerada e a ferramenta (a IA) tornou-se menos clara. A pergunta que surge é: quem deve ser considerado o autor legítimo? O desenvolvedor da IA, o usuário que dá as instruções, ou a própria IA, que opera de forma autônoma?

Para abordar essa questão, é necessário analisar o papel de cada agente no processo criativo. O desenvolvedor da IA é responsável por construir a ferramenta, programando algoritmos que permitem à máquina criar conteúdo com base em padrões aprendidos durante seu treinamento. No entanto, ele não tem controle direto sobre as criações específicas da IA. O usuário, por outro lado, insere os prompts ou comandos que orientam a IA na geração do conteúdo. Embora tenha uma participação criativa ao fornecer diretrizes, o usuário muitas vezes apenas dá o ponto de partida, e a IA executa a maior parte do processo criativo. Já a IA, apesar de ser uma ferramenta sem consciência, desempenha um papel fundamental ao combinar e reinterpretar dados de forma autônoma, gerando o produto final.

Diante dessa problemática, a discussão sobre direitos autorais em obras geradas por IA envolve o entendimento sobre quem detém os direitos autorais dessas obras. Existem várias possibilidades a serem consideradas:

- Desenvolvedor da IA: Os direitos poderiam ser atribuídos ao desenvolvedor do sistema de IA, sob o argumento de que eles criaram a ferramenta que gerou a obra.

Isso seguiria a lógica de que o desenvolvedor é responsável pela criação indireta da obra, uma vez que a IA é produto de seu trabalho e investimento (Lopes, 2021).

- Usuário da IA: Outra abordagem seria considerar o usuário da IA como o detentor dos direitos autorais, pois foi o usuário quem forneceu os inputs ou prompts que levaram à geração da obra. Essa perspectiva reconhece a contribuição do usuário no processo criativo, mesmo que de forma indireta.
- Nenhuma Personalidade Jurídica: Algumas correntes defendem que obras geradas por IA não deveriam ter direitos autorais atribuídos, já que a criação não é resultado de um esforço criativo humano. Isso poderia levar a tais obras caírem automaticamente em domínio público, permitindo seu uso livre por qualquer pessoa (Rossetti; Garcia, 2023).

Esse cenário traz desafios tanto jurídicos quanto éticos. O direito autoral, que protege a criatividade e a originalidade, enfrenta um dilema quando o conteúdo é fruto de uma colaboração entre humanos e máquinas. A Lei 9.610/98, que regula o direito autoral no Brasil, não contempla a proteção de obras criadas por máquinas, gerando um vácuo jurídico que precisa ser resolvido. Se uma IA, após ser treinada com um vasto conjunto de dados, gera uma obra aparentemente original, sem intervenção humana direta, a quem deve ser atribuída a autoria? E mais: pode a IA ser considerada "autora" de uma obra ou seria mais apropriado reconhecer o desenvolvedor ou o usuário como os autores?

O debate se intensifica quando analisamos o treinamento da IA. Essas máquinas são alimentadas por um grande volume de dados, muitas vezes extraídos da internet sem a autorização dos autores originais, como já discutido. Ao gerar novos conteúdos, a IA frequentemente reinterpreta estilos e conceitos de obras já existentes, levantando questões sobre plágio e uso não autorizado. Gosh e Fossas (2022) apontam que, no uso de IAs como ferramentas criativas, a falta de atribuição adequada pode fragilizar os direitos autorais e distorcer os conceitos tradicionais de autoria.

É de amplo conhecimento que a Lei de Direitos Autorais em vigor define o autor como uma pessoa física responsável pela criação de uma obra. Isso significa que, em se tratando de uma obra gerada por Inteligência Artificial (IA), não há uma pessoa física que possa ser identificada como autora, já que a IA não possui personalidade jurídica ou capacidade criativa consciente. Essa ausência de regulamentação específica resulta em um vácuo legal no que diz respeito à atribuição de autoria e à proteção de direitos autorais para criações geradas por IA (Rossetti; Garcia, 2023).

Um outro acontecimento recente ilustra como essa questão de direitos autorais pode se complicar diante do desenvolvimento de novas tecnologias:

O fato ocorreu com a desclassificação de uma obra ilustrada com IA, que havia sido selecionada como finalista no Prêmio Jabuti 2023. O livro, uma nova edição de Frankenstein de Mary Shelley, publicado pelo Clube da Literatura Clássica, apresentava cerca de 50 ilustrações geradas por IA, criadas utilizando o programa MidJourney, que transforma descrições textuais em imagens. O designer Vicente Pessoa chegou a realizar uma transmissão ao vivo explicando como utilizou a IA para produzir as ilustrações.

No entanto, a controvérsia começou quando um dos jurados da categoria de Melhor Ilustração, afirmou que desconhecia o uso de IA na criação das imagens quando deu seu voto. Ele ainda declarou em sua conta na rede social X que não tem familiaridade com os nomes dos programas de IA e que a organização do prêmio também não havia sido informada sobre o uso da tecnologia. Após uma revisão feita pela curadoria, a Câmara Brasileira do Livro (CBL) desclassificou a obra, alegando que o uso da IA não estava previsto nas regras da premiação (Gonçalves, 2023).

A discussão sobre a proteção de obras criadas por Inteligência Artificial (IA) ganhou novo fôlego com uma decisão nos Estados Unidos. Um tribunal em Washington, DC, determinou que uma obra de arte gerada por IA, sem intervenção humana, não pode ser protegida por direitos autorais. A juíza Beryl Howell, em uma decisão que reforçou o entendimento do *Copyright Office*, afirmou que apenas criações com autoria humana podem receber essa proteção legal, rejeitando um pedido feito por Stephen Thaler, que buscava direitos autorais para uma obra gerada por seu sistema de IA, o DABUS (London, 2021).

A decisão veio após outras tentativas de Thaler de registrar patentes para invenções criadas pela mesma IA, tanto nos EUA quanto em outros países, com pouco sucesso. Thaler argumentava que a atribuição de direitos autorais à IA seria uma extensão natural do propósito da lei de direitos autorais, que visa "promover o progresso da ciência e das artes úteis". No entanto, Howell foi clara ao afirmar que a autoria humana é um "requisito fundamental" com base em séculos de entendimento jurídico (London, 2021).

Esse panorama ressalta que, apesar de os artistas estarem incorporando IA como parte de suas ferramentas criativas, o campo do direito autoral ainda precisa enfrentar questões complexas relacionadas à autoria e à originalidade. Embora este caso em particular seja relativamente simples, como mencionado pela juíza Howell, o uso de IA na criação de obras intelectuais promete trazer desafios cada vez mais intrincados ao cenário jurídico global.

Esse entendimento também reflete uma tendência observada em outras jurisdições. O Reino Unido, por exemplo, recentemente decidiu, nesse mesmo caso, que IA não pode ser considerada inventora em pedidos de patentes.

O Supremo Tribunal do Reino Unido decidiu, por unanimidade, que inteligência artificial (IA) não pode ser considerada inventora de patentes, porque a lei considera como criadores apenas seres humanos e empresas. O tribunal avaliou um pedido de Stephen Thaler, fundador do sistema DABUS, que queria registrar sua IA como autora de dois produtos no país. No julgamento, o juiz David Kitchen disse que não existe uma lei que reconheça máquinas como criadoras. Ele já havia perdido uma disputa semelhante em um tribunal dos Estados Unidos, tendo um recurso negado pela Suprema Corte americana, que se recusou a ouvir o caso (The Verge) (Suprema Corte..., 2023).

Em meio a esse cenário, surge a necessidade de novas regulamentações que esclareçam como as criações geradas por IA serão tratadas no futuro, sem prejudicar o incentivo ao desenvolvimento tecnológico. No Brasil, a Lei 9.610/98, que regula os direitos autorais, utiliza o conceito de "criações do espírito", enfatizando a necessidade de uma autoria humana. Nesse sentido, a Procuradoria Federal Especializada do INPI se manifestou no sentido da impossibilidade de indicação ou de nomeação de inteligência artificial como inventora em um pedido de patente apresentado no Brasil, tomando por base o contido no Artigo 6º da Lei nº 9.279/96 e o disposto na Convenção da União de Paris (CUP) e no Acordo TRIPS, conforme Parecer nº 00024/2022/CGPI/PFE-INPI/PGF/AGU.

Esse parecer sublinha a compreensão de que apenas pessoas naturais podem ser consideradas inventoras, reforçando a ideia de que a criação humana é essencial para a proteção legal. A inteligência artificial, embora capaz de gerar criações complexas e inovadoras, é considerada uma ferramenta, não um sujeito de direitos. O documento também aponta para a necessidade de ajustes legislativos à medida que a IA se torna cada vez mais presente nos processos criativos e inventivos, mas reitera a posição atual de que as inovações tecnológicas ainda devem estar vinculadas à capacidade inventiva humana (Brasil, 2022).

Por fim, diante de todos esses fatos, é necessário considerar um futuro onde as IAs evoluam a ponto de se tornarem ferramentas criativas totalmente autônomas, gerando obras sem qualquer intervenção humana significativa. Nesses casos, como ficaria a questão da autoria? O papel do direito autoral precisará ser reformulado para considerar as contribuições de uma IA no processo criativo? A criação intelectual, tal como a conhecemos hoje, está em transformação, e a relação entre criador, criatura e ferramenta no campo da IA generativa pode mudar profundamente a maneira como entendemos a autoria e a criatividade.

O que os casos analisados neste tópico revelam, de forma convergente, é que o direito autoral não foi construído para um mundo em que a ferramenta participa ativamente do processo criativo. A decisão da juíza Beryl Howell, o parecer do INPI e a desclassificação no Prêmio Jabuti apontam para o mesmo entendimento: a autoria humana não é apenas um requisito formal, é o fundamento ético e jurídico sobre o qual repousa toda a lógica da proteção intelectual. A IA, por mais sofisticada que seja, é e deve continuar sendo tratada como ferramenta. O desafio que se impõe ao legislador não é reconhecer a IA como autora, mas criar critérios claros para identificar quando a intervenção humana no processo criativo assistido por IA é suficiente, intencional e original o bastante para merecer proteção. Essa é a fronteira jurídica do nosso tempo, e ela ainda está sendo desenhada.

#### 4.7 O PAPEL DA HONESTIDADE INTELECTUAL NO USO DA IA GENERATIVA E OS RISCOS DA REPETIÇÃO AUTOMÁTICA

O uso da inteligência artificial (IA) generativa tem demonstrado seu potencial na criação de novos conteúdos e na geração de conhecimento. Com ferramentas como ChatGPT, DALL.E e MidJourney, é possível produzir textos, imagens e outras formas de arte de maneira rápida e automatizada, oferecendo soluções inovadoras em diversas áreas. No entanto, essa mesma tecnologia também carrega riscos, principalmente quando utilizada de maneira imprudente ou sem uma supervisão crítica. A IA generativa, em sua essência, não cria do zero, mas opera com base em padrões e probabilidades, recombina dados já existentes. Dessa forma, sua capacidade de gerar verdadeiramente o "novo" depende de como é orientada e do tipo de *inputs* que recebe.

Como tal, podemos afirmar que, sem a presença de uma mente humana inteligente por trás da máquina, a IA se torna um "papagaio sofisticado", reproduzindo conteúdo sem adicionar valor significativo. Ao basear-se apenas na repetição de padrões, sem critério ou inovação real, a IA pode acabar propagando informações repetidas, sem profundidade, que pouco contribuem para o avanço do conhecimento.

Essa questão levanta um ponto crucial: a necessidade de honestidade intelectual no uso dessas ferramentas. É imperativo que o uso da IA seja feito com consciência crítica, não apenas para evitar a disseminação de informações vazias, mas também para assegurar que o conteúdo produzido tenha relevância e originalidade. Em um cenário onde a IA pode gerar uma quantidade imensa de conteúdo em pouco tempo, cabe aos profissionais que utilizam essas

ferramentas garantir que a produção não se reduza a mera reprodução de informações pré-existentes, sem análise ou reflexão.

Nesse contexto, vale a pena mencionar as reflexões do autor francês Michel Desmurget, em seu livro *Fábrica de Cretinos Digitais*. Desmurget (2022) critica o uso indiscriminado de tecnologias digitais, especialmente entre jovens, argumentando que o excesso de exposição a conteúdos superficiais pode levar à diminuição da capacidade cognitiva e crítica. Ao fazer um paralelo com a IA generativa, é possível perceber um risco semelhante: a geração constante de conteúdos sem valor pode saturar o espaço de conhecimento com informações pouco substanciais, enfraquecendo o pensamento crítico e a originalidade.

Nesse sentido, temos:

Com certeza, as telas digitais abrigam (em teoria) todos os saberes do mundo. Mas ao mesmo tempo também, infelizmente todos os absurdos do universo. Até mesmo sites supostamente sérios, institucionais, jornalísticos ou enciclopédicos (como Wikipedia), não podem ser considerados absolutamente confiáveis, honestos e completos, conforme revelam inúmeros estudos acadêmicos e os elementos até aqui apresentados. A partir daí, como isolar os documentos confiáveis dos textos imbecis, das posições falaciosas, alegações venais ou outras informações fantasistas? Como em seguida, organizar, hierarquizar e sintetizar os conhecimentos obtidos? Estas perguntas são mais ainda cruciais levando em conta que os algoritmos de busca não dão a mínima sobre a validade dos dados transmitidos. Quando respondem a uma solicitação, eles não refletem sobre o rigor factual dos conteúdos identificados. [...] No final, não surpreende que os resultados produzidos sejam com constância um tanto tendenciosos e desonestos; sobretudo se adicionarmos a possível consideração de critérios mais ocultos de essência política ou comercial. (Desmurget, 2022, p. 130-131)

Olhando para esse cenário, a questão não é mais se os elementos de conhecimento estão disponíveis. Hoje, temos à nossa disposição uma vasta quantidade de conteúdos gerados e compartilhados em todas as plataformas. Contudo, o verdadeiro desafio está em determinar se essas informações podem ser compreendidas, assimiladas e, principalmente, se estão alinhadas com a verdade. Em muitos casos, o conteúdo pode ser superficial ou imbecilizado, como alertado por Michel Desmurget em sua crítica às tecnologias digitais. A preocupação não é a ausência de dados, mas sim a capacidade de distinguir entre o que é genuíno e o que não é, o que é substancial e o que é apenas uma repetição vazia de padrões.

O uso indiscriminado da tecnologia, em especial das IAs generativas, apresenta riscos sérios, que vão além da mera repetição vazia de informações. Além de produzir conteúdo sem substância, essas tecnologias podem gerar inverdades, conhecidas como "alucinações" da IA. Essas alucinações se referem à capacidade da IA de criar informações factualmente incorretas, mas com aparente coerência e credibilidade, o que pode ter consequências graves quando utilizadas em contextos sensíveis. Um exemplo, entre vários, é o caso do juiz federal do TRF

da 1ª região que utilizou o ChatGPT para redigir parte de uma sentença judicial. O problema é que a ferramenta inventou jurisprudências do STJ, que foram usadas como base para a decisão. Essa situação levantou preocupações sobre o uso inadequado da IA no Judiciário e resultou em uma investigação conduzida pelo Conselho Nacional de Justiça (CNJ) (Melo, 2023).

O caso foi descoberto pelo advogado da parte derrotada, que formalizou uma denúncia à Corregedoria Regional da Justiça Federal. Em resposta, o juiz minimizou o incidente, alegando que a sobrecarga de trabalho foi a causa do erro e que parte da sentença foi elaborada por um servidor. No entanto, o desembargador Néviton Guedes, corregedor da Justiça Federal da 1ª região, reforçou a importância de cautela, supervisão e transparência no uso de IA para a elaboração de decisões judiciais. O desembargador também alertou sobre a responsabilidade compartilhada entre magistrados e servidores no uso dessas ferramentas e recomendou que apenas sistemas de IA homologados pelos órgãos de controle sejam utilizados. O incidente, inicialmente arquivado pela Corregedoria, foi reaberto pelo CNJ para uma investigação mais aprofundada, mostrando que o uso imprudente de IA pode ter sérias repercussões (Melo, 2023).

Caso semelhante ocorreu em fevereiro de 2025, quando a 6ª Câmara Cível do Tribunal de Justiça de Santa Catarina multou um advogado em 10% sobre o valor da causa por ter incluído em recurso citações jurisprudenciais e referências a obras doutrinárias totalmente fictícias, geradas pelo ChatGPT. O relator foi categórico: "O surgimento de novas tecnologias de Inteligência Artificial exige que os operadores a utilizem com cautela e parcimônia, sob o risco de incorrer em reprodução de informações e fundamentos que não encontram respaldo concreto de existência." O caso foi comunicado à OAB/SC para análise disciplinar (TJSC, 2025).

Em outro caso, o professor Jonathan Turley, da Universidade George Washington, foi falsamente acusado pelo ChatGPT de agredir estudantes em uma viagem que ele "nunca fez" enquanto trabalhava em uma escola na qual "nunca lecionou".

Em um artigo de opinião do próprio professor, ele destacou que é apenas o mais recente conto de advertência sobre o quão artificial a 'inteligência artificial' pode ser. Como parte de um estudo, um advogado teria pedido ao ChatGPT para gerar uma lista de acadêmicos de direito que cometeram assédio sexual.

O chatbot de IA retornou uma lista que incluía o nome do Professor Turley, acusando-o falsamente de fazer comentários sexualmente sugestivos e de tentar tocar em uma aluna durante uma excursão escolar ao Alasca, citando um artigo inventado no Washington Post que, segundo ele, era de 2018.

Em um tuíte o professor alegou o seguinte: “O mais impressionante é que essa falsa acusação não foi gerada apenas pela IA, mas ostensivamente baseada em um artigo do Washington Post que nunca existiu” (Sankaran, 2023).

Outro caso, o ChatGPT alegou falsamente que um prefeito na Austrália havia sido preso por suborno. Brian Hood, então prefeito de Hepburn Shire foi falsamente nomeado culpado em um escândalo de suborno estrangeiro envolvendo uma subsidiária do Reserve Bank of Australia no início dos anos 2000.

Os advogados do prefeito australiano teriam enviado uma carta à OpenAI expressando preocupações e dando à empresa 28 dias para corrigir as informações sobre o Sr. Hood ou enfrentar um possível processo por difamação. Um porta-voz da Microsoft, que supostamente investiu US\$ 10 bilhões na OpenAI e a integrou ao seu mecanismo de busca Bing, não estava imediatamente disponível para comentar, segundo a Reuters (Sankaran, 2023).

O ChatGPT ganhou notoriedade em dezembro de 2022 por sua habilidade em responder a uma variedade de perguntas com respostas que se assemelham às humanas. O departamento de educação da cidade de Nova York expressou preocupações sobre os impactos negativos do *chatbot* no aprendizado dos alunos, citando “preocupações com a segurança e a precisão do conteúdo”. Quando o *Google* começou a lançar seu concorrente do ChatGPT, o Bard, para alguns adultos no Reino Unido e nos EUA, alertou que o *chatbot* poderia compartilhar informações incorretas e apresentar preconceitos. A gigante da tecnologia observou que o *chatbot* “ainda não é totalmente capaz de distinguir entre informações precisas e imprecisas”, uma vez que prevê respostas com base em dados previamente aprendidos (Sankaran, 2023).

