Universidade Estadual de Londrina

FELIPE GOMES SILVA

PROTEÇÃO DE DADOS E INTELIGÊNCIA ARTIFICIAL: TRANSPARÊNCIA E *ACCOUNTABILITY* NA MITIGAÇÃO DA OPACIDADE ALGORÍTMICA E DA ASSIMETRIA INFORMACIONAL

Londrina
2026

FELIPE GOMES SILVA

PROTEÇÃO DE DADOS E INTELIGÊNCIA ARTIFICIAL: TRANSPARÊNCIA E *ACCOUNTABILITY* NA MITIGAÇÃO DA OPACIDADE ALGORÍTMICA E DA ASSIMETRIA INFORMACIONAL

Dissertação apresentada ao Programa de Pós-graduação em Direito Negocial da Universidade Estadual de Londrina-UEL como requisito parcial para a obtenção do título de Mestre.

Orientador: Profa. Dra. Tania Lobo Muniz

Londrina
2026

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

SILVA, FELIPE GOMES

A TUTELA JURÍDICA DA PROTEÇÃO DE DADOS PESSOAIS NA INTELIGÊNCIA ARTIFICIAL: A CENTRALIDADE DOS PRINCÍPIOS DE TRANSPARÊNCIA E *ACCOUNTABILITY* NA MITIGAÇÃO DA OPACIDADE ALGORÍTMICA E DA ASSIMETRIA INFORMACIONAL / FELIPE GOMES SILVA. - Londrina, 2025.

160 f.

Orientadora: TANIA LOBO MUNIZ.

Dissertação (Mestrado em Direito Negocial) - Universidade Estadual de Londrina, Centro de Estudos Sociais Aplicados, Programa de Pós-Graduação em Direito Negocial, 2025.

Inclui bibliografia.

1. Proteção de dados pessoais; Inteligência artificial; Transparência; *Accountability*; Governança global. - Tese. I. MUNIZ, TANIA LOBO. II. Universidade Estadual de Londrina. Centro de Estudos Sociais Aplicados. Programa de Pós-Graduação em Direito Negocial. III. Título.

CDU 34

FELIPE GOMES SILVA

PROTEÇÃO DE DADOS E INTELIGÊNCIA ARTIFICIAL: TRANSPARÊNCIA E *ACCOUNTABILITY* NA MITIGAÇÃO DA OPACIDADE ALGORÍTMICA E DA ASSIMETRIA INFORMACIONAL

Dissertação apresentada ao Programa de Pós-graduação em Direito Negocial da Universidade Estadual de Londrina-UEL como requisito parcial para a obtenção do título de Mestre.

BANCA EXAMINADORA

Profa. Orientadora
Tania Lobo Muniz
Universidade Estadual de Londrina - UEL

Prof. Membro 2
Patrícia Ayub da Costa
Universidade Estadual de Londrina - UEL

Prof. Membro 3
Renata Capriolli Zocatelli Queiroz
Universidade Estadual de Londrina - UEL

Londrina, 24 de fevereiro de 2025

GRADECIMENTOS

A Yasmin, minha noiva, pelo amor paciente e pela generosidade de compreender cada dia dedicado à redação desta dissertação. Obrigado por acolher minhas ausências, apoiar meus propósitos e estar sempre ao meu lado, mesmo nos momentos em que isso significou renúncias.

À minha família, que foi o alicerce silencioso e constante de toda a minha trajetória. Pelas oportunidades, pelos valores e pela confiança que me permitiram estudar, perseverar e chegar até aqui.

À Professora Renata Capriolli Zocatelli Queiroz, fonte permanente de inspiração intelectual e humana. Foi ela quem me incentivou a ingressar no mestrado e acreditou que eu poderia trilhar esse caminho. Sou profundamente grato.

À minha orientadora, Professora Tania Lobo Muniz, pela compreensão nos meus sumiços, pelas prorrogações de prazo acolhidas com serenidade e pela orientação precisa que tornou possível finalizar esta dissertação da melhor forma. Agradeço por sua paciência, pedagogia e rigor crítico.

À Professora Patrícia Ayub da Costa, que foi muito mais do que membro da banca: foi guia, apoio e presença constante em todas as etapas do mestrado. Seu cuidado, suas orientações e seu incentivo tiveram impacto decisivo na construção deste trabalho.

Por fim, agradeço a Deus, pela força nos momentos difíceis, pela clareza nos caminhos escolhidos e pela oportunidade de concluir mais esta etapa da minha vida com serenidade e propósito.

Escravidão com uma aparência de liberdade é a escravidão mais cruel – George Orwell

SILVA, Felipe Gomes. **Proteção de dados e inteligência artificial: transparência e *accountability* na mitigação da opacidade algorítmica e da assimetria informacional**. 2025. 160 f. Dissertação (Mestrado em Direito Negocial) – Universidade Estadual de Londrina, Centro de Estudos Sociais Aplicados, Programa de Pós-Graduação em Direito Negocial, Londrina, 2025.

RESUMO

Investiga como estruturar a efetiva tutela jurídica dos dados pessoais