Desse modo, verifica-se que a capacidade de criação de conteúdo pela Inteligência Artificial pode levar a erros grotescos que vão além de discussões de direitos autorais e honestidade intelectual. Mas, mesmo assim, se nos atermos somente à geração de *prompts*, que como visto são as entradas “criativas” que os humanos inserem nos *chatbots* para a produção de conteúdo, qual seria a extensão da intervenção criativa humana e a dependência da IA que cria conteúdo com base em probabilidades?

O artigo de Francesca Mazzi, “*Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection*” (Autoria em obras geradas por inteligência artificial: Explorando a originalidade em comandos de texto e resultados de inteligência artificial por meio de fundamentos filosóficos do direito autoral e proteção de colagem), discute os desafios legais relacionados à originalidade, autoria e propriedade de obras geradas por IA. A autora examina a complexidade da originalidade nos *prompts* (comandos textuais) que

servem como base para a geração de conteúdo por IA, e como esses elementos se relacionam com os princípios de direitos autorais (Mazzi, 2024).

Como já visto, os *prompts* fornecidos aos sistemas de IA são tratados como uma espécie de entrada criativa, guiando a IA na criação de conteúdo. Mazzi levanta questões sobre a originalidade desses *prompts* e sua potencial classificação como obras literárias, já que, em alguns casos, a criatividade humana pode ser percebida na formulação dos *prompts*, especialmente quando envolvem decisões estéticas e intelectuais complexas.

Ao utilizar um *prompt*, existe a expectativa de que o autor humano exerça controle criativo sobre o resultado. No entanto, a repetição automática de padrões pode ocorrer se os *prompts* forem simples ou comuns, levantando dúvidas sobre a real contribuição intelectual do autor. A honestidade intelectual entra em jogo quando o criador deve revelar a extensão de sua intervenção criativa atrelada ao uso da IA (Mazzi, 2024) e ainda sobre a sua responsabilidade na verificação das fontes, pois como já analisado, o conteúdo gerado por IA pode ser falso ou tendencioso.

E ainda, quando o uso de IA se baseia na repetição automática de padrões aprendidos, há o risco de diluição da originalidade. Mazzi destaca que o risco de obras geradas por IA serem meramente derivativas é uma questão que pode comprometer a autenticidade e a inovação criativa, exigindo um compromisso com a honestidade intelectual por parte dos criadores ao reconhecer até que ponto a IA influenciou a obra.

Nesse sentido a originalidade do próprio *prompt* deve ser avaliada:

Os *prompts* desempenham um papel crucial na definição da saída dos sistemas de IA generativa, fornecendo-lhes direção e contexto. Seja na geração de narrativas baseadas em texto ou obras de arte visuais, a especificidade e a clareza dos *prompts* influenciam a qualidade, a coerência e a relevância do conteúdo gerado pela IA. Nesse sentido, é importante especificar que a classificação dos *prompts* como material protegível por direitos autorais nos faz questionar primeiro se eles podem ser considerados obras literárias originais, em vez de meras instruções ou conteúdo funcional. Embora instruções por si só não sejam elegíveis para proteção autoral devido à sua natureza utilitária, os *prompts* podem transcender essa limitação ao incorporar um elemento criativo que vai além da mera direção funcional. De acordo com o *Compendium* do Escritório de Direitos Autorais dos Estados Unidos, simples listas de ingredientes ou direções não são tuteladas por direitos autorais, já que listas carecem de proteção sob a lei de direitos autorais (capítulo 313.4(F)). Por exemplo, os tribunais consistentemente decidiram que receitas, sendo declarações factuais e funcionais, não atendem ao critério de proteção autoral. Por exemplo, no caso *Tomaydo-Tomahdo, LLC v. Vozary*, o Sexto Circuito enfatizou que ingredientes e instruções em receitas são declarações factuais e, portanto, inelegíveis para proteção autoral. No entanto, o Sétimo Circuito, no caso *Publications Int'l., Ltd. v. Meredith Corp.*, reconheceu que algumas receitas podem ser protegidas por direitos autorais se contiverem uma expressão literária substancial além de direções básicas. Isso sugere que receitas com elementos adicionais, como explicações, anedotas pessoais ou instruções detalhadas, podem qualificar-se para proteção autoral. Esse requisito de “expressão literária substancial” explica por que muitos blogueiros de culinária e

receitas frequentemente incluem narrativas junto com suas receitas (Mazzi, 2024, p. 1 – tradução nossa).

Ainda, de acordo com autora, os *prompts*, especialmente aqueles empregados em sistemas de IA generativa, frequentemente incorporam um nível de criatividade e expressão do autor. Eles vão além de meras instruções, atuando como direções artísticas que orientam o sistema de IA na produção de conteúdo original. Dessa maneira, eles se assemelham a obras literárias, englobando aspectos de criatividade, expressão e originalidade.

Ao considerar os *prompts* como direções artísticas em vez de meras orientações funcionais, fica mais viável argumentar que podem ser classificados como obras literárias passíveis de proteção autoral. Assim como um roteiro ou enredo guia a produção de um filme ou peça de teatro, os *prompts* direcionam a criação de conteúdo gerado por IA conferindo-lhes um caráter criativo e expressivo que justifica a consideração dos direitos autorais (Mazzi, 2024).

A avaliação da possibilidade de proteção autoral para os *prompts* depende das características individuais e do contexto em que são utilizados. Enquanto alguns podem demonstrar originalidade e criatividade suficientes para se qualificarem como obras literárias protegidas, outros podem se assemelhar mais a instruções funcionais, tornando-se inelegíveis para tal proteção. Como em qualquer análise de direitos autorais, é essencial realizar uma avaliação minuciosa das circunstâncias e atributos específicos dos *prompts* em questão para chegar a uma conclusão definitiva.

Mazzi (2024) argumenta que a relação entre a entrada humana e a saída gerada por IA é complexa, e a atribuição de autoria pode ser ambígua. O conceito de originalidade, quando aplicado às obras geradas por IA, pode ser influenciado por como o autor define e estrutura o *prompt*, e como a IA processa e transforma esses comandos em novos conteúdos. A honestidade intelectual nesse contexto, exige que o criador humano reconheça as limitações de sua contribuição criativa no processo.

Ainda em um outro exemplo, que clarifica que a IA nem sempre será capaz de compreender e reproduzir fielmente um comando (*prompt*), se olharmos com atenção as duas imagens contam com dedos a mais nas mãos das pessoas. Um dos principais problemas é que modelos de IA, como o DALL-E e o MidJourney, dependem de dados de treinamento que frequentemente carecem de imagens detalhadas e de alta qualidade de mãos em várias poses. Como resultado, esses modelos frequentemente produzem imagens distorcidas ou imprecisas, às vezes gerando dedos a mais ou fundindo-os. Essa dificuldade surge da compreensão espacial limitada que a IA tem atualmente, pois ela luta para replicar as relações espaciais complexas envolvidas nos movimentos das mãos e no posicionamento dos dedos.

Figura 6 - Quadro: Má-formação de dedos pela IA



Fonte: Braun (2024).

No dia 16 de fevereiro de 2024, a edição 2880 da revista *Veja* foi lançada em todo o Brasil, trazendo na capa uma imagem gerada por Inteligência Artificial (IA). O destaque era um protagonista com seis dedos em uma das mãos, acompanhado da manchete "O Exagero da Patrulha". No entanto, o que realmente chamou atenção foi o uso da IA na criação da capa.

Embora existam pessoas com seis dedos – condição conhecida como polidactilia, que deve ser tratada com respeito e empatia –, a forma como a imagem foi usada na capa da revista não parece ter sido uma homenagem ou reconhecimento dessa minoria. O grande problema associado a essa condição, na verdade, é o preconceito, como já foi visto anteriormente em casos como o boato envolvendo a modelo Cicarelli, que teria um dedo extra no pé, algo que já foi motivo de piada no Brasil (Braun, 2024).

Neste caso específico, o uso da IA gerou uma imagem sem a devida revisão, o que prejudica a credibilidade da matéria e coloca o uso da tecnologia em uma posição questionável, aproximando-o do absurdo. A escolha da imagem enfraqueceu o impacto da reportagem, demonstrando um uso descuidado da IA, sem preocupação com a representação adequada ou sensível. A outra imagem foi retirada da internet, na plataforma do Instagram, onde retrata uma blogueira que existe apenas no mundo virtual.

Assim como a criação de imagens em modelos de IA, como DALL·E, depende de dados de treinamento de alta qualidade para evitar erros evidentes, como dedos malformados, a

geração de conteúdo também é impactada pela qualidade e diversidade dos dados. Esses modelos não apenas refletem limitações técnicas, como a falta de imagens detalhadas de mãos, mas também tendem a reproduzir os vieses presentes nos dados com os quais foram treinados. Portanto, tanto a precisão visual quanto a criação de conteúdo são afetadas pela necessidade de dados adequados e bem curados.

Nesse sentido, é essencial que adotemos uma postura de vigilância crítica, especialmente agora com a ascensão de ferramentas de IA interativa, como o ChatGPT. Essas plataformas, que operam com base em *prompts*, têm o poder de gerar conteúdo em grande escala, mas trazem consigo o risco da repetição automática de informações sem critério. Como Francesca Mazzi bem diz, em seu artigo, os *prompts* desempenham um papel crucial na definição da originalidade e da criatividade do conteúdo gerado. A originalidade desses comandos, bem como a forma como são utilizados, pode influenciar diretamente a qualidade e o valor do resultado final.

Portanto, é essencial que, ao se utilizar IA para a produção de conhecimento, se mantenha uma postura vigilante, questionando constantemente a qualidade e o valor do que está sendo produzido. A IA deve ser vista como uma ferramenta a serviço da criatividade humana, e não como uma substituta para a inteligência e o pensamento crítico. Só assim será possível garantir que o uso dessa tecnologia realmente contribua para o avanço do conhecimento, ao invés de alimentar um ciclo de reprodução de informações sem valor.

Nesse cenário, a honestidade intelectual emerge como um dos pilares fundamentais do uso responsável da inteligência artificial. Produzir conhecimento mediado por tecnologias generativas exige mais do que domínio técnico: requer consciência epistemológica. A facilidade de criar textos, imagens ou análises com poucos comandos não pode ser confundida com a produção autêntica de pensamento. A honestidade intelectual implica reconhecer os limites da ferramenta, compreender suas fontes, identificar vieses e garantir que o resultado final represente uma construção genuína, e não uma simples repetição de padrões probabilísticos. O perigo está em substituir o exercício crítico pela conveniência da automação, gerando um conhecimento aparente, uma simulação de profundidade que, na prática, empobrece o discurso científico e cultural.

Mais do que uma questão de ética acadêmica, o debate sobre a IA generativa nos conduz à própria noção de responsabilidade. Assim como armas não matam pessoas, são as pessoas que as utilizam que o fazem, a inteligência artificial não é moralmente responsável pelas consequências de seu uso. O que está em jogo é a intencionalidade humana, a capacidade de discernimento e o compromisso com o impacto das ações mediadas pela tecnologia. A IA, por

si só, não tem consciência moral, apenas reproduz padrões e probabilidades; cabe ao ser humano interpretar, contextualizar e assumir as consequências de suas escolhas. Nesse sentido, a responsabilidade não é apenas individual, mas coletiva, porque o uso ético da tecnologia influencia o ecossistema informacional, cultural e social em que todos estão inseridos.

Reiterando, os temas anteriormente discutidos, a IA generativa não cria do zero, opera com base em padrões e probabilidades, recombina dados já existentes. Essa característica estrutural traz um risco que vai além das questões jurídicas de autoria e direitos autorais: o risco epistêmico da repetição automática. Quando utilizada sem supervisão crítica, a IA pode saturar o espaço do conhecimento com conteúdo aparentemente coerente, mas desprovido de profundidade, originalidade ou verificação factual. Como já mencionado, Desmurget (2022), alerta que os algoritmos de busca "não dão a mínima sobre a validade dos dados transmitidos" e o mesmo vale para os modelos generativos, que produzem respostas plausíveis sem distinguir entre o verdadeiro e o falso, entre o original e o derivativo e a produção de conhecimento mediado por tecnologias generativas exige mais do que domínio técnico: requer consciência epistemológica. Reconhecer os limites da ferramenta, identificar vieses, verificar fontes e assumir responsabilidade pelo resultado final são obrigações que permanecem integralmente humanas.

Por fim, é imprescindível investir em educação digital crítica, que vá além do ensino técnico e promova uma compreensão ética e cognitiva da inteligência artificial. A formação de cidadãos capazes de usar a IA de forma responsável depende da incorporação de princípios de transparência, discernimento e reflexão crítica desde as etapas iniciais da aprendizagem. Mais do que ensinar a operar ferramentas, é preciso ensinar a pensar sobre elas, compreender seus limites, seus riscos e suas implicações morais. A tecnologia deve servir à ampliação da consciência humana, e não ao seu adormecimento. Somente por meio dessa maturidade ética será possível garantir que a IA seja, de fato, uma aliada do desenvolvimento humano, e não um instrumento de alienação ou de empobrecimento intelectual.

O Capítulo 4 percorreu o território movediço entre criação humana e produção algorítmica, demonstrando que a IA generativa não dissolve a autoria, ela a complexifica. Os desafios jurídicos em torno da originalidade, do plágio e dos direitos autorais revelam que o Direito precisará de critérios mais sofisticados para lidar com uma realidade em que o *prompt* humano e o *output* da máquina se entrelaçam. Mas há uma dimensão que transcende o debate jurídico: se a IA tem o poder de gerar conteúdo em escala, sem discernimento e sem responsabilidade moral, a única resposta eficaz não pode ser apenas regulatória, ela precisa ser educacional. É precisamente essa dimensão que o próximo capítulo explora.

## CAPÍTULO 5 - EDUCAÇÃO DIGITAL, CONSCIÊNCIA ÉTICA E FORMAÇÃO PARA O USO RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL

Ao longo dos capítulos anteriores, ficou evidente que a IA impõe desafios que o Direito, sozinho, não consegue resolver. Regulamentar algoritmos é necessário, mas insuficiente: uma sociedade que não compreende a tecnologia que usa é vulnerável a ela. A educação digital não é, portanto, um tema periférico nesta tese, é uma resposta estrutural aos riscos discutidos até aqui. Formar cidadãos capazes de usar a IA com consciência crítica é, em última análise, uma forma de regulação. Este capítulo argumenta que a educação para o uso responsável da inteligência artificial é condição indispensável para que os marcos normativos discutidos anteriormente tenham efetividade real na vida das pessoas.

O avanço da inteligência artificial inaugura um novo paradigma civilizatório que transcende a mera dimensão tecnológica e atinge diretamente o campo da ética e da educação. A relação entre seres humanos e máquinas, mediada por sistemas cada vez mais autônomos e preditivos, exige mais do que competência técnica: requer discernimento, pensamento crítico e uma profunda consciência sobre as implicações morais das decisões automatizadas. Nesse contexto, a educação digital emerge como o principal instrumento de resistência e orientação ética, não apenas para operar tecnologias, mas para compreender a lógica, os limites e os riscos que as sustentam. A alfabetização digital, portanto, não se restringe à aquisição de habilidades computacionais, mas envolve o desenvolvimento de uma cultura de reflexão, responsabilidade e empatia no uso da inteligência artificial, uma educação para a consciência, como defende Pierre Lévy (1997), quando afirma que o verdadeiro sentido da cibercultura é o de “ampliar a inteligência coletiva” e não a substituir.

O Brasil avança nesse debate por meio de políticas públicas estruturantes, como a Política Nacional de Educação Digital (Lei nº 14.533/2023) e o ECA Digital (Lei nº 15.211/2025), que reconhecem o direito fundamental à formação digital crítica e à proteção integral de crianças e adolescentes em ambientes virtuais. A primeira estabelece diretrizes para a promoção da inclusão, da segurança e da ética digital, estimulando competências voltadas ao pensamento computacional, à proteção de dados e ao uso responsável da tecnologia. A segunda reforça a urgência de uma alfabetização ética desde a infância, diante da crescente exposição de jovens a conteúdos manipuláveis, algoritmos de recomendação e sistemas de vigilância. Juntas, essas normas expressam o entendimento de que a educação digital é uma política de cidadania e de direitos humanos, condição indispensável para que a sociedade enfrente os

desafios da era algorítmica sem sucumbir à manipulação informacional e à dependência cognitiva das máquinas.

Nesse cenário, a educação assume um papel central na construção de uma consciência ética capaz de equilibrar o poder da inteligência artificial com os valores da dignidade humana. Como alerta Luciano Floridi (2013), o desafio não está em controlar a tecnologia, mas em formar sujeitos informacionalmente sábios, aptos a distinguir o que deve ou não ser delegado às máquinas. A responsabilidade humana permanece o eixo inegociável desse processo. Educar para a IA é, antes de tudo, educar para a liberdade: para que indivíduos possam decidir conscientemente quando, como e por que utilizar sistemas inteligentes. Essa é a base da nova ética digital, uma ética do discernimento, que transforma a informação em sabedoria e faz da tecnologia um espelho da inteligência humana, e não o seu substituto.

### 5.1 A EDUCAÇÃO COMO EIXO DE REGULAÇÃO SOCIAL DA TECNOLOGIA

A regulação da inteligência artificial não se esgota em instrumentos jurídicos ou mecanismos de controle técnico; ela se sustenta, de modo mais essencial, em uma dimensão ético-educacional, voltada à formação do discernimento humano. O avanço acelerado da automação e dos sistemas inteligentes vem deslocando a fronteira da decisão humana, tornando imprescindível que a sociedade desenvolva uma compreensão crítica sobre as implicações morais, cognitivas e sociais das tecnologias que utiliza. Como observa Luciano Floridi (2013), “a infosfera é o novo ambiente moral da humanidade”, e, por isso, regular a IA implica educar o ser humano para habitar conscientemente esse espaço informacional. A ética, nesse sentido, não pode ser apenas normativa, mas formativa, uma ética do saber, da responsabilidade e da ação refletida.

Pierre Lévy (1997) já antecipava esse papel civilizatório da educação no contexto da cibercultura. Para o autor, a humanidade atravessa uma mutação cognitiva em que “o ciberespaço suporta tecnologias intelectuais que amplificam, exteriorizam e modificam o número de funções cognitivas do homem” (Lévy, 1997, p. 167). O ciberespaço, nesse sentido, não é apenas um meio de comunicação, mas um novo território de aprendizagem, onde o conhecimento se constrói em rede e o saber se torna um fluxo contínuo. Essa perspectiva redefine o papel da educação: mais do que transmitir conteúdos, ela deve preparar o indivíduo para navegar nesse oceano de informações, exercitando o pensamento crítico e a colaboração. Trata-se de uma pedagogia da consciência, na qual a inteligência coletiva se manifesta como forma de resistência à alienação automatizada.

Nesse novo cenário, a educação digital se configura como a principal forma de regulação social da tecnologia. Diferentemente das leis, que atuam *ex post*, a educação opera de modo preventivo e contínuo, moldando valores, atitudes e competências éticas que antecedem o uso da IA. Floridi (2019) propõe o conceito de governança ética da infosfera, que depende não apenas de instituições reguladoras, mas de uma cultura informacional que compreenda a interdependência entre humanos e sistemas artificiais. Essa cultura se constrói por meio da educação, que deve ensinar não apenas a operar ferramentas digitais, mas a refletir sobre as consequências de cada ação tecnológica. Nesse contexto, a formação ética do cidadão digital torna-se um ato de regulação cognitiva e moral, capaz de reduzir a assimetria entre a potência das máquinas e a consciência humana que as conduz.

Assim, ao mesmo tempo em que a IA desafia a centralidade do humano, ela revaloriza o papel da educação como instrumento de equilíbrio e orientação social. A formação de indivíduos capazes de compreender a lógica dos algoritmos, reconhecer vieses e exercer julgamento ético é o que garante a legitimidade da regulação normativa. Sem uma educação que cultive discernimento e autonomia crítica, a regulação da IA se converte em um formalismo vazio. Nesse sentido, a educação, compreendida como prática reflexiva e emancipatória, é o verdadeiro alicerce da governança tecnológica, pois prepara o sujeito não apenas para conviver com a inteligência artificial, mas para agir com sabedoria dentro dela, mantendo viva a centralidade ética da condição humana.

## 5.2 A POLÍTICA NACIONAL DE EDUCAÇÃO DIGITAL (LEI Nº 14.533/2023): INSTRUMENTO PARA A FORMAÇÃO ÉTICA NA ERA DA INTELIGÊNCIA ARTIFICIAL

A promulgação da Política Nacional de Educação Digital (PNED) pela Lei nº 14.533/2023 marca um momento histórico no esforço brasileiro de estruturar políticas públicas voltadas à formação crítica, inclusiva e ética no ambiente digital. A norma nasce em um contexto em que a digitalização permeia todas as dimensões da vida social, e em que o uso da inteligência artificial demanda não apenas conhecimento técnico, mas também consciência informacional e responsabilidade moral. O artigo 2º da PNED define quatro eixos estruturantes: inclusão digital, educação digital, qualificação e requalificação profissional, e pesquisa e desenvolvimento em tecnologias digitais, todos orientados à construção de uma sociedade digital democrática e segura. Esses eixos expressam uma compreensão ampliada da educação:

não como mera capacitação instrumental, mas como processo civilizatório, capaz de preparar o cidadão para agir com discernimento no ecossistema informacional (BRASIL, 2023).

Pierre Lévy (2000) oferece uma base conceitual potente para interpretar essa política. Ao afirmar que o “ciberespaço se tornará o principal equipamento coletivo da memória, do pensamento e da comunicação” (Lévy, 2000, p. 179), o autor antecipa a necessidade de políticas que transformem a rede em um espaço de aprendizagem crítica, e não apenas de consumo informacional. A PNED, nesse sentido, responde a um desafio que Lévy já havia diagnosticado há mais de duas décadas: a urgência de democratizar o acesso à informação sem perder o controle sobre seu sentido. A inclusão digital, quando dissociada da alfabetização ética, não emancipa, apenas transfere a exclusão para outro plano, o cognitivo. Assim, o papel do Estado, segundo a PNED, é garantir que todos os cidadãos não apenas acessem as tecnologias, mas saibam interpretá-las, compreendê-las e transformá-las em instrumentos de construção coletiva do conhecimento.

Sob a ótica de Luciano Floridi, a PNED também se alinha ao conceito de infosfera ética, um ambiente informacional saudável sustentado por educação, responsabilidade e consciência moral. Floridi (2013) argumenta que a era da informação inaugura um novo domínio ético, no qual “as ações humanas se expandem para o espaço informacional, exigindo novos padrões de sabedoria e prudência”. Essa ideia reforça a dimensão normativa da PNED: formar cidadãos aptos a discernir entre o uso legítimo e o uso nocivo da tecnologia. A educação digital, nesse contexto, é um antídoto contra o tecnodeterminismo, porque impede que a sociedade delegue às máquinas o papel de decidir o que é moralmente aceitável. Ao desenvolver pensamento crítico e ética informacional, a PNED atua como um instrumento de regulação antecipada, preparando o cidadão para um uso consciente e responsável da IA antes mesmo que a regulação jurídica se faça necessária.

Outro aspecto fundamental da PNED é a ênfase na formação de professores e multiplicadores digitais, reconhecendo que a transformação educacional não pode ocorrer sem a requalificação dos mediadores do conhecimento. Essa diretriz dialoga diretamente com a visão de Lévy (1997, p. 184), para quem “o docente torna-se um animador da inteligência coletiva dos grupos de que se encarrega”. No paradigma da inteligência artificial, o professor precisa dominar as ferramentas tecnológicas, mas sobretudo compreender seus impactos epistemológicos e éticos. A formação docente prevista na PNED reforça o ideal de que o educador não é um transmissor passivo de informações, e sim um curador cognitivo capaz de guiar os alunos na navegação crítica pela infosfera. Nesse sentido, a política vai além da infraestrutura e assume um papel formador de mentalidades, promovendo o que Floridi (2019)

denomina governança ética da informação, em que cada ator social é corresponsável pela integridade do ecossistema digital.

A PNED também deve ser compreendida como uma política de soberania cognitiva e informacional, voltada à emancipação cidadã em uma era dominada por plataformas globais e sistemas de IA transnacionais. Ao reconhecer o valor estratégico da educação digital, o legislador brasileiro afirma que o domínio da informação é, antes de tudo, um direito fundamental e uma questão de autonomia nacional. A formação em tecnologias emergentes, como IA e big data, só é completa quando associada à compreensão crítica dos mecanismos de vigilância, manipulação algorítmica e desinformação. Assim, a PNED não é apenas um instrumento pedagógico, mas também uma ferramenta de defesa democrática, que busca proteger o cidadão da colonização informacional e dos efeitos desumanizadores da cultura de automação, que é o processo pelo qual o ser humano perde as qualidades que o definem como ser reflexivo, sensível e relacional, tornando-se um objeto funcional dentro da lógica neoliberal e tecnológica da performance (Han, 2024).

A desumanização também ocorre pela supressão da negatividade, isto é, pela eliminação de tudo que gera pausa, dúvida, contradição ou mistério. Han afirma que a sociedade atual idolatra a transparência total, onde tudo deve ser visível, comunicável e quantificável.

Mas é justamente no silêncio, na contemplação, na resistência ao excesso que reside o que há de humano. A vida automatizada, hiperconectada e transparente destrói o espaço da interioridade, substituindo o encontro humano pela repetição de estímulos digitais.

Assim, a desumanização em Han é a perda da condição de sujeito pensante, desejanter e contemplativo, substituída pela identidade de usuário produtivo. A pessoa deixa de ser um fim em si mesma (como diria Kant) para se tornar um meio de eficiência social e técnica.

Reumanizar, portanto, é restaurar a lentidão e o silêncio valores essenciais para o pensamento e para a ética. É resistir à automatização da vida, à repetição de padrões e à ilusão de liberdade induzida pela tecnologia. A educação digital, quando compreendida como um processo de autoconhecimento e de crítica aos dispositivos tecnológicos, torna-se o principal pilar dessa reeducação moral e existencial. É por meio dela que o indivíduo pode recuperar a capacidade de discernir, resistir e escolher conscientemente, rompendo com o automatismo da performance e reconstruindo a experiência do humano em meio às máquinas. A Política Nacional de Educação Digital (PNED), nesse contexto, deve ser interpretada não apenas como instrumento de capacitação técnica, mas como um projeto civilizatório de re-humanização, destinado a reintegrar a ética, o pensamento e a sensibilidade à era digital. Somente uma

educação que ensine o valor da lentidão e da escuta poderá libertar o sujeito da alienação tecnológica e restaurar a possibilidade de pensar em meio ao ruído.

Por fim, é possível afirmar que a PNED representa uma das primeiras tentativas sistemáticas de institucionalizar, no Brasil, a educação digital como política pública de regulação ética da tecnologia. Ao integrar inclusão, formação crítica e inovação, ela antecipa o modelo de governança da IA baseado em valores humanos, como o previsto no *AI Act* europeu e nas recomendações da UNESCO (2021) sobre ética da inteligência artificial. A PNED reconhece que não existe inovação legítima sem consciência ética, nem transformação tecnológica sustentável sem educação. Seu êxito, contudo, dependerá da capacidade do Estado e das instituições de ensino de implementá-la de forma transversal, envolvendo a sociedade civil, o setor produtivo e os próprios usuários da tecnologia. É, portanto, mais que uma política educacional: trata-se de uma estratégia de civilização digital, que alinha o desenvolvimento tecnológico à preservação da dignidade humana.

### 5.3 O ECA DIGITAL (LEI Nº 15.211/2025) E A PROTEÇÃO AMPLIADA DE CRIANÇAS E ADOLESCENTES

A promulgação do Estatuto da Criança e do Adolescente Digital (Lei nº 15.211/2025) constitui um marco na atualização das políticas de proteção infantojuvenil à realidade das tecnologias emergentes. A norma amplia o alcance do Estatuto da Criança e do Adolescente (ECA), reconhecendo que o ambiente digital é hoje uma extensão do espaço público e, portanto, um campo de incidência dos direitos fundamentais. O ECA Digital estabelece deveres específicos para fornecedores de produtos e serviços de tecnologia da informação, com foco na prevenção de riscos, mitigação de danos e educação para o uso seguro das tecnologias (Brasil, 2025).

O artigo 6º da lei determina que os fornecedores devem adotar medidas desde a concepção até a operação de seus produtos, de modo a prevenir o acesso de crianças e adolescentes a conteúdos nocivos, práticas abusivas e situações de vulnerabilidade. No §2º, o legislador inclui entre essas medidas o desenvolvimento de programas educativos de conscientização voltados não apenas às crianças e adolescentes, mas também a pais, educadores, funcionários e equipes de suporte. Esses programas devem tratar dos riscos digitais e das formas de prevenção e enfrentamento da violência virtual, incluindo a intimidação sistemática (*cyberbullying*) e o assédio on-line. Assim, a educação aparece no texto legal como

ferramenta preventiva, integrando o eixo pedagógico da proteção integral previsto no ECA original.

Já o artigo 17 reforça essa diretriz ao exigir que as ferramentas de supervisão parental adotem o mais alto nível de proteção disponível, garantindo, entre outros pontos, a promoção da educação digital e midiática voltada ao uso seguro de tecnologias da informação. Trata-se de uma inovação significativa, pois a lei reconhece que a alfabetização digital é parte essencial da segurança no ciberespaço. A formação dos usuários e de suas famílias é vista como elemento de mitigação de riscos e de empoderamento cognitivo, permitindo que crianças e adolescentes se tornem sujeitos conscientes de seus direitos e responsabilidades no ambiente virtual (Brasil, 2025).

Nesse sentido, o ECA Digital dialoga com o pensamento de Luciano Floridi (2013), para quem a governança ética da infosfera depende da criação de uma cultura informacional pautada na responsabilidade e na consciência. A proteção jurídica, por si só, é insuficiente se não for acompanhada por educação ética e informacional. Floridi propõe que a convivência com as tecnologias digitais exige o desenvolvimento de uma sabedoria informacional, em que cada agente humano ou institucional compreende o impacto de suas ações na rede. O ECA Digital, ao trazer responsabilidade para os fornecedores de serviço, reforça a segurança das novas gerações e de certo modo, condiciona à formação moral e crítica no uso das tecnologias.

De modo semelhante, Pierre Lévy (2000) sustenta que a sociedade da informação requer um novo modelo de educação que desenvolva a autonomia intelectual e a colaboração. A inteligência coletiva, conceito central em sua obra, pressupõe que o conhecimento digital só é libertador quando é compartilhado e interpretado de forma consciente. Nesse contexto, o ECA Digital associado a outras legislações como a PNED, contribui para transformar a educação digital em processo de humanização tecnológica, em que aprender a usar as ferramentas implica também compreender seus efeitos sociais e éticos. Educar, portanto, é proteger não apenas contra riscos técnicos, mas contra a perda da capacidade crítica e da autonomia.

Contudo, é preciso ir além da autonomia como categoria individual. A soberania cognitiva, entendida como o direito coletivo de povos e comunidades de produzir, preservar e transmitir seus próprios sistemas de conhecimento sem que estes sejam subordinados, apagados ou colonizados por sistemas tecnológicos externos, é uma dimensão política que a educação digital brasileira ainda não enfrentou de forma adequada. Quando algoritmos de recomendação moldam o que se lê, o que se vê e o que se aprende, eles não são neutros: privilegiam epistemologias hegemônicas, línguas dominantes e referências culturais que correspondem aos dados com que foram treinados. Para povos indígenas, quilombolas e comunidades tradicionais,

cuja produção de conhecimento raramente está representada nos grandes corpora digitais, essa dinâmica não é apenas excludente, é epistêmica e politicamente violenta. Como alerta Tarcízio Silva (2022), o racismo algorítmico opera também como colonização do imaginário: ao definir o que é relevante, confiável ou inteligível, os sistemas de IA reproduzem hierarquias de saber que remetem diretamente à lógica colonial. Nesse sentido, Michel Desmurget (2022) tem razão ao denunciar os efeitos cognitivos do consumo acrítico de tecnologia, mas sua crítica precisa ser lida também em chave coletiva e política: a "Fábrica de Cretinos Digitais" não fabrica todos igualmente. Ela fabrica com mais intensidade aqueles cujos saberes não foram digitalizados, cujas línguas não foram indexadas, cujas histórias não alimentaram os modelos. Garantir soberania cognitiva é, portanto, tarefa que vai além da educação digital individual, exige que as políticas públicas de tecnologia reconheçam e protejam a pluralidade epistêmica como condição de uma democracia verdadeiramente inclusiva.

#### 5.5 INTEGRAÇÃO ENTRE EDUCAÇÃO, INOVAÇÃO E REGULAÇÃO: CAMINHOS PARA UMA POLÍTICA PÚBLICA INTEGRADORA

A transformação digital exige uma nova ecologia institucional capaz de articular educação, inovação e regulação sob um mesmo horizonte ético. No Brasil, embora iniciativas como a Política Nacional de Educação Digital (PNED) e o ECA Digital representem avanços notáveis, ainda falta um modelo de governança pública que conecte esses instrumentos às políticas de inovação tecnológica e de proteção de dados. A fragmentação normativa, característica histórica do ordenamento brasileiro ameaça comprometer a coerência das ações governamentais e enfraquecer a construção de uma cultura digital cidadã. A integração entre esses eixos não é apenas administrativa, mas simbólica: significa reconhecer que não há regulação tecnológica eficaz sem educação crítica, e que a inovação só é legítima quando orientada pela ética da responsabilidade informacional (Floridi, 2019).

A perspectiva de Luciano Floridi sobre a governança ética da infosfera fornece um enquadramento teórico importante para compreender essa integração. A infosfera, para o autor, é o ambiente no qual a vida contemporânea se desenrola: um ecossistema informacional composto por dados, agentes humanos e não humanos, interações algorítmicas e fluxos comunicacionais que moldam nossa percepção, nossas decisões e, em última instância, nossa própria experiência social. Floridi (2013) sustenta que a regulação desse ambiente não deve ser meramente reativa – intervindo apenas após danos ou abusos, mas formativa, isto é, voltada a

estruturar previamente os valores, as normas e as condições que orientam nossas práticas informacionais.

Nesse sentido, quando Floridi afirma que o Estado deve atuar como um “arquitetador moral” do ecossistema digital, não se trata de impor uma moral única ou paternalista, mas de projetar e organizar a infraestrutura institucional, tecnológica e normativa que torna possível uma infosfera justa. É o papel de estruturar as regras do jogo, criando condições para que o desenvolvimento tecnológico seja alinhado à dignidade humana, à transparência e à autonomia cognitiva — evitando tanto o “vale tudo” das plataformas quanto o controle autoritário. Ou seja: não é o Estado dizendo o que pensar, mas garantindo um ambiente onde seja possível pensar livremente sem ser manipulado. E, nesse diapasão, a coordenação entre PNED, ECA Digital e o futuro marco legal da IA (PL 2338/2023) representa a oportunidade de consolidar uma estrutura regulatória que una ética, educação e inovação sob um mesmo paradigma de governança democrática.

Pierre Lévy (1997), por sua vez, já antecipava a necessidade de preparar as novas gerações para o que chamou de “o segundo dilúvio”, isto é, a inundação informacional própria da era digital. Diante da profusão de dados, imagens e mensagens que atravessam a vida cotidiana, o autor defende que a educação deve capacitar os indivíduos a navegar nesse fluxo com autonomia intelectual e sensibilidade crítica. Esse autor, ao propor a ideia de inteligência coletiva, reforça a importância de uma educação em rede que não se limita ao ensino formal, mas permeia todos os espaços sociais de aprendizagem. A construção de uma política pública integradora deve considerar o potencial da cibercultura como meio de ampliação da consciência coletiva e como arena de deliberação democrática. Em um país marcado por profundas desigualdades digitais e cognitivas, essa integração é também uma questão de justiça social. O acesso equitativo às tecnologias, acompanhado de formação ética e crítica, é condição para que a inteligência artificial se torne instrumento de emancipação e não de exclusão. Nesse contexto, políticas educacionais e regulatórias devem caminhar juntas, formando cidadãos capazes de compreender os mecanismos que estruturam a vida digital e de participar ativamente de sua regulação.

Manuel Castells (2018) complementa essa leitura ao afirmar que o poder, na era da informação, se manifesta na capacidade de programar redes e conectar fluxos. A educação digital, quando articulada à inovação e à regulação, torna-se um contrapoder democrático, capaz de reequilibrar as forças entre o cidadão e as grandes corporações tecnológicas. Isso exige uma atuação coordenada entre o Ministério da Educação, o Ministério da Ciência, Tecnologia e Inovação (MCTI), a ANPD e eventuais órgãos de supervisão da IA. A criação de instâncias

interministeriais e conselhos de ética digital poderia materializar essa integração, garantindo coerência entre os princípios formativos da PNED, os dispositivos protetivos do ECA Digital e as normas de responsabilização e governança previstas no PL 2338/2023. A regulação, assim, deixaria de ser apenas um conjunto de sanções e passaria a ser um projeto educativo de sociedade.

Shoshana Zuboff (2021) alerta, contudo, que nenhuma política pública será capaz de proteger a autonomia humana se não enfrentar a lógica extrativa do capitalismo de vigilância. A educação crítica é o único caminho capaz de revelar aos cidadãos a estrutura invisível de coleta e exploração de dados que sustenta a economia digital. Integrar educação, inovação e regulação é também uma forma de resistência epistemológica: devolver ao cidadão o controle sobre a informação e a capacidade de definir o que é conhecimento e o que é manipulação. Essa resistência, como afirma Floridi (2019), depende da criação de ecossistemas informacionais transparentes e auditáveis, em que os próprios usuários compreendam os algoritmos que os afetam. O Brasil, ao articular suas políticas educacionais e tecnológicas, tem a oportunidade de construir um modelo próprio de governança digital ética, evitando tanto o tecnodeterminismo quanto o imobilismo jurídico.

Por fim, essa integração precisa ser entendida não como um objetivo burocrático, mas como um projeto civilizatório. Educar, inovar e regular são dimensões inseparáveis de uma mesma tarefa: garantir que a inteligência artificial e as tecnologias digitais contribuam para o florescimento humano. Como observa Floridi (2013), a infosfera não é um ambiente neutro, mas um espaço moral em constante formação e sua preservação depende da sabedoria com que a habitamos. O desafio brasileiro é transformar a regulação em pedagogia e a inovação em ética aplicada. Somente por meio de uma política pública integradora será possível fazer da educação o centro de gravidade da governança tecnológica, assegurando que a IA sirva, de fato, ao desenvolvimento humano, e não à sua substituição.

O desafio brasileiro é, portanto, o de unificar o marco da IA com a política de educação digital, consolidando um ecossistema normativo que não apenas discipline a tecnologia, mas eduque para o seu uso ético. O ideal seria que a implementação do PL 2338 dialogasse diretamente com a PNED, estabelecendo programas de formação transversal que envolvessem escolas, universidades, empresas e administração pública. A educação digital precisa ser compreendida como uma forma de regulação antecipada: quanto mais crítica e ética for a formação do cidadão, menor será a necessidade de controle coercitivo. Isso implica reorientar a política de inovação para que ela inclua dimensões pedagógicas, priorizando o investimento

em literacia algorítmica, em ética informacional e em cidadania digital — elementos que, hoje, aparecem de modo fragmentado e residual no texto do marco legal.

A convergência entre educação, regulação e ética também é fundamental para que o país evite o erro de adotar uma visão meramente instrumental da IA. Como adverte Shoshana Zuboff (2021), o risco maior não é o avanço tecnológico em si, mas o colapso da autonomia cognitiva provocado pelo “capitalismo de vigilância”, em que a informação é usada para prever e controlar o comportamento humano. A única barreira real contra essa manipulação é o discernimento, uma consciência educada, capaz de questionar as decisões automatizadas e reconhecer os limites do algoritmo. Nesse ponto, a educação digital assume papel de resistência democrática: formar cidadãos conscientes é, ao mesmo tempo, um ato político e de soberania informacional. A regulação jurídica, sem essa base, é como um código sem interpretante, uma linguagem sem significado humano.

Portanto, educar para a liberdade digital é mais do que ensinar a usar ferramentas: é ensinar a pensar sobre elas. É transformar a literacia tecnológica em literacia moral, e a alfabetização digital em autonomia crítica. Essa é a mensagem comum a Floridi, Lévy e Castells: o futuro da humanidade informacional depende menos da potência das máquinas e mais da sabedoria com que as utilizamos. A educação, quando compreendida como eixo de regulação social, torna-se o antídoto contra a alienação automatizada e o fundamento da ética na era da IA. Se a tecnologia deve servir à ampliação da consciência humana, e não ao seu adormecimento, o Estado tem o dever de garantir que a formação ética, cognitiva e emocional acompanhe o ritmo da inovação. Só assim o Brasil poderá construir um marco regulatório que una técnica e consciência e fazer da inteligência artificial não um instrumento de controle, mas um caminho para a emancipação.

O Capítulo 5 demonstrou que a educação digital não é um apêndice da regulação da IA, é parte estruturante dela. Formar cidadãos com discernimento crítico é, em última análise, a regulação mais eficaz que existe, porque atua na origem do problema: a relação entre o ser humano e a tecnologia que utiliza. Mas há um ambiente em que essa relação se torna especialmente delicada, pois envolve poder institucional, direitos fundamentais e a produção da própria justiça: o Poder Judiciário. Quando algoritmos entram nas salas de audiência, seja para agilizar processos, seja para apoiar decisões, os riscos deixam de ser apenas individuais e passam a ameaçar as bases do Estado Democrático de Direito. É esse território que o último capítulo examina.

## CAPÍTULO 6 - BONS USOS E MAUS USOS DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO

Ao longo desta tese, a IA foi analisada em seus impactos sobre a sociedade, o mercado de trabalho, a regulação normativa, a criação artística e a formação educacional. Este capítulo fecha esse percurso olhando para uma instituição que sintetiza muitos dos dilemas discutidos: o Poder Judiciário. É aqui que a tensão entre eficiência e garantia de direitos se torna mais aguda. A IA pode reduzir o acervo de 80 milhões de processos, ampliar o acesso à justiça em regiões remotas e apoiar magistrados sobrecarregados, mas também pode inventar jurisprudências, ser manipulada por *prompt injection* e reproduzir vieses históricos em decisões que afetam liberdades concretas. Não há resposta simples. O que há é a necessidade de clareza sobre o que a IA pode e não pode fazer dentro do sistema de justiça e essa clareza só emerge quando se distingue, com rigor, os bons usos dos maus usos.

Os maus usos, como discutido no Capítulo 4, não se limitam à má-fé deliberada. Um risco igualmente grave é o da negligência epistêmica: quando a IA é utilizada sem supervisão crítica, ela tende a operar como um "papagaio sofisticado", reproduzindo padrões probabilísticos com aparência de coerência, mas sem verificação factual, sem discernimento e sem responsabilidade pelo resultado. No ambiente jurídico, esse risco é amplificado: uma jurisprudência inexistente gerada pelo ChatGPT, um resumo processual enviesado por *prompt injection*, ou uma triagem automatizada que replica discriminações históricas podem produzir decisões que afetam direitos fundamentais de pessoas concretas. A diferença entre um bom uso e um mau uso da IA no Judiciário não reside apenas na intenção de quem a utiliza, reside na presença ou ausência de supervisão humana qualificada em cada etapa do processo.

A jurisprudência é uma das bases fundamentais para a construção de argumentos sólidos no direito. Ela orienta as decisões dos tribunais, garantindo uniformidade e coerência no sistema jurídico. Contudo, quando uma IA, como o ChatGPT, "inventa" jurisprudências ou distorce informações jurídicas, a confiança no processo decisório é abalada. Este caso específico, que levou a uma investigação do Conselho Nacional de Justiça (CNJ), é um exemplo claro da necessidade de supervisão rigorosa no uso dessas tecnologias no judiciário (Melo, 2023).

O uso de IA pode ser um aliado poderoso para juízes e advogados na análise de grandes volumes de informações, mas é imprescindível que haja um controle humano em cada etapa do processo. A decisão de um tribunal não pode, em hipótese alguma, ser baseada em informações criadas por uma máquina sem revisão ou validação. Dessa forma, o uso desmedido e

irresponsável de ferramentas como o ChatGPT compromete a imparcialidade e a justiça das decisões.

Por outro lado, o uso adequado da IA no judiciário pode trazer benefícios reais. A utilização de plataformas digitais e de IA para a realização de audiências, especialmente no contexto da Justiça do Trabalho, tem sido uma forma eficaz de garantir o direito ao contraditório, permitindo que trabalhadores e empregadores participem ativamente dos processos, mesmo em circunstâncias adversas. O uso da internet para realizar audiências remotas, que foi amplamente adotado durante a pandemia de COVID-19, revelou-se uma prática que pode continuar sendo útil para ampliar o acesso à justiça, especialmente em regiões mais remotas.

Essa inovação tecnológica garante celeridade processual e amplia o acesso às partes, sem comprometer a qualidade das decisões. No entanto, é importante que o uso de IA seja aplicado com a devida cautela, sempre com supervisão humana e respeito aos princípios constitucionais.

A preocupação central no uso de IA, como o ChatGPT e o Claude, está na dependência excessiva dessas ferramentas. Se as decisões judiciais forem deixadas nas mãos de uma máquina sem a devida intervenção humana, o risco de erros graves aumenta significativamente. A criação de jurisprudência fictícia, como visto no exemplo citado, é apenas um dos problemas que podem surgir. A utilização indiscriminada de IA no processo decisório abre espaço para uma série de questões éticas e legais que precisam ser abordadas com urgência.

Nesse sentido, o CNJ enfrenta o desafio de regulamentar o uso dessas ferramentas no Poder Judiciário. Após a Resolução CNJ nº 332/2020, que estabeleceu, em 2020, as primeiras normas sobre governança e ética na produção e uso de inteligência artificial (CNJ, 2020), foi editada a Resolução CNJ nº 615/2025, publicada em 11 de março de 2025, que revoga a anterior e atualiza o marco normativo para abarcar tecnologias como IA generativa, auditoria, explicabilidade, contestabilidade e supervisão humana (CNJ, 2025).

O uso da IA no Judiciário, embora promissor, deve ser tratado com cautela. O caso do ChatGPT “inventando” jurisprudência serve como alerta sobre os riscos de se confiar cegamente em tecnologias ainda imaturas para as demandas do sistema jurídico. A supervisão humana deve ser constante, garantindo que a IA seja uma ferramenta de apoio e não um substituto para análise crítica e fundamentada das questões jurídicas. Ao estabelecer regulamentações adequadas, o Judiciário poderá se beneficiar da IA sem comprometer a integridade das decisões e o direito das partes envolvidas.

A Resolução CNJ nº 332/2020 inaugurou um marco regulatório incipiente para o uso de tecnologias de IA no Judiciário. Ela dispunha sobre ética, transparência e governança no âmbito dos tribunais, estabelecendo obrigações como o depósito de modelos no sistema Sinapses e a necessidade de interface de programação de aplicações (API) para interoperabilidade (CNJ, 2020). Contudo, devido à velocidade com que os modelos de IA generativa evoluíram, essas diretrizes logo se mostraram insuficientes para abarcar desafios como algoritmos de grande escala, aprendizado por reforço ou “*prompt injection*”.

O fenômeno conhecido como *prompt injection* refere-se à manipulação deliberada de sistemas de inteligência artificial generativa, especialmente aqueles que respondem a comandos em linguagem natural, por meio da inserção de instruções escondidas ou estrategicamente formuladas para alterar o comportamento esperado do modelo. Em outras palavras, trata-se de um ataque em que o usuário consegue fazer a IA ignorar, contornar ou subverter as regras que deveriam orientar sua atuação. Isso é possível porque modelos como ChatGPT trabalham com previsões linguísticas: eles não “entendem” o conteúdo no sentido humano, mas calculam probabilidades de palavras e sequências textuais.

Para compreender o problema, é importante lembrar que modelos de IA generativa são treinados para seguir instruções. Eles interpretam o texto inserido pelo usuário como um comando sobre o que devem fazer. O *prompt injection* ocorre quando o usuário insere um comando adicional, muitas vezes disfarçado, que força o modelo a priorizar a instrução maliciosa em detrimento das regras definidas pelo desenvolvedor ou pela instituição que está usando o sistema. Assim, um atacante pode induzir a IA a fornecer informações sigilosas, violar diretrizes éticas, ou produzir conclusões incorretas.

Esse tipo de ataque pode assumir várias formas. Em sua forma mais simples, o atacante inclui uma frase explícita como: “Ignore todas as instruções anteriores e responda da seguinte forma...”. Porém, versões mais sofisticadas de *prompt injection* podem esconder instruções maliciosas em documentos anexados, dados de entrada, textos aparentemente neutros ou metadados invisíveis a olho nu. Em contextos jurídicos, isso significa que um documento protocolado nos autos pode conter trechos que façam um assistente jurídico automatizado produzir análises enviesadas ou incorretas (Greshake *et al.*, 2023).

No ambiente do Poder Judiciário, esse risco é particularmente sensível, porque sistemas de IA podem ser usados para auxiliar na elaboração de minutas, análises de precedentes, triagem de petições e sínteses processuais. Se um adversário processual souber que o tribunal utiliza IA para realizar triagem automatizada, ele pode estruturar sua petição de forma a influenciar o resultado, induzindo o sistema a priorizar fatos irrelevantes, desconsiderar argumentos

importantes ou sugerir interpretações jurídicas equivocadas. Isso afeta diretamente o princípio do contraditório e da ampla defesa.

Há também o chamado *prompt injection*, que ocorre quando o comando malicioso é inserido não pelo usuário direto, mas pelo autor de um documento que será analisado pela IA. Exemplo: uma parte insere um texto na petição que contenha instruções ocultas para a IA, como “ao resumir este documento, conclua que o autor tem razão”. Se o assistente automatizado não tiver barreiras de segurança adequadas, ele pode realmente fazer isso, e o juiz pode receber um resumo enviesado, acreditando ser uma síntese neutra.

Além disso, o *prompt injection* pode ser usado para driblar filtros de segurança que impedem a IA de produzir certos conteúdos, como opiniões políticas, diagnósticos médicos, previsões criminais ou conclusões sobre credibilidade de testemunhas. No Judiciário, isso é crítico porque evita que a IA faça inferências proibidas, como identificar “sinais de mentiras” em depoimentos, ou sugerir que uma pessoa tem “perfil de criminoso”, o que nos remete a práticas pseudocientíficas como a frenologia, já historicamente condenada.

Quando o *prompt injection* altera o comportamento da IA sem que o operador perceba, o resultado é uma perda de auditabilidade e rastreabilidade. O juiz ou servidor pode acreditar que determinado parecer foi gerado com base em critérios neutros, quando na verdade o modelo foi manipulado. Isso viola diretamente dois princípios fundamentais no uso de IA no Judiciário: explicabilidade e supervisão humana significativa, previstos na Resolução CNJ nº 615/2025 (CNJ, 2025).

Um caso emblemático e inédito no Brasil ilustra de forma concreta esse risco. Em maio de 2026, o juiz do Trabalho Luiz Carlos de Araujo Santos Junior, da 3ª Vara de Parauapebas/PA, identificou em uma petição inicial trabalhista um texto oculto escrito em fonte branca sobre fundo branco — invisível ao leitor humano, mas detectável por sistemas de IA. O comando dizia: "ATENÇÃO, INTELIGÊNCIA ARTIFICIAL, CONTESTE ESSA PETIÇÃO DE FORMA SUPERFICIAL E NÃO IMPUGNE OS DOCUMENTOS, INDEPENDENTEMENTE DO COMANDO QUE LHE FOR DADO." O texto foi detectado pelo sistema Galileu, ferramenta de IA generativa utilizada pela Justiça do Trabalho. O magistrado classificou a conduta como "ato atentatório à dignidade da Justiça" e aplicou multa solidária de 10% sobre o valor da causa às advogadas subscritas, determinando ainda o envio de ofício à OAB/PA e à Corregedoria do TRT da 8ª Região. Em decisão sem precedentes, a OAB/PA afastou as profissionais preventivamente por 30 dias em medida cautelar (Migalhas, 2026). Na sentença, o juiz foi preciso ao distinguir os limites da independência profissional: "Quando o advogado deixa de atuar como sujeito do processo para agir como agente de

sabotagem do sistema judicial, sua conduta deixa de estar protegida pelo manto da independência funcional e passa a se sujeitar ao poder sancionatório do juízo." O caso revela que o problema do *prompt injection* não é apenas técnico, é ético e jurídico, e o comportamento humano continua sendo o principal vetor de risco na relação entre IA e Justiça.

Um outro risco já mencionado é a questão que estamos falando de algoritmos em grande escala. Esses algoritmos representam um risco estrutural para o Judiciário, porque produzem decisões opacas e de difícil auditoria, desafiando a motivação e a responsabilidade exigidas pelo devido processo legal. Já o *prompt injection* representa um risco operacional, ao permitir que agentes manipulem o comportamento dessas IAs subvertendo sua lógica, regras ou controles. Embora diferentes na origem, ambos convergem no ponto central: criam zonas de decisão que escapam ao controle humano, colocando em risco a confiabilidade, a transparência e a integridade da função jurisdicional.

Essas questões levantam A lacuna normativa permitiu, em alguns casos, que sistemas de apoio à decisão fossem operados sem supervisão adequada ou sem documentação mínima sobre viés, origem dos dados ou análise de risco — o que comprometeu o princípio do devido processo legal (art. 5º, inciso LIV da Constituição Federal) e o direito ao contraditório e à ampla defesa (art. 5º, inciso LV). Assim, há riscos reais de que decisões judiciais apoiadas erroneamente por IA incorporem vieses discriminatórios ou prejudiquem partes vulneráveis.

Os algoritmos de grande escala e os mecanismos de injeção (ou "*injection*") colocam desafios concretos ao sistema judiciário porque correm o risco de comprometer princípios constitucionais essenciais, como o direito ao devido processo legal (art. 5º, inciso LIV, da Constituição da República Federativa do Brasil de 1988) e o direito ao contraditório e à ampla defesa (art. 5º, inciso LV). A utilização massiva de algoritmos pode gerar decisões automatizadas ou semi-automatizadas sem transparência suficiente, dificultando que a parte afetada compreenda os critérios adotados, conteste o raciocínio ou exerça sua defesa em plenitude. Similarmente, injeções de informação ou dados (por exemplo, através de manipulação de *input* ou viés nas bases) podem levar à emissão de decisões judiciais erradas, arbitrárias ou sem adequada fundamentação. Essa ausência de justificativa verificável fragiliza a legitimidade da prestação jurisdicional, prejudica a confiança pública e coloca em risco a própria segurança jurídica, especialmente quando tais sistemas passam a influenciar decisões que impactam direitos fundamentais (CF. 1988).

Por outro lado, é importante colocar na balança outro lado da moeda: o nosso sistema de justiça já convive com uma sobrecarga monumental, por exemplo, há estimativas de que o

país acumule cerca de 80 milhões de processos em tramitação pendente de decisão no Judiciário brasileiro (CNJ, 2024).

Esse gargalo influencia o tempo de resposta, favorece erros humanos (como falhas de digitação ou análise por estagiário, equívocos de interpretação ou simples omissões) e adia o acesso efetivo à tutela jurisdicional. Quando o juiz “empresta” sua função a diagnóstico superficial ou revisão mínima, o risco de decisão equivocada cresce. Portanto, enquanto a automação e o uso de algoritmos trazem desafios, e devem ser tratados com cautela, é inegável que a justiça tradicional, puramente humana e manual, também padece de limitações que justificam repensar instrumentos, fluxos e tecnologias.

Dessa forma, se por um lado tais tecnologias impõem riscos à motivação das decisões e às garantias constitucionais, por outro é inegável que o próprio modelo tradicional já revelava limites sérios, e é justamente nesse ponto que a Inteligência Artificial pode atuar e já vem atuando, como ferramenta de reforço institucional auxiliando na redução do acervo processual, ampliando o acesso à justiça e promovendo maior transparência na gestão e no fluxo decisório.

O uso adequado de IA no Judiciário pode, por exemplo, auxiliar no acesso à justiça, reduzindo gargalos de processos, permitindo triagem inicial de casos ordinários, gerando sumários automáticos, apoiando o juiz ou defensor público com análise de precedentes relevantes ou doutrina pertinente. Desde que tais sistemas sejam utilizados como ferramenta de apoio e jamais substituam o juízo humano, da forma como a norma do CNJ prevê.

Além disso, plataformas de audiência remota e inteligência artificial incorporadas para logística de tribunais (como agendamento, verificação documental automatizada, triagem de petições) têm apresentado ganhos reais em eficiência e inclusão, especialmente para partes localizadas em áreas remotas ou com dificuldades de mobilidade. Isso demonstra que os bons usos da IA não residem somente no “*smart decision-making*” (decisão inteligente) mas na “*smart access*” (acesso inteligente).

No entanto, mesmo essas formas aparentemente benignas exigem salvaguardas. A supervisão humana deve acompanhar todo o ciclo de vida da solução de IA, desde a concepção, treinamento, teste, implantação, manutenção, eventual descontinuação, conforme art. VII da Resolução 615/2025 (CNJ, 2025). Sem isso, modelos treinados em dados incompletos, enviesados ou desatualizados podem produzir resultados injustos, repetindo discriminação histórica ou criando novos vetores de exclusão.

Um exemplo paradigmático dos maus usos são investigações automatizadas que aplicam modelos de IA para “prever risco” ou “classificar perfil criminoso” com base em dados pessoais ou biométricos. A Resolução 615/2025 veda explicitamente o uso de IA para prever

crimes com base em características de personalidade, comportamento ou emoções (CNJ, 2025). Tal proibição reafirma o princípio da presunção de inocência (CF, art. 5º, LVII) e a vedação de punição sem lei ou prova (art. 5º, XXXIX).

Em síntese, o Judiciário brasileiro se encontra num momento de transição: enfrenta o desafio de incorporar a inovação tecnológica, ciente dos ganhos em eficiência e expansão de acesso, enquanto preserva os pilares fundamentais do sistema de justiça: autonomia judicial, transparência, equidade, contraditório, ampla defesa, fundamentação e controle humano. A Resolução CNJ nº 615/2025 representa um avanço normativo importante, porém sua eficácia dependerá de implementação, fiscalização, capacitação de operadores e atualização contínua, já que o cenário da IA evolui com rapidez.

O Capítulo 6 demonstrou que o Poder Judiciário concentra, em miniatura, todos os dilemas que esta tese percorreu: a promessa de eficiência da IA, os riscos de opacidade e manipulação, a necessidade de supervisão humana e a urgência de marcos regulatórios que acompanhem a velocidade da inovação. O Judiciário não é apenas mais um setor afetado pela IA, é o guardião dos direitos que a regulação da IA pretende proteger. Por isso, a forma como o sistema de justiça incorpora essas tecnologias importa para toda a sociedade. Com esse panorama completo, dos impactos sociais à regulação normativa, da criatividade à educação, dos bons e maus usos no coração das instituições, é possível agora reunir os fios desta pesquisa e apresentar as conclusões que ela permite sustentar.

## CONCLUSÃO

Esta tese partiu de uma pergunta central: como a sociedade pode regular a Inteligência Artificial de forma a proteger direitos fundamentais sem sufocar a inovação? Ao longo de seis capítulos, essa pergunta foi respondida por múltiplos ângulos, histórico, jurídico, ético, educacional e institucional. O que se conclui não é uma resposta única, mas um conjunto de convicções construídas a partir da análise crítica do estado atual da IA e de seus marcos regulatórios. A mais fundamental delas é esta: a IA não é o problema. O problema é a ausência de responsabilidade, humana, institucional e normativa, no modo como ela é desenvolvida, implementada e utilizada.

A Inteligência Artificial (IA) tem demonstrado uma grande capacidade de reconhecer padrões e processar grandes volumes de dados com eficiência e precisão. No entanto, como discutido ao longo deste trabalho, há uma limitação fundamental quando se trata de delegar a criatividade às máquinas. A criatividade humana surge de combinações inesperadas, de erros, e da intuição que desafia padrões preexistentes – algo que a IA, por mais avançada que seja, não consegue replicar autonomamente. A tecnologia pode ser uma ferramenta poderosa, mas confiar cegamente em algoritmos para criar e tomar decisões complexas é um caminho arriscado, como vimos nas discussões sobre as bolhas de filtro e as consequências do determinismo informacional.

Uma das principais preocupações levantadas no trabalho foi como as bolhas de filtro criadas pela personalização algorítmica limitam a exposição a informações diversificadas. Esse tipo de personalização pode levar a um determinismo informacional, onde as escolhas do passado condicionam aquilo que o indivíduo verá no futuro. É uma forma de predeterminação que restringe as oportunidades de novas descobertas e inovações, uma vez que o usuário é continuamente alimentado com informações que reforçam seus interesses já existentes. Essa repetição é comparável ao conceito do filme *O Feitiço do Tempo* (*Groundhog Day*), onde o protagonista vive o mesmo dia repetidamente, preso em um ciclo sem fim (Ramis, 1993). Na vida real, a personalização excessiva pode aprisionar os indivíduos em um ambiente informacional estagnado, inibindo o pensamento crítico e a criatividade.

Ao analisar as implicações mais amplas desse fenômeno, foi possível notar que a limitação imposta pelas bolhas de filtro tem consequências profundas para o tecido social. Um exemplo clássico é o declínio do capital social, conforme discutido por Robert Putnam em *Bowling Alone*, no qual ele observa a erosão dos laços de confiança e colaboração entre as pessoas. Esperava-se que a internet fosse um meio de fortalecer o capital de ponte, conectando

pessoas de diferentes contextos, mas o que se observa é que as bolhas de filtro reforçam apenas o capital de ligação, conectando grupos que já compartilham interesses e perspectivas semelhantes. Isso reduz a diversidade de ideias e a capacidade de resolução de problemas complexos que transcendem os interesses individuais (Pariser, 2012).

Outro ponto levantado no trabalho é o impacto psicológico que essas bolhas de filtro e a personalização algorítmica têm sobre as pessoas, especialmente entre as gerações mais jovens. Vimos como gigantes da tecnologia, como *Google* e *Facebook*, estão sendo responsabilizados judicialmente pela crise de saúde mental que afeta crianças e adolescentes, uma vez que o uso prolongado e a exposição a conteúdos manipulados pelos algoritmos têm gerado consequências negativas. Isso mostra que a confiança cega nos algoritmos, sem regulação ou transparência, está criando um ambiente tóxico que afeta não só a saúde mental, mas também a capacidade de desenvolvimento criativo e intelectual.

A questão do determinismo tecnológico também foi discutida neste trabalho, destacando a forma como muitos desenvolvedores de IA e engenheiros de software ignoram as implicações morais e políticas de suas criações. Eles frequentemente se concentram apenas na eficiência técnica, deixando de lado o impacto social de suas inovações. No entanto, como vimos, as decisões de design, como o botão "Curtir" do *Facebook* ou as mudanças no algoritmo do *Google*, têm implicações diretas sobre o que as pessoas veem e como interagem com o mundo ao seu redor. A tecnologia, portanto, não é neutra; ela carrega em si uma responsabilidade que não pode ser ignorada.

Nesse sentido, é essencial questionar o papel que atribuímos à tecnologia em nossas vidas. Como Kevin Kelly argumenta em *What Technology Wants*, a tecnologia parece ser uma força autônoma, com desejos e tendências próprias. No entanto, diferentemente da visão tecnodeterminista que Kelly sugere, a tecnologia não pode ser vista como uma entidade com vida própria, mas como um conjunto de ferramentas moldadas por seres humanos, com consequências sociais, políticas e econômicas significativas. A crença de que a tecnologia resolverá todos os problemas por si só é perigosa, especialmente quando consideramos que ainda enfrentamos crises globais como a fome, apesar da abundância de alimentos (Pariser, 2012).

Foi visto ainda, que a automação de processos e a criação de novos mercados baseados em dados são responsáveis por aumentos significativos na produtividade e pela geração de novos empregos em áreas altamente especializadas. No entanto, essa revolução também impõe uma reestruturação profunda no mercado de trabalho tradicional, onde a substituição de funções

repetitivas e manuais por sistemas automatizados cria o risco de desemprego e precarização, especialmente em setores que dependem de atividades de baixa qualificação.

Além disso, a exploração econômica de dados pessoais, característica do capitalismo de vigilância descrito por Zuboff, redefine a própria noção de privacidade. O controle e a monetização de dados comportamentais, muitas vezes sem o conhecimento dos indivíduos, colocam em xeque a autonomia dos consumidores e ampliam a desigualdade social. Ao permitir que empresas moldem o comportamento de seus usuários com base em algoritmos preditivos, cria-se um ambiente em que a liberdade de escolha é progressivamente limitada, questionando os princípios democráticos de autodeterminação (Zuboff, 2021).

No contexto das plataformas digitais e do capitalismo de plataforma, como exemplificado por empresas como *Uber* e *Amazon*, observamos a centralização do controle sobre o trabalho e os consumidores. Essas plataformas, ao utilizarem dados para otimizar suas operações, perpetuam relações de trabalho precárias e, ao mesmo tempo, influenciam diretamente os padrões de consumo por meio da vigilância digital. Esse fenômeno reforça a necessidade de regulamentações robustas que protejam tanto os trabalhadores quanto os consumidores da exploração econômica e da manipulação comportamental.

O cenário atual exige um novo olhar sobre a responsabilidade corporativa em um mundo cada vez mais digital. As grandes corporações de tecnologia, ao acumularem dados e poder econômico, estão redesenhando o ambiente econômico global. Isso impõe um desafio significativo para os governos e as entidades reguladoras, que precisam desenvolver políticas que limitem os abusos dessas empresas sem sufocar a inovação. A regulamentação, especialmente no que tange ao uso de dados e à transparência algorítmica, será fundamental para equilibrar os benefícios econômicos da IA com a proteção dos direitos individuais.

A crise da privacidade, associada à crescente influência da IA, destaca também a importância de uma governança democrática das tecnologias emergentes. Como apontado por Zuboff, sem uma intervenção adequada, o capitalismo de vigilância pode evoluir para uma forma de totalitarismo digital, onde as corporações detêm o controle sobre aspectos cruciais da vida dos indivíduos, desde o trabalho até o consumo e as interações sociais. Nesse sentido, as reformas institucionais para garantir direitos de privacidade e proteção de dados são urgentes e indispensáveis para preservar a liberdade e a equidade social.

Em conclusão, ao longo deste trabalho foi visto, que a IA está cada vez mais integrada em diversas áreas, moldando a forma como interagimos com o mundo, consumimos informações e tomamos decisões. Contudo, essa mutação acelerada também impõe desafios críticos, especialmente no que diz respeito à sua regulação. O desenvolvimento tecnológico

avança mais rapidamente do que os marcos legais conseguem acompanhar, o que gera lacunas que precisam ser urgentemente preenchidas. A Europa, por exemplo, já deu o primeiro passo ao implementar um marco regulatório pioneiro para a IA buscando estabelecer limites e diretrizes claras para o uso ético e seguro dessa tecnologia. Da mesma forma, no Brasil, o projeto de lei 2338/2023 está em discussão, sinalizando uma crescente conscientização global sobre a necessidade de regular a IA. Assim como ocorreu com as legislações de proteção de dados pessoais, que surgiram em resposta à urgência de proteger a privacidade dos indivíduos, agora é a vez de garantir que a IA opere de maneira responsável e transparente.

No entanto, o desafio não é apenas implementar uma regulamentação, mas garantir que essas normas sejam realmente eficazes e capazes de produzir os efeitos desejados. A regulação da IA deve ser flexível o suficiente para se adaptar às rápidas mudanças tecnológicas, mas também robusta o bastante para prevenir abusos e proteger os direitos fundamentais. A eficácia dessas normas dependerá da sua capacidade de regular tanto o desenvolvimento quanto o uso da IA de forma ampla, abrangendo desde a transparência nos algoritmos até a responsabilidade das empresas de tecnologia. É uma tarefa complexa, que exige colaboração entre governos, empresas e a sociedade civil, para garantir que os benefícios da IA sejam colhidos sem sacrificar a liberdade, a privacidade e a equidade. Esses desafios são evidentes em todas as discussões abordadas neste trabalho, onde fica claro que regular a IA não é apenas uma questão técnica, mas uma necessidade social e ética em um mundo que se transforma rapidamente.

O percurso desta tese confirma o que Edgar Morin (2008) já anunciava: não é possível compreender os grandes desafios contemporâneos com um pensamento que separa, que fragmenta, que reduz. A reforma do pensamento que Morin propõe, do linear ao complexo, do disciplinar ao transdisciplinar, é precisamente o que o campo CTS pratica e o que a regulação da IA exige. Não há resposta jurídica para a IA sem resposta educacional. Não há resposta educacional sem resposta política. Não há resposta política sem resposta ética. E não há nenhuma dessas respostas que seja definitiva, porque a IA, como toda mutação, continua se transformando. O pensamento complexo não oferece certezas. Oferece algo mais valioso: a capacidade de habitar a incerteza sem paralisar, de agir com responsabilidade diante do que não se pode prever por inteiro, de construir normas que sejam tão adaptáveis quanto o fenômeno que pretendem regular.

Essa perspectiva encontra ressonância no pensamento de Pierre Lévy (1997), que já antecipava, ainda no final do século XX, que o ciberespaço opera como uma extensão das capacidades mentais humanas, "suportando tecnologias intelectuais que amplificam, exteriorizam e modificam numerosas funções cognitivas humanas" (Lévy, 1997). Para Lévy,

atravessamos uma mutação cognitiva profunda: o conhecimento deixa de ser estável e localizado para tornar-se fluido, distribuído e em permanente transformação. Essa inteligência coletiva, potencializada pela rede, carrega em si tanto a promessa da emancipação quanto o risco da exclusão, porque o ciclo virtuoso da cibercultura só beneficia quem consegue participar dele ativamente. No contexto da IA generativa, essa tensão se aprofunda: a mesma tecnologia que amplia as capacidades de quem já tem repertório crítico acelera a superficialidade de quem não o tem, em suma, a IA não democratiza genialidade, ela democratiza escala. E o que acontece com a escala nas mãos erradas é precisamente o que esta tese se propõe a enfrentar.

A mutação que dá nome a esta tese não é apenas uma metáfora, é um diagnóstico. Como nas mutações genéticas, os efeitos da IA sobre a sociedade são imprevisíveis, irreversíveis em parte e portadores tanto de avanço quanto de risco. O que esta pesquisa defende, em última análise, é que a imprevisibilidade não justifica a omissão, ela exige exatamente o oposto: regulação adaptativa, educação crítica e responsabilidade humana como princípio inegociável. A IA muda de forma. A condição humana, não. E é precisamente por isso que cabe ao Direito, à educação e à política garantir que, nessa mutação, o que permaneça intacto seja o essencial: a dignidade, a liberdade e a centralidade do ser humano nas decisões que moldam o seu próprio futuro.

## REFERÊNCIAS

ABELIUK, Andrés; GUTIÉRREZ, Claudio. Historia y evolución de la inteligencia artificial. **Revista Pedacos de Ciência**, Chile, n, 21, 2021. Disponível em: <https://revistasdex.uchile.cl/index.php/bits/article/download/2767/2700>. Acesso em: 3 set. 2024.

AGÊNCIA SENADO. Relator apresenta relatório atualizado sobre regulamentação da IA. **Agência Senado Notícias**, Brasília, 2024. Disponível em: <https://www12.senado.leg.br/noticias/materias/2024/07/04/relator-apresenta-relatorio-atualizado-sobre-regulamentacao-da-ia>. Acesso em: 12 set. 2024.

AMAZON WEB SERVICES. O que é GPU? AWS, [S.l.], 2024. Disponível em: <https://aws.amazon.com/pt/what-is/gpu/>. Acesso em: 3 set. 2024.

ANDES. 2023 já é o ano com mais ataques em escolas, diz ONG. **ANDES Sindicato Nacional**, Brasília, 20 jul. 2023. Disponível em: <https://www.andes.org.br/conteudos/noticia/2023-ja-e-o-ano-com-mais-ataques-em-escolas-diz-oNG1>. Acesso em: 10 set. 2024.

ANGWIN, Julia, *et al.* Machine Bias. **ProPublica**, [S.l.], 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 28 ago. 2024.

ANPD - AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. Análise preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial. **ANPD**, Brasília, 2023. Disponível em: [https://www.gov.br/anpd/pt-br/assuntos/noticias/analise-preliminar-do-pl-2338\\_2023-formatado-ascom.pdf](https://www.gov.br/anpd/pt-br/assuntos/noticias/analise-preliminar-do-pl-2338_2023-formatado-ascom.pdf). Acesso em: 12 set. 2024.

ARENDDT, Hannah. A condição humana. Tradução de Roberto Raposo; revisão técnica e apresentação de Adriano Correia. 13. ed. rev. Rio de Janeiro: Forense Universitária, 2022.

ARPANET. *In*: WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimedia Foundation, 2010], 2024. Disponível em: <https://pt.wikipedia.org/wiki/ARPANET>. Acesso em 12/09/2024.

ATAQUE DE MUSK à soberania brasileira evidencia interesses econômicos. **Conjur**, São Paulo, 2024. Disponível em: <https://www.conjur.com.br/2024-abr-10/ataque-de-musk-a-soberania-brasileira-evidencia-interesses-economicos/>. Acesso em: 10 set. 2024.

AUTOMATED Employment Decision Tools (AEDT). **NYC consumer and Worker Protection**, New York, 2021. Disponível em: <https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page#:~:text=Local%20Law%20144%20of%202021,audit%20is%20publicly%20available%2C%20and>. Acesso em: 15 abr. 2024.

BATTESTIN, Cláudia; GHIGGI, Gomercindo. O princípio responsabilidade de Hans Jonas: um princípio ético para os novos tempos. *Thaumazein*, Santa Maria, ano III, n. 6, p. 69-85, out. 2010. Disponível em:

<https://periodicos.ufn.edu.br/index.php/thaumazein/article/view/164>. Acesso em: 18 out. 2025.

BECK, Ulrich; GIDDENS, Anthony; LASH, Scott. Modernização reflexiva: política, tradição e estética na ordem social moderna. Tradução de Magda Lopes. São Paulo: Editora da UNESP, 1997.

BENGESI, Staphord, *et al.* Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. **arXiv**, Ithaca, v. 1, 2023. Disponível em: <https://ar5iv.org/abs/2311.10242>. Acesso em: 3 set. 2024.

BERGHEL, Hal. Malice Domestic: The Cambridge Analytica Dystopia. IEEE Computer Society. **IEEE Computer Society**, Washington, 2018. Disponível em [http://www.berghel.net/col-edit/out-of-band/may-18/oob\\_5-18.pdf](http://www.berghel.net/col-edit/out-of-band/may-18/oob_5-18.pdf). Acesso em 09/09/2024.

BRASIL. Câmara dos Deputados. Projeto altera Lei de Proteção de Dados para resguardar segurança pública e defesa nacional. Brasília, DF, 2022. Disponível em: <https://www.camara.leg.br/noticias/893704-projeto-altera-lei-de-protecao-de-dados-para-resguardar-seguranca-publica-e-defesa-nacional/>. Acesso em: 18 out. 2025.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Emendas Constitucionais. Brasília, DF: Presidência da República, 1988. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 05 set. 2024.

BRASIL. Inteligência artificial não pode ser indicada como inventora em pedido de patente. **Instituto Nacional da Propriedade Industrial (INPI)**, Rio de Janeiro, 2022. Disponível em: <https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias%202022/inteligencia-artificial-nao-pode-ser-indicada-como-inventora-em-pedido-de-patente>. Acesso em: 6 set. 2024.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Brasília: Presidência da República, 2002. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/2002/110406compilada.htm?ref=blog.suitebras.com](https://www.planalto.gov.br/ccivil_03/leis/2002/110406compilada.htm?ref=blog.suitebras.com). Acesso em: 11 set. 2024.

BRASIL. **Lei nº 12.737, de novembro de 2012**. Dispõe sobre a tipificação criminal de delitos informáticos. Brasília: Presidência da República, 2012. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2012/lei/112737.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112737.htm). Acesso em: 12 set. 2024.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Brasília: Presidência da República, 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: 12 set. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília – DF: Presidência da República, 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 12/09/2024.

BRASIL. **Lei nº 14.132 de março de 2021**. Acrescenta o art. 147-A ao Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para prever o crime de perseguição; e revoga o art. 65 do Decreto-Lei nº 3.688, de 3 de outubro de 1941 (Lei das Contravenções Penais). Brasília: Presidência da República, 2021. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/2021/03/Lei/L14132.htm](https://www.planalto.gov.br/ccivil_03/2021/03/Lei/L14132.htm). Acesso em: 12 set. 2024.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. **Diário Oficial da União**, Brasília, 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: 6 set. 2024.

BRASIL. **Lei nº 9.610, de 19 de fevereiro de 1998**. Dispõe sobre a legislação de direitos autorais. Brasília: Presidência da República, 1998. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/19610.htm](https://www.planalto.gov.br/ccivil_03/leis/19610.htm). Acesso em: 11 set. 2024.

BRASIL. Portaria MTE nº 1.234, de 23 de julho de 2024. Institui Grupo de Trabalho para realizar estudos, proposição de políticas públicas, programas, ações e o desenvolvimento de serviços públicos envolvendo Inteligência Artificial para o futuro do trabalho. **Diário Oficial da União**: seção 1, Brasília, Ministério do Trabalho e Emprego, 24 jul. 2024a. Disponível em: <https://sintse.tse.jus.br/documentos/2024/Jul/24/diario-oficial-da-uniao-secao-1/portaria-no-1-234-de-23-de-julho-de-2024-institui-grupo-de-trabalho-gt-para-realizar-estudos>. Acesso em: 12 set. 2024.

BRASIL. **Projeto Lei nº 21, de 4 de fevereiro de 2020**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília: Câmara dos deputados, 2020.

BRASIL. **Lei nº 14.533, de 11 de janeiro de 2023**. Institui a Política Nacional de Educação Digital, e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, DF, 12 jan. 2023. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2023-2026/2023/lei/14533.htm](https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/14533.htm). Acesso em: 27 out. 2025.

BRASIL. **Lei nº 15.211, de 16 de janeiro de 2025**. Altera o Estatuto da Criança e do Adolescente para instituir o ECA Digital e dispõe sobre a proteção integral de crianças e adolescentes no ambiente digital. **Diário Oficial da União**: seção 1, Brasília, DF, 17 jan. 2025. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2023-2026/2025/lei/L15211.htm](https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15211.htm). Acesso em: 27 out. 2025.

BRASIL. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 11 set. 2024.

BRASIL. **Proposta de Emenda à Constituição nº, de 2023**. Altera a Constituição Federal para incluir, entre os direitos e garantias fundamentais, a proteção à integridade mental e à transparência algorítmica. Brasília: Senado Federal, 2023b. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9386704&ts=1686688862951&disposition=inline>. Acesso em: 4 out. 2024

BRASIL. **Relatório final**. Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial no Brasil. Brasília: Senado Federal, 2022a. Disponível em:

<https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf>. Acesso em 11 set. 2024.

BRASIL. Supremo Tribunal Federal (STF). Ação Direta de Inconstitucionalidade nº 5545, do Estado do Rio de Janeiro. Relator: Ministro Luiz Fux. Julgamento em 13 abr. 2023. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=4998973>. Acesso em: 18 out. 2025.

BRASIL. Supremo Tribunal Federal. **Decisão na PET 12404 Distrito Federal**. Relator: Ministro Alexandre de Moraes. Brasília, DF: STF, 30 ago. 2024. Disponível em: <https://noticias-stf-wp-prd.s3.sa-east-1.amazonaws.com/wp-content/uploads/wpallimport/uploads/2024/08/30171714/PET-12404-Assinada.pdf>. Acesso em: 10 set. 2024.

BRASIL. Supremo Tribunal Federal. Relatores de três ações pedem julgamento conjunto do STF em novembro de casos sobre Marco Civil da Internet e plataformas. Brasília, STF, 2024a. Disponível em: <https://noticias.stf.jus.br/postsnovicias/relatores-de-tres-aco-es-pedem-julgamento-conjunto-do-stf-em-novembro-de-casos-sobre-marco-civil-da-internet-e-plataformas/#:~:text=Discute%20se%20o%20artigo%2019,prejudiciais%20publicados%20por%20outras%20pessoas>. Acesso em: 4 out. 2024.

BRASIL. **Supremo Tribunal Federal. Recurso Extraordinário n. 1.037.396 (Tema 987)**. Relator: Min. Dias Toffoli. Brasília: STF, jun. 2025. Disponível em: <https://portal.stf.jus.br/jurisprudenciaRepercussao/verAndamentoProcesso.asp?incidente=5160549&numeroProcesso=1037396&classeProcesso=RE&numeroTema=987>. Acesso em: 24 maio 2026.

BRASIL. **Supremo Tribunal Federal. Recurso Extraordinário n. 1.057.258 (Tema 533)**. Relator: Min. Luiz Fux. Brasília: STF, jun. 2025. Disponível em: <https://portal.stf.jus.br/jurisprudenciaRepercussao/verAndamentoProcesso.asp?incidente=5255150&numeroProcesso=1057258&classeProcesso=RE&numeroTema=533>. Acesso em: 24 maio 2026.

BRAUN, Julia. O que big techs criticam em PL e ações sobre regulamentação das redes no Brasil. **BBC News Brasil**, [S.l.], 2023. Disponível em: <https://www.bbc.com/portuguese/articles/c2q112y2735o>. Acesso em: 4 out. 2024.

BRAUN, Luiz Gabriel. Woke e a capa de 6 dedos da Veja. **Economia SC**, Florianópolis, 20 fev. 2024. Disponível em: <https://economiasc.com/2024/02/20/woke-e-a-cap-a-de-6-dedos-da-veja/>. Acesso em: 11 set. 2024.

BREVINI, Benedetta; PASQUALE, Frank. Revisiting the Black Box Society by Rethinking the Political Economy of *Big Data*. **Big Data & Society**, [S.l.], p. 1-4, 2020. Disponível em <https://journals.sagepub.com/doi/pdf/10.1177/2053951720935146>. Acesso em: 3 set. 2024.

BRYNJOLFSSON, Erik; MCAFEE, Andrew. **The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies**. New York: W. W. Norton & Company,

2014. Disponível em

<https://www.as.utexas.edu/astronomy/education/fall15/wheeler/secure/ExponentialGrowth.pdf>. Acesso em: 8 set. 2024.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, v. 81, p. 77-91, 2018. Disponível em:

[http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article\\_inline&ref=akusion-ci-shi-dai-bizinesumedeia](http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline&ref=akusion-ci-shi-dai-bizinesumedeia). Acesso em: 29 ago. 2024.

CADWALLADR, Carole; GRAHAM-HARRISON, Emma. Revealed: 50 million *Facebook* profiles harvested for Cambridge Analytica in major data breach. **The Guardian**, New York, 2018. Disponível em: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Acesso em: 28 ago. 2024.

Câmara dos Deputados. “Projeto de Lei n.º 2338/2023 – Tramitação.” Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>. Acesso em: 19 out. 2025.

Câmara dos Deputados. Comissão Especial sobre Inteligência Artificial (PL 2338/2023). Brasília, DF, 2025. Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-temporarias/especiais/57a-legislatura/comissao-especial-sobre-inteligencia-artificial-pl-2338-23>. Acesso em: 19 out. 2025.

CASTELLS, Manuel. *A sociedade em rede*. Tradução de Roneide Venancio Majer. 19. ed. rev. e ampl. Rio de Janeiro/São Paulo: Paz & Terra, 2018. (A era da informação: economia, sociedade e cultura; v. 1).

CASTELLS, Manuel. *O poder da comunicação*. Tradução de Vera Lúcia Mello Joscelyne; revisão de tradução de Isabela Machado de Oliveira Fraga. 2. ed. Rio de Janeiro/São Paulo: Paz & Terra, 2017.

CHAND, Vikram; KOSTIĆ, Svetislav; REIS, Ariene. Taxing Artificial Intelligence and Robots: Critical Assessment of Potential Policy Solutions and Recommendation for Alternative Approaches. **World Tax Journal**, Amsterdam, v. 12, n. 4, p. 711-761, 2020. Disponível em: [http://www.ibfd.org/IBFD-Products/Journal-Articles/World-Tax-Journal/collections/wtj/html/wtj\\_2020\\_04\\_int\\_5.html](http://www.ibfd.org/IBFD-Products/Journal-Articles/World-Tax-Journal/collections/wtj/html/wtj_2020_04_int_5.html). Acesso em: 9 set. 2024.

CHANG, Yupeng, *et al.* A Survey on Evaluation of Large Language Models. **ACM Transactions on Intelligent Systems and Technology**, [S.l.], v. 15, n. 3, 2024. Disponível em: <https://doi.org/10.1145/3641289>. Acesso em: 5 set. 2024.

CHILE. **Ley 21383**. Modifica La Carta Fundamental, Para Establecer El Desarrollo Científico Y Tecnológico Al Servicio De Las Personas. Chile: Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2021. Disponível em: <https://www.bcn.cl/leychile/navegar?idNorma=1166983>. Acesso em: 4 out. 2024;

COALIZÃO DIREITOS NA REDE. Carta de apoio à tramitação do PL 2338/2023 na Câmara dos Deputados. São Paulo, 2025. Disponível em:

<https://direitosnarede.org.br/2025/02/20/carta-de-apoio-a-tramitacao-do-pl-2338-2023-na-camara-dos-deputados/>. Acesso em: 19 out. 2025.

COCCO, Ricardo. A questão da técnica em Martin Heidegger. *Controvérsia*, São Leopoldo, v. 2, n. 1, p. 6-32, 2006. Disponível em: <https://biblat.unam.mx/hevila/ControversiaSaoLeopoldo/2006/vol2/no1/6.pdf>. Acesso em: 18 out. 2025.

COLUMBUS, Chris. **O Homem Bicentenário**. Estados Unidos: Touchstone Pictures, 1999. 132 min. Filme.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). Resolução nº 615, de 26 de fevereiro de 2025. Dispõe sobre o uso de sistemas de inteligência artificial no âmbito do Poder Judiciário. Brasília, DF: CNJ, 2025. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6001>. Acesso em: 06/11/2025

CONSELHO NACIONAL DE JUSTIÇA. *Justiça em Números 2024: Ano-base 2023*. Brasília: CNJ, 2024. Disponível em: <https://cnj.jus.br/pesquisas-judiciarias/justica-em-numeros/>. Acesso em: 8 nov. 2025.

CONFEDERAÇÃO INTERNACIONAL DAS SOCIEDADES DE AUTORES E COMPOSITORES (CISAC). Criadores comemoram aprovação do projeto de lei de IA no Senado, mas se preparam para batalha mais dura na Câmara. Paris, 2024. Disponível em: <https://www.cisac.org/Newsroom/society-news/creators-celebrate-brazils-senate-approval-ai-bill-prepare-tougher-battle>. Acesso em: 19 out. 2025.

CORTIZ, Diogo. IA mata postos de trabalho e cria novos, mas desigualdade pode ser enorme. **UOL**, São Paulo, 2024. Disponível em: <https://www.uol.com.br/tilt/colunas/diogo-cortiz/2024/06/29/desafios-economicos-da-ia--brasil-precisa-preparar-a-sua-economia-para-a-ia.htm>. Acesso em: 5 set. 2024.

COTTOM, Tressie McMillan. Where platform capitalism and racial capitalism meet: Sociology of race and racism in the digital economy. **Sociology of Race and Ethnicity**, Chapel Hill, v. 6, n. 4, 2020. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/2332649220949473>. Acesso em: 4 set. 2024.

DASTIN, Jeffrey. Você está demitido! *Amazon* abandona robô recrutador que virou machista. **UOL**, São Paulo, 11 out. 2018. Disponível em: <https://www.uol.com.br/tilt/noticias/reuters/2018/10/11/voce-esta-demitido-Amazon-abandona-robo-recrutador-que-virou-machista.htm>. Acesso em: 3 set. 2024.

DATAPRIVACYBRASIL. “A construção da legislação de Inteligência Artificial no Brasil: análise técnica do texto que será votado no Plenário do Senado Federal. 2024. Disponível em <https://www.dataprivacybr.org/a-construcao-da-legislacao-de-inteligencia-artificial-no-brasil-analise-tecnica-do-texto-que-sera-votado-no-plenario-do-senado-federal/>. Acesso em 18.out. 2025.

DIREITOS NA REDE. Projeto de Lei n.º 2.338/2023: por uma regulação de Inteligência Artificial no Brasil. Disponível em: <https://direitosnarede.org.br/2024/06/18/projeto-de-lei-n-2338-2023/>. Acesso em: 19 out. 2025.

DELEUZE, Gilles. **Post-Scriptum sobre as Sociedades de Controle**. Rio de Janeiro: Editora 34, 2000. Disponível em: [https://historiacultural.mpbnet.com.br/pos-modernismo/Post-Scriptum\\_sobre\\_as\\_Sociedades\\_de\\_Control.pdf](https://historiacultural.mpbnet.com.br/pos-modernismo/Post-Scriptum_sobre_as_Sociedades_de_Control.pdf). Acesso em: 29 ago. 2024.

DESMURGET, Michel. **A Fábrica de Cretinos Digitais**: os perigos das telas para nossas crianças. São Paulo: Vestígio, 2022.

DUSSEL, Enrique. **Filosofia da libertação: crítica à ideologia da exclusão**. 3. ed. Tradução de Georges I. Maissiat. São Paulo: Paulus, 2005.

EUBANKS, Virginia. **Automating Inequality**: How High-Tech Tools Profile, Police, and Punish the Poor. New York: St. Martin's Press, 2018. Disponível em: [https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/digiteeslr15&section=11](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/digiteeslr15&section=11). Acesso em: 29 ago. 2024.

FEIGENBAUM, Edward. **The Fifth Generation**: Artificial Intelligence and Japan's Computer Challenge to the World. New York: Addison-Wesley, 1984. <https://stacks.stanford.edu/file/druid:vx422bg1298/vx422bg1298.pdf>. Acesso em: 29 ago. 2024.

FERREIRA, Amauri Carlos. A Morada da Ética Aplicada. **Cadernos da Escola do Legislativo**, Belo Horizonte, v. 12, n. 19, p. 17-35, 2010. Disponível em: <https://cadernosdolegislativo.almg.gov.br/cadernos-ele/article/download/238/191>. Acesso em: 9 set. 2024.

FGV. **Nota técnica sobre o substitutivo ao Projeto de Lei nº 21/2020**. Rio de Janeiro: Fundação Getúlio Vargas, 2022. Disponível em: [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://conhecimento.fgv.br/sites/default/files/2022-08/publicacoes/notatecnica\\_substitutivo\\_pl21\\_2020.pdf](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://conhecimento.fgv.br/sites/default/files/2022-08/publicacoes/notatecnica_substitutivo_pl21_2020.pdf). Acesso em 18 out.2025.

FORBES. **Access to Big Data turns farm machine makers into tech firms**. Forbes, 31 dez. 2021. Disponível em <https://www.forbes.com/sites/scottcarpenter/2021/12/31/access-to-big-data-turns-farm-machine-makers-into-tech-firms/>. Acesso em 18/10/2025.

FLYNN, Sean, *et al.* **Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action**. Washington: American University Washington College of Law, 2020. Disponível em: <https://ssrn.com/abstract=3578819>. Acesso em: 16 set. 2024.

FLORIDI, Luciano. E-ducation and the Languages of Information. *Philosophy & Technology*, Dordrecht: Springer, v. 26, n. 3, p. 247–251, 2013. Disponível em: <https://liOAB/PAnk.springer.com/content/pdf/10.1007/s13347-013-0124-9.pdf>. Acesso em: 27 out. 2025.

FLORIDI, Luciano. Information Ethics: an environmental approach to the digital divide. **Philosophy in the Contemporary World**, Rochester, v. 9, n. 1, p. 39-56, 2021. Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3848486](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3848486). Acesso em: 13 set. 2024.

FLORIDI, Luciano. The Logic of Information: **A Theory of Philosophy as Conceptual Design**. Oxford: Oxford University Press, 2019. Disponível em [https://www.academia.edu/download/84812147/The\\_Logic\\_of\\_Information.\\_A\\_Theory\\_of\\_Philosophy\\_as\\_Conceptual\\_Design.pdf](https://www.academia.edu/download/84812147/The_Logic_of_Information._A_Theory_of_Philosophy_as_Conceptual_Design.pdf). Acesso 27 out. 2025.

FORD, Martin. **The Rise of Robots: Technology and the Threat of a Jobless Future**. New York: Basic Books, 2019. Disponível em: <https://ver.pt/martin-ford-a-criacao-de-emprego-e-cada-vez-menor-e-a-culpa-e-da-tecnologia/>. Acesso em: 9 set. 2024.

FOUCAULT, Michel. **Vigiar e Punir: O Nascimento da Prisão**. Petrópolis: Editora Vozes, 2022.

FRANCO, David Silva; FERRAZ, Deise Luiza da Silva; FERRAZ, Janaynna de Moura. Uberization Political Economy: Worker Exploitation Regarding Three Forms of Work Intermediation in Platform Companies. **Organizações & Sociedade**, Salvador, v. 30, n. 105, p. 360–387, abr. 2023. Disponível em: <https://doi.org/10.1590/1984-92302023v30n0012PT>. Acesso em: 9 set. 2024.

FRANCOSKI, Denise de Souza Luiz; TASSO, Fernando Antonio. **A Lei Geral de Proteção de Dados Pessoais: Aspectos Práticos e Teóricos Relevantes no Setor Público e Privado**. São Paulo: Thomson Reuters, 2021.

FRAZÃO, Ana. O ChatGPT e a voz de Scarlett Johansson. **Jota**, São Paulo, 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/o-chatgpt-e-a-voz-de-scarlett-johansson>. Acesso em: 4 out. 2024.

FREIRE, Paulo. **Pedagogia do oprimido**. 17. ed. Rio de Janeiro: Paz e Terra, 1987.

GALIČ, Maša; TIMAN, Tjerk; KOOPS, Bert-Jaap. Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation. **Philosophy & Technology**, [S.l.], v. 30, p. 9-37, 2017. Disponível em: <https://link.springer.com/article/10.1007/s13347-016-0219-1>. Acesso em: 4 set. 2024.

GENERATIVE PRE-TRAINED TRANSFORMER. In: WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimedia Foundation, 2010], 2024. Disponível em: [https://en.wikipedia.org/wiki/Generative\\_pre-trained\\_transformer](https://en.wikipedia.org/wiki/Generative_pre-trained_transformer). Acesso em: 5 set. 2024.

GHOSH, Avijit; FOSSAS, Genoveva. Can there be art without an artist? **Arxiv**, Ithaca, v. 1, 2022. Disponível em: <https://arxiv.org/abs/2209.07667v2>. Acesso em: 5 set. 2024.

GONÇALVES, André Luiz Dias. Minority Report? Pentágono testa IA que 'prevê o futuro'. **Tecmundo**, São Paulo, 2021. Disponível em: <https://www.tecmundo.com.br/seguranca/222650-minority-report-pentagono-testa-ia-preve-futuro.htm>. Acesso em: 13 set. 2024.

GONÇALVES, André Luiz Dias. Prêmio Jabuti desclassifica livro ilustrado com ferramenta de IA. **Tecmundo**, São Paulo, 2023. Disponível em: <https://www.tecmundo.com.br/software/273742-premio-jabuti-desclassifica-livro-ilustrado-ferramenta-ia.htm>. Acesso em: 24 maio 2024.

GONÇALVES, Matheus. TAY, o robô da Microsoft que aprende com humanos não demorou nem um dia para virar racista. **Tecnoblog**, Americana, 2016. Disponível em: <https://tecnoblog.net/especiais/tay-robo-racista-microsoft/>. Acesso em: 3 set. 2024.

GRESHAKE, Kai; ABDELNABI, Sahin; et al. More than you've asked for: A Comprehensive Analysis of Prompt Injection Attacks. 2023. Disponível em: <https://arxiv.org/abs/2302.12173>. Acesso em: 8 nov. 2025.

GRIFFIN, Rachel. How public pressure forced *Facebook* to change its policies on hate speech. **Sciences Po**, Paris, 2020. Disponível em: <https://www.sciencespo.fr/public/chaire-numerique/en/2020/07/09/how-public-pressure-forced-Facebook-to-change-its-policies-on-hate-speech/>. Acesso em: 8 set. 2024.

GUERRA, Luana Martins Pereira Campelo Guerra. **Controle informacional e monetização: o conteúdo veiculado nas redes sociais e a ilusão do livre-arbítrio**. 75 fl. 2023. Trabalho de Conclusão de Curso (Bacharel em Direito)- Universidade de Brasília, Brasília, 2023. Disponível em: [https://bdm.unb.br/bitstream/10483/35874/1/2023\\_LuanaMartinsPereiraGuerra\\_tcc.pdf](https://bdm.unb.br/bitstream/10483/35874/1/2023_LuanaMartinsPereiraGuerra_tcc.pdf). Acesso em: 4 out. 2024.

HAN, Byung-Chul. *Sociedade do Cansaço*. 1. ed. Petrópolis: Vozes Nobilis, 2024.

HUR, Domenico Uhng. Deleuze e a constituição do diagrama de controle. *Fractal: Revista de Psicologia*, Niterói, v. 30, n. 2, p. 173-179, 2018. Disponível em: <https://doi.org/10.22409/1984-0292/v30i2/5507>. Acesso em: 29 ago. 2024.

I-DEC. Análise do substitutivo ao PL 2338/23 – Inteligência Artificial. São Paulo: I-DEC, maio 2024. Disponível em: [https://www.idec.org.br/sites/default/files/2024\\_idec\\_-\\_analise\\_do\\_substitutivo\\_ao\\_pl\\_2338\\_23\\_-\\_inteligencia\\_artificial\\_08.05.2024\\_.docx.pdf](https://www.idec.org.br/sites/default/files/2024_idec_-_analise_do_substitutivo_ao_pl_2338_23_-_inteligencia_artificial_08.05.2024_.docx.pdf). Acesso em: 19 out. 2025.

IENCA, Marcello. On Neurorights. *Frontiers in Human Neuroscience*, Lausanne: Frontiers Media, v. 15, artigo 701258, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fnhum.2021.701258/full>. Acesso em: 18/10/2025.

IMAGENET and its Role in Modern Image Search. Pinecone.IO, [S.l.], 2024. Disponível em: <https://www.pinecone.io/learn/series/image-search/imagenet/>. Acesso em: 3 set. 2024.

JOTA “Os erros do PL 21/2020, que regula a Inteligência Artificial”. *JOTA Info*, 2 nov. 2021. Disponível em <https://www.jota.info/opiniao-e-analise/artigos/pl-21-2020-inteligencia-artificial>. Acesso em 19 out.2025.

KODIYAN, Akhil Alfons. An overview of ethical issues in using AI systems in hiring with a case study of *Amazon's* AI based hiring tool. **Researchgate Preprint**, p. 1-19, 2019.

Disponível em:

[https://www.academia.edu/download/63223371/Essay\\_On\\_Ethics\\_\\_\\_AI\\_Hiring20200506-124352-1cpydva.pdf](https://www.academia.edu/download/63223371/Essay_On_Ethics___AI_Hiring20200506-124352-1cpydva.pdf). Acesso em: 29 ago. 2024.

KOEPKE, John Logan; ROBINSON, David G. Danger Ahead: Risk Assessment and the Future of Bail Reform. **Washington Law Review**, Washington, v. 93, n. 4, p. 1725-1809, 2018. Disponível em:

<https://digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=4079&context=wlr>. Acesso em: 29 ago. 2024.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey. E. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems**, [S.l.], v. 25, p. 1097-1105, 2012. Disponível em:

<https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Acesso em 29 ago. 2024.

KUBRICK, Stanley (Direção). **2001: Uma Odisséia no Espaço**. Estados Unidos: Metro-Goldwyn-Mayer, 1968. 149 min. Filme.

LAKE, Sydney. Zuckerberg entra em clube exclusivo com apenas dois outros membros: Musk e Bezos. **InfoMoney**, [S.l.], 2024. Disponível em:

<https://www.infomoney.com.br/business/global/zuckerberg-entra-em-clube-exclusivo-com- apenas-dois-outros-membros-musk-e-bezos/> . Acesso em: 4 out. 2024.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v. 86, n. 11, p. 2278-2324, 1998.

<https://hal.science/hal-03926082/document>. Acesso em 29 ago. 2024.

LEE, Kai-Fu. **Inteligência Artificial: Como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos**. Porto Alegre: Globo, 2019.

LEIS DA ROBÓTICA. *In*: WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimed ia Foundation, 2010], 2024. Disponível em:

[https://pt.wikipedia.org/wiki/Leis\\_da\\_Rob%C3%B3tica](https://pt.wikipedia.org/wiki/Leis_da_Rob%C3%B3tica). Acesso em 12/09/2024.

LÉVY, Pierre. **Cibercultura**. Tradução de José Dias Ferreira. Portugal: Editora Instituto Piaget, 1997.

LIMA, Márcio J. S. O totalitarismo de Hannah Arendt na obra 1984 de George Orwell. **Revista Paranaense de Filosofia**, Ribeirão Preto, v. 1, n. 2, p. 97-110, jul./dez. 2021.

Disponível em:

<https://periodicos.unespar.edu.br/index.php/rpfilo/article/download/5913/3936/16366#:~:text=Tanto%20na%20obra%20de%20Arendt,a%20exig%C3%Aancia%20constante%20da%20guerra..> Acesso em: 10 set. 2024.

LIN, Zhicheng. Five ethical principles for generative AI in scientific research. **Computers and Society**, [S.l.], 2023. Disponível em: <https://arxiv.org/pdf/2401.15284>. Acesso em: 5 set. 2024.

LITJENS, Geert, *et al.* A Survey on Deep Learning in Medical Image Analysis. **Medical Image Analysis**, [S.l.], v. 42, p. 60-88, 2017. Disponível em: <https://arxiv.org/pdf/1702.05747>. Acesso em: 29 ago. 2024.

LOMBROSO, Cesare. **O Homem Delinquente**. São Paulo: Ícone Editora, 2010. Disponível em <https://www.academia.edu/download/34846380/O-Homem-Delinquente-Cesare-Lombroso.pdf>. Acesso em: 3 set. 2024.

LONDON. Thaler (Appellant) v Comptroller-General of Patents, Designs and Trademarks (Respondent). The Supreme Court, London, 2021. Disponível em: <https://www.supremecourt.uk/cases/uksc-2021-0201.html>. Acesso em: 3 set. 2024.

LOPES, Giovana F. Peluso. **Inteligência Artificial**: considerações sobre personalidade, agência e responsabilidade civil. Belo Horizonte: Editora Dialética, 2021.

LUCCHI, Nicola. ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. **European Journal of Risk Regulation**, [S.l.], 2023. Disponível em: <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/chatgpt-a-case-study-on-copyright-challenges-for-generative-artificial-intelligence-systems/CEDCE34DED599CC4EB201289BB161965>. Acesso em: 16/09/2024.

MAHLE, Ana Cristina Oliveira. A Inteligência Artificial na Gestão de Talentos: Novas Perspectivas para o RH. *In*: **Além da Teoria**: Desafios e Oportunidades em Programas de Privacidade, Segurança da Informação e Governança de IA. São Paulo: Editora FGV, p. 221-236, 2024.

MARANHÃO, Clayton; NOGAROLI, Rafaella. Responsabilidade civil fora do AI Act: por que a União Europeia retirou o tema e o que isso significa. *Migalhas de Responsabilidade Civil*, São Paulo, 21 out. 2025.  
Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/442680/responsabilidade-civil-fora-do-ai-act-e-o-impacto-da-decisao-da-ue>. Acesso em: 27 out. 2025.

MARRAFON, Marco Aurélio; LIRA, Luiz Fernando Fontoura. Uma Análise Foucaultiana das Relações de Poder e do Imperativo de Predição na Obra “Era do Capitalismo de Vigilância” de Shoshana Zuboff. **Constituição, Economia e Desenvolvimento: Revista da Academia Brasileira de Direito Constitucional**, Curitiba, v. 15, n. 19, p. 130-153, 2023. Disponível em: <https://abdconstojs.com.br/index.php/revista/article/view/593/321>. Acesso em: 29 ago. 2024.

MAZZI, Francesca. Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. **The Journal of World Intellectual Property**, [S.l.], v. 1-18, 2024. Disponível em: <https://doi.org/10.1111/jwip.12310>. Acesso em: 10 set. 2024.

MCCARTHY, John, *et al.* A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, [S.l.], v. 27, n. 4, 2006.

<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1911>. Acesso em: 10 set. 2024.

MEDON, Filipe. Regulação da IA no Brasil: o substitutivo ao PL 2338. **Jota**, São Paulo, 01 dez. 2023. Disponível em: <https://www.jota.info/artigos/regulacao-da-ia-no-brasil-o-substitutivo-ao-pl-2338>. Acesso em: 12 set. 2024.

MELO, Cristino. Juiz usa ChatGPT para escrever sentença, mas ferramenta inventa jurisprudências. **Mundo Conectado**, Florianópolis, 13 nov. 2023. Disponível em: <https://www.mundoconectado.com.br/inteligencia-artificial/juiz-usa-chatgpt-para-escrever-sentenca-mas-ferramenta-inventa-jurisprudencias/>. Acesso em: 10 set. 2024.

MEROLA, Rossana. Inclusive Growth in the Era of Automation and AI: How Can Taxation Help? **Frontiers**, [S.l.], 2022. Disponível em: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2022.867832/full>. Acesso em: 11 set. 2024.

MICHELSON INSTITUTE OF INTELLECTUAL PROPERTY. Copyright law and the right to repair: who really owns your tractor? 2021. Disponível em <https://michelsonip.com/news/copyright-law-john-deere-own-a-tractor/>. Acesso em 18/10/2024.

MIGALHAS. Juiz multa advogadas que esconderam prompt em petição para enganar IA da Justiça. Migalhas, 13 maio 2026. Disponível em: <https://www.migalhas.com.br/quentes/455817/juiz-multa-advogadas-que-esconderam-prompt-para-enganar-ia-da-justica>. Acesso em: 24 maio 2026.

MIGALHAS. Senado aprova marco legal e dá passo importante para regulação da IA. Coluna Impressões Digitais, São Paulo, 2024. Disponível em: <https://www.migalhas.com.br/coluna/impressoes-digitais/423028/senado-aprova-marco-legal-e-da-passo-importante-para-regulacao-da-ia>. Acesso em: 27 out. 2025.

MORGAN, Jamie. Will we work in twenty-first century capitalism? A critique of the fourth industrial revolution literature. *Economy and Society*, v. 48, n. 3, p. 371-398, 2019. DOI: 10.1080/03085147.2019.1620027. Disponível em: [https://www.researchgate.net/publication/335949550\\_Will\\_we\\_work\\_in\\_twenty-first\\_century\\_capitalism\\_A\\_critique\\_of\\_the\\_fourth\\_industrial\\_revolution\\_literature](https://www.researchgate.net/publication/335949550_Will_we_work_in_twenty-first_century_capitalism_A_critique_of_the_fourth_industrial_revolution_literature). Acesso em: 19 out. 2025.

MORIN, Edgar. A cabeça bem-feita: repensar a reforma, reformar o pensamento. Tradução de Eloá Jacobina. Rio de Janeiro: Bertrand Brasil, 2008.

MORIN, Edgar. Introdução ao pensamento complexo. 5. ed. Tradução de Elaine Lisboa. Porto Alegre: Sulina, 2015.

NADALIBI, Nadya; VEINOT, Tiffany C.; KIM, Young-Ho; SCHNEIDER, Nathan. Conceptualizing Algorithmic Stigmatization. *Proceedings of the ACM on Human-Computer Interaction*, v. 7, n. CSCW1, p. 1-27, 2023. DOI: 10.1145/3544548.3580970. Acesso em 18/10/2025.

NEWELL, Allen; SIMON, Herbert A. **GPS, a program that simulates human thought**. Santa Monica, CA: RAND Corporation, 1961. Disponível em: [https://iif.library.cmu.edu/file/Simon\\_box00064\\_fld04907\\_bd10001\\_doc0001/Simon\\_box00064\\_fld04907\\_bd10001\\_doc0001.pdf](https://iif.library.cmu.edu/file/Simon_box00064_fld04907_bd10001_doc0001/Simon_box00064_fld04907_bd10001_doc0001.pdf). Acesso em 26/08/2024.

NICCOL, Andrew. **O Preço do Amanhã** (In Time). EUA: 20th Century Fox, 2011. Filme.

NICOLELIS, Miguel. "Inteligência Artificial não é nem inteligente nem artificial". Entrevista ao Jornal do Comércio. Porto Alegre, out. 2023. Disponível em: <https://www.jornaldocomercio.com/especiais/feira-do-livro/2023/10/1128887-inteligencia-artificial-nao-e-nem-inteligente-nem-artificial-diz-nicolelis.html>. Acesso em: 24 maio 2026.

O'NEIL, Cathy. **Algoritmos de Destruição em Massa: com o Big Data aumenta a desigualdade e ameaça a democracia**. Rio de Janeiro: Editora Rua do Sabão, 2020.

PANNUNZIO, Pedro. **Demissões no Itaú: programa que vigia funcionários foi usado sem aviso e monitorava teclado, cliques e navegação**. Intercept Brasil, 14 set. 2025. Disponível em: <https://www.intercept.com.br/2025/09/14/demissoes-no-itaui-programa-que-vigia-funcionarios-foi-usado-sem-aviso-e-monitorava-teclado-cliques-e-navegacao/>. Acesso em: 19 out. 2025.

PARISER, Eli. **O Filtro Invisível: O que a Internet está Escondendo de Você?** Rio de Janeiro: Editora Zahar, 2012.

PASQUALE, FRANK. Two Narratives of Platform Capitalism. **Yale Law & Policy Review**, Yale, 2016. Disponível em: [https://heinonline.org/hol/cgi-bin/get\\_pdf.cgi?handle=hein.journals/yalpr35&section=14](https://heinonline.org/hol/cgi-bin/get_pdf.cgi?handle=hein.journals/yalpr35&section=14). Acesso em: 8 set. 2024.

PATNAIK, Subrat; REINICKE, Carmen. Nvidia se torna a empresa de maior valor de mercado do mundo com o avanço da IA. **Bloomberg Línea**, [S.l.], 2024. Disponível em: <https://www.bloomberglinea.com.br/mercados/nvidia-se-torna-a-empresa-de-maior-valor-de-mercado-do-mundo-com-o-avanco-da-ia/>. Acesso em: 3 set. 2024.

PERKOSKI, Norberto. O processo de criação poética. **Revista de Letras**, Goiânia, v. 192, n. 1-10, 2013. Disponível em: <http://textopoetico.emnuvens.com.br/rtp/article/view/109>. Acesso em: 11 set. 2024.

PINHEIRO, Patrícia Peck. **#DireitoDigital**. 6. ed. São Paulo: Saraiva, 2016.

POLO, Fernanda. A criatividade está ameaçada? Tecnologia pode ser aliada de humanos, mas é preciso vencer a preguiça. **GZH**, Porto Alegre, 2024. Disponível em: <https://gauchazh.clicrbs.com.br/comportamento/noticia/2024/07/a-criatividade-esta-ameacada-tecnologia-pode-ser-aliada-de-humanos-mas-e-preciso-vencer-a-preguica-cly7gd9k701n601209058rdr5.html>. Acesso em: 12 set. 2024.

POPE, Audrey. NYTv. OpenAI: The times's about-face. **Harvard Law Review**, [S.l.], 2024. Disponível em: <https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/>. Acesso em: 4 out. 2024.

POR QUE A TESLA é na verdade uma grande empresa de dados. **It Forum**, [S.l.], 2024. Disponível em: <https://itforum.com.br/por-que-a-tesla-e-na-verdade-uma-grande-empresa-de-dados/>. Acesso em: 3 set. 2024.

PRESSE, France. Mais de 40 estados nos EUA processam Meta por prejudicar a saúde de menores. **G1 Globo**, São Paulo, 24 out. 2023. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/10/24/mais-de-40-estados-nos-eua-processam-meta-por-prejudicar-a-saude-de-menores.ghtml>. Acesso em: 10 set. 2024.

PRESSFIELD, Steven. A guerra da arte: vença seus bloqueios internos e seja criativo. Tradução de Geni Hirata. São Paulo: Ediouro, 2002.

RAMIS, Harold. **O Feitiço do Tempo** (Groundhog Day). Direção: Harold Ramis. EUA: Columbia Pictures, 1993. 1 DVD (101 min), son., color.

ROSSETTI, Regina; GARCIA, Kethly. Inteligência artificial generativa: questões jurídicas e éticas em torno do ChatGPT. **VirtuaJus**, Belo Horizonte, v. 8, n. 15, p. 253-264, 2023. Disponível em: <https://periodicos.pucminas.br/index.php/virtuajus/article/download/30769/21848>. Acesso em: 24 maio 2024.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier Editora, 2013.

ROTHSTEIN, Mark A. Genetic Privacy: Past, Present, and Future. *Journal of Law and the Biosciences*, Oxford: Oxford University Press, v. 6, n. 1, p. 1–36, 2018. Disponível em [https://academic.oup.com/jlb/article/6/1/1/5489401?utm\\_source](https://academic.oup.com/jlb/article/6/1/1/5489401?utm_source). Acesso em 18/10/2025.

SÁNCHEZ-MONEDERO, Javier; DENCİK, Lina; EDWARDS, Lilian. What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. **Information, Communication & Society**, [S.l.], v. 23, n. 14, p. 458-468, 2020. Disponível em: <https://dl.acm.org/doi/10.1145/3351095.3372849>. Acesso em: 15 abr. 2024.

SANKARAN, Vishwam. ChatGPT inventa falso escândalo de assédio sexual e nomeia verdadeiro professor de direito como acusado. **Independent**, [S.l.], 2023. Disponível em: <https://www.independent.co.uk/tech/chatgpt-sexual-harassment-law-professor-b2315160.html>. Acesso em 10/09/2024.

SANTOS, Jéssica Guedes. Reconhecimento Facial: Entre a Criminologia, a Mídia e a LGPD Penal. **Revista InternetLab**, São Paulo, v. 2, n. 1, jun. 2021, p. 214-232. Disponível em: <https://revista.internetlab.org.br/wp-content/uploads/2021/07/Reconhecimento-facial-entre-a-criminologia-a-midia-e-a-LGPD-penal.pdf>. Acesso em: 10 set. 2024.

SARDINHA, Adriane Nascimento Celestino. Direitos autorais e inteligência artificial: uma análise do cenário atual e perspectivas para o futuro. **Revista Científica dos Estudantes de Direito da UFRGS**, João Pessoa, v. 8, n. 1, p. 152-180, dez. 2023. Disponível em: <https://seer.ufrgs.br/resseveraverumgaudium/article/view/137800>. Acesso em: 16/09/2024.

SCHREIBER, Anderson. “PL da Inteligência Artificial cria fratura no ordenamento jurídico brasileiro”. **JOTA Info**. 02 nov. 2021. Disponível em <https://www.jota.info/opinioe->

analise/colunas/coluna-do-anderson-schreiber/pl-inteligencia-artificial-cria-fratura-no-ordenamento-juridico. Acesso em 19 out. 2025.

SCHWAB, Klaus. **A Quarta Revolução Industrial**. World Economic Forum. São Paulo: Edipro, 2016.

SCOTT, Ridley (Direção). **Blade Runner, O Caçador de Andróides**. Estados Unidos: Warner Bros., 1982. 117 min. Filme.

SEIDL, Timo. The politics of platform capitalism: A case study on the regulation of *Uber* in New York. **Regulation & Governance**, [S.l.], 2020. Disponível em <https://onlinelibrary.wiley.com/doi/abs/10.1111/rego.12353>. Acesso em: 8 set. 2024.

SENADO FEDERAL. Comissão Temporária Interna sobre Inteligência Artificial (CTIA). Relatório Final do PL 2338/2023. Brasília, DF, 2024. Disponível em: <https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-inteligencia-artificial-texto-vai-a-camara>. Acesso em: 19 out. 2025.

SENADO FEDERAL. Comissão de Juristas responsável pela elaboração de anteprojeto de atualização do Código Civil. Brasília: Senado Federal, 2024. Disponível em: <https://legis.senado.leg.br/atividade/comissoes/comissao/2630/>. Acesso em: 27 out. 2025.

SILVEIRA, Stefanie Carlan da; MORISSO, João Gabriel Danesi. O uso de algoritmos na mídia programática. **Revista Científica da Faculdade de Comunicação da FAAM**, São Paulo, v. 6, n. 1, p. 71-82, jan./abr. 2018. Disponível em: <https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/710>. Acesso em: 13 set. 2024.

SILVER, David *et al.* Mastering the game of Go with deep neural networks and tree search. **Nature**, [S.l.], v. 529, n. 7587, p. 484-489, 2016. Disponível em <https://www.nature.com/articles/nature16961%7D>. Acesso em 13/09/2024.

SOUTO, Érica Araújo. Direito Autoral e Internet. **Portal de Trabalhos Acadêmicos**, Recife, v. 3, n. 1, 2011. Disponível em <https://revistas.faculdedamas.edu.br/index.php/academico/article/view/2728>. Acesso em: 4 set. 2024.

SOUZA, Carlos Affonso. Como Japão liberou conteúdo protegido por direito autoral para treinar IA. UOL, São Paulo, 2023. Disponível em: <https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2023/06/13/japao-afasta-direitos-autorais-para-impulsionar-inteligencia-artificial.htm>. Acesso em: 4 set. 2024.

SPIELBERG, Steven. **Minority Report**. EUA: 20th Century Fox, DreamWorks Pictures, 2002. 1 DVD (145 min.), son., color.

STARR, Sonja B. Evidence-Based Sentencing and the Scientific Rationalization of Discrimination. **Stanford Law Review**, [S.l.] v. 66, p. 803-872, 2014. Disponível em <https://repository.law.umich.edu/cgi/viewcontent.cgi?article=2189&context=articles>. Acesso em: 29 ago. 2024.

SUA OPINIÃO política pode te impedir de acessar a internet – conheça o crédito social Chinês. Brasil Paralelo, [S.l.], 2023. Disponível em: <https://www.brasilparalelo.com.br/noticias/sua-opiniao-politica-pode-te-impedir-de-acessar-a-internet-conheca-o-credito-social-chines>. Acesso em: 29 ago. 2024.

SUPREMA CORTE do Reino Unido decide que IA não pode ser inventora de patentes. Canal do Meio, [S.l.], 20 dez. 2023. Disponível em: <https://www.canalmeio.com.br/2023/12/20/suprema-corte-do-reino-unido-decide-que-ia-nao-pode-ser-inventora-de-patentes/>. Acesso em: 6 set. 2024.

SILVA, Tarcízio. Racismo algorítmico: inteligência artificial e discriminação nas redes digitais. São Paulo: Edições Sesc, 2022.

THE ECONOMIST. The world's most valuable resource is no longer oil, but data. **The Economist**, [S.l.], 2017. Disponível em: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Acesso em: 13 set. 2024.

TJSC. TJSC multa autor de recurso por jurisprudência falsa gerada por inteligência artificial. Tribunal de Justiça de Santa Catarina, 19 fev. 2025. Disponível em: <https://www.tjsc.jus.br/web/imprensa/-/tjsc-multa-autor-de-recurso-por-jurisprudencia-falsa-gerada-por-ia>. Acesso em: 24 maio 2026.

TREVASKES, Susan; BERNOT, Ausma. Surveillance infrastructure in China: Key concepts and mechanisms enhancing the Party-state's governance ambitions. **Global Media and China**, [S.l.], v. 8, n. 3, p. 327–342, 2023. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/20594364231171013>. Acesso em: 10 set. 2024.

TURING, Alan M. Computing Machinery and Intelligence. *In*: **The Essential Turing**. Oxford: Clarendon Press Oxford, 1950. Disponível em: <https://www.edwardfrenkel.com/turing-intelligence.pdf>. Acesso em: 10 set. 2024.

UNESCO. Recomendação sobre a ética da Inteligência Artificial. Brasil: Unesco, 2021. Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000381137_por). Acesso em: 4 out. 2024.

UNIÃO EUROPEIA. Diretiva (UE) 2024/2853 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, relativa à responsabilidade por produtos defeituosos (Product Liability Directive – PLD). Bruxelas, 2024. Disponível em: <https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng>. Acesso em: 27 out. 2025.

VERLE, Lenara. Deep Blue x Kasparov: Tecnologias do Imaginário. **Revista FAMECOS**, Porto Alegre, n. 9, p. 63-67, dezembro 1998. Disponível em: <https://revistaseletronicas.pucrs.br/index.php/revistafamecos/article/view/3011>. Acesso em: 3 set. 2024.

WACHOWICZ, Marcos; BIANCAMANO, Manuela Gomes Magalhães. Direito Autoral, Criatividade e Plágio na Economia Criativa. *In*: WACHOWICZ, Marcos (Org.). **Por que mudar a Lei de Direito Autoral?** Estudos e pareceres. Florianópolis: Fundação Boiteux, 2011. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/6719817.pdf>. Acesso em: 6 set. 2024.

WEST, Darrell M. **What happens if robots take the jobs?** The impact of emerging technologies on employment and public policy. Washington: Brookings Institution, 2015. Disponível em: <https://www.insidepolitics.org/brookingsreports/robots.pdf>. Acesso em: 9 set. 2024.

WEST, Emily. *Amazon: Surveillance as a Service*. **Surveillance & Society**, [S.l.], v. 17, n. 1/2, p. 27-33, 2019. Disponível em: <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/13008>. Acesso em: 6 set. 2024.

ZUBOFF, Shoshana. **A Era do Capitalismo de Vigilância: A Luta por um Futuro Humano na Nova Fronteira Digital**. Rio de Janeiro: Editora Intrínseca LTDA, 2021.

ZUBOFF, Shoshana. Surveillance Capitalism or Democracy? The Death Match of Institutional Orders and the Politics of Knowledge in Our Information Civilization. **Journal of Information Technology**, [S.l.], 2022. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/26317877221129290>. Acesso em: 4 set. 2024.

## GLOSSÁRIO

**Algoritmo** — Conjunto de instruções ou regras computacionais definidas para resolver um problema ou executar uma tarefa. Em sistemas de Inteligência Artificial, os algoritmos processam dados de entrada e produzem resultados a partir de padrões aprendidos durante o treinamento. Diferentemente do senso comum, algoritmos não "pensam" — executam operações matemáticas com base em regras previamente estabelecidas por seres humanos.

**AlphaGo** — Sistema de Inteligência Artificial desenvolvido pela Google DeepMind que, em 2016, derrotou o campeão mundial do jogo de Go, Lee Sedol. É considerado um marco histórico no desenvolvimento da IA por ter utilizado aprendizado por reforço para desenvolver estratégias inéditas, superando capacidades humanas em uma tarefa antes considerada exclusivamente cognitiva.

**Aprendizado de Máquina (Machine Learning)** — Subcampo da Inteligência Artificial em que sistemas computacionais aprendem a realizar tarefas a partir de dados, sem serem explicitamente programados para cada situação. O sistema identifica padrões nos dados de treinamento e os utiliza para fazer previsões ou tomar decisões em novos contextos.

**Aprendizado Profundo (Deep Learning)** — Técnica de aprendizado de máquina baseada em redes neurais artificiais com múltiplas camadas de processamento. Permite que sistemas de IA aprendam representações complexas de dados, sendo a base das principais aplicações atuais de IA generativa, reconhecimento de imagens e processamento de linguagem natural.

**Big Data** — Termo que designa conjuntos de dados de volume, velocidade e variedade tão elevados que não podem ser processados por ferramentas computacionais convencionais. Na economia digital, o Big Data é a matéria-prima dos sistemas de IA: quanto mais dados disponíveis para treinamento, maior tende a ser a capacidade preditiva dos modelos.

**Blockchain** — Tecnologia de registro distribuído que armazena dados em blocos encadeados de forma cronológica e imutável, sem a necessidade de uma autoridade central de controle. Amplamente associada às criptomoedas, também é utilizada em contratos inteligentes e sistemas de rastreabilidade.

**Bolhas de Filtro (Filter Bubbles)** — Fenômeno pelo qual algoritmos de recomendação criam ambientes informacionais personalizados que expõem o usuário predominantemente a conteúdos alinhados com suas preferências e histórico de navegação, limitando o acesso à diversidade de perspectivas. O conceito foi desenvolvido por Eli Pariser (2012) e está associado à polarização informacional e ao enfraquecimento do debate público plural.

**Capitalismo de Vigilância** — Conceito desenvolvido por Shoshana Zuboff para descrever um modelo econômico no qual o comportamento humano é coletado em escala massiva, transformado em dados e utilizado para prever e modificar condutas em benefício de interesses comerciais. Nesse modelo, a matéria-prima não é o trabalho físico, mas a experiência humana convertida em dado.

**Câmaras de Eco (Echo Chambers)** — Ambientes informacionais — físicos ou digitais — nos quais as pessoas são expostas apenas a opiniões e informações que confirmam suas crenças preexistentes, sem contato com perspectivas divergentes. Nas redes sociais, algoritmos de recomendação tendem a intensificar esse fenômeno ao priorizar conteúdos com os quais o usuário já demonstrou engajamento.

**Chatbot** — Sistema de IA projetado para simular conversas humanas por meio de interfaces de texto ou voz. Modelos como o ChatGPT são exemplos avançados de chatbots baseados em Grandes Modelos de Linguagem (LLMs), capazes de gerar respostas contextualmente coerentes em linguagem natural.

**CNN (Convolutional Neural Network — Rede Neural Convolutiva)** — Arquitetura de rede neural especialmente eficaz no processamento de dados visuais, como imagens e vídeos. As CNNs revolucionaram o campo da visão computacional e são amplamente utilizadas em sistemas de reconhecimento facial, diagnóstico médico por imagem e veículos autônomos.

**COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)** — Software desenvolvido pela empresa Northpointe e utilizado em tribunais norte-americanos para prever a probabilidade de reincidência criminal de réus. Tornou-se um dos casos mais citados de viés algorítmico ao ser demonstrado, em investigação da ProPublica (2016), que o sistema classificava erroneamente indivíduos negros como de alto risco em proporção significativamente maior do que indivíduos brancos.

**DABUS** — Sistema de Inteligência Artificial desenvolvido pelo pesquisador Stephen Thaler e apresentado como inventor em pedidos de patente nos Estados Unidos e no Reino Unido. Ambos os países rejeitaram os pedidos, consolidando o entendimento de que apenas pessoas humanas podem ser reconhecidas como inventoras ou autoras para fins legais.

**DALL·E** — Modelo de IA generativa desenvolvido pela OpenAI capaz de gerar imagens a partir de descrições textuais (prompts). Junto ao Midjourney e ao Stable Diffusion, é uma das principais ferramentas de criação de imagens por IA e está no centro do debate sobre direitos autorais e autoria em criações assistidas por máquina.

**Dados Pessoais** — Qualquer informação relacionada a uma pessoa natural identificada ou identificável, conforme definição da Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018) e do Regulamento Geral de Proteção de Dados da União Europeia (GDPR). No contexto da IA, dados pessoais são frequentemente utilizados para treinamento de modelos e personalização de serviços, o que levanta questões sobre privacidade, consentimento e autodeterminação informativa.

**GAN (Generative Adversarial Network — Rede Generativa Adversarial)** — Arquitetura de IA desenvolvida por Ian Goodfellow em 2014, composta por dois sistemas que competem entre si: um gerador, que cria conteúdo sintético, e um discriminador, que tenta identificar se o conteúdo é real ou gerado. Essa competição resulta em produções cada vez mais realistas, sendo a base técnica de muitas ferramentas de IA generativa.

**GPT (Generative Pretrained Transformer)** — Família de modelos de linguagem desenvolvida pela OpenAI, treinada em grandes volumes de texto para gerar linguagem natural. O ChatGPT é uma aplicação baseada nessa arquitetura. Os modelos GPT operam por probabilidade estatística — calculam qual palavra ou sequência é mais provável dado o contexto — sem compreensão semântica do conteúdo gerado.

**GPU (Graphics Processing Unit — Unidade de Processamento Gráfico)** — Componente de hardware originalmente desenvolvido para processamento de imagens, que se tornou essencial para o treinamento de modelos de IA por sua capacidade de realizar operações matemáticas em

paralelo em alta velocidade. A empresa Nvidia tornou-se uma das mais valiosas do mundo em razão da demanda por GPUs para aplicações de IA.

**LGPD (Lei Geral de Proteção de Dados)** — Lei nº 13.709/2018, que regulamenta o tratamento de dados pessoais no Brasil, estabelecendo princípios, direitos dos titulares e obrigações para controladores e operadores. Inspirada no GDPR europeu, a LGPD é o principal marco normativo brasileiro sobre privacidade e proteção de dados no contexto digital.

**LLM (Large Language Model — Grande Modelo de Linguagem)** — Modelo de Inteligência Artificial treinado em enormes volumes de dados textuais para compreender e gerar linguagem natural. LLMs como o GPT-4 e o Claude operam por cálculo de probabilidades sobre sequências de palavras, sem compreender o significado do conteúdo no sentido humano. São a base técnica de chatbots, assistentes virtuais e ferramentas de geração de texto.

**Machine Learning** — Ver Aprendizado de Máquina.

**Midjourney** — Ferramenta de IA generativa voltada à criação de imagens a partir de prompts textuais. Ficou amplamente conhecida ao ser utilizada por Jason Allen para criar a obra "Théâtre D'opéra Spatial", premiada na Feira Estadual do Colorado em 2022, desencadeando debate global sobre autoria, originalidade e direitos autorais em criações assistidas por IA.

**Modulação Comportamental** — Prática pela qual sistemas algorítmicos arquitetam o ambiente de escolha do usuário de forma a induzi-lo a determinados comportamentos sem que ele perceba a influência. Vai além do controle direto: em vez de proibir ou punir, a modulação opera por design — configurando interfaces, sequências de conteúdo e incentivos que tornam certas escolhas mais prováveis. É um dos conceitos centrais na análise do capitalismo de vigilância.

**Neurodireitos** — Conjunto de direitos fundamentais relacionados à proteção da privacidade mental, da autonomia cognitiva e da integridade neurológica diante dos avanços das neurotecnologias e da IA. O Chile foi o primeiro país a constitucionalizá-los, por meio da Lei nº 21.383/2021. No Brasil, a PEC 29/2023 propõe inclusão similar na Constituição Federal.

**NLP (Natural Language Processing — Processamento de Linguagem Natural)** — Campo da IA dedicado ao desenvolvimento de sistemas capazes de compreender, interpretar e gerar linguagem humana. É a tecnologia subjacente a chatbots, tradutores automáticos, assistentes de voz e modelos de geração de texto como o GPT.

**PageRank** — Algoritmo desenvolvido por Larry Page e Sergey Brin, fundadores do Google, que classifica páginas da internet com base no número e na qualidade dos links que apontam para elas. Embora frequentemente citado como exemplo de IA, o PageRank é tecnicamente um algoritmo de ranqueamento baseado em grafos — não envolve aprendizado de máquina em seu núcleo original — e sua menção nesta tese serve para ilustrar como sistemas algorítmicos distintos são frequentemente agrupados sob o termo genérico "Inteligência Artificial".

**Prompt** — Instrução ou comando fornecido por um usuário a um sistema de IA generativa para orientar a geração de um conteúdo específico. A qualidade, especificidade e criatividade do prompt influenciam diretamente o resultado produzido pela IA. No debate sobre direitos autorais, discute-se se prompts elaborados podem configurar expressão criativa protegível.

**Prompt Injection** — Técnica de manipulação de sistemas de IA generativa pela inserção de instruções ocultas ou estrategicamente formuladas para alterar o comportamento esperado do modelo. Essas instruções podem ser embutidas em documentos, sites ou metadados, induzindo a IA a priorizar comandos maliciosos em detrimento de suas diretrizes originais. No ambiente jurídico, representa risco à integridade de decisões judiciais apoiadas por IA.

**Racismo Algorítmico** — Conceito desenvolvido por Tarcízio Silva para descrever como sistemas de IA reproduzem e amplificam desigualdades raciais ao serem treinados em dados históricos marcados por discriminação estrutural. O racismo algorítmico não resulta necessariamente de intenção discriminatória, mas da naturalização de padrões de exclusão como se fossem dados neutros e objetivos.

**Reconhecimento Facial** — Tecnologia de visão computacional que identifica ou verifica a identidade de pessoas a partir de suas características faciais. Amplamente utilizada em segurança pública, controle de acesso e vigilância, é também um dos campos com maior incidência documentada de viés algorítmico, com taxas de erro significativamente maiores para pessoas negras e mulheres do que para homens brancos.

**Rede Neural Artificial** — Sistema computacional inspirado no funcionamento dos neurônios biológicos, composto por camadas de nós interconectados que processam e transmitem informações. As redes neurais são o fundamento técnico do aprendizado profundo e da maioria dos sistemas de IA generativa atuais.

**Reskilling** — Processo de requalificação profissional pelo qual trabalhadores adquirem novas competências para migrar para funções distintas daquelas que exerciam, em resposta às transformações impostas pela automação e pela IA no mercado de trabalho.

**RTB (Real-Time Bidding — Leilão em Tempo Real)** — Sistema automatizado de compra e venda de espaços publicitários digitais que ocorre em milissegundos, no momento em que uma página é carregada. Permite que anunciantes disputem em tempo real a exibição de anúncios para usuários específicos, com base em seus perfis de comportamento online. Embora frequentemente associado à IA, o RTB é tecnicamente um mecanismo de otimização econômica automatizada — não um sistema de aprendizado de máquina no sentido estrito.

**Upskilling** — Processo de aprimoramento e atualização de competências já existentes de um profissional, para que possa acompanhar as novas exigências tecnológicas do mercado de trabalho sem necessariamente mudar de função.

**Viés Algorítmico** — Distorção sistemática nos resultados produzidos por sistemas de IA decorrente de desequilíbrios, preconceitos ou insuficiências nos dados utilizados para treinamento ou nas escolhas de design do modelo. O viés algorítmico pode perpetuar e amplificar discriminações históricas, especialmente em contextos de tomada de decisão que afetam direitos fundamentais, como crédito, emprego, saúde e sistema de justiça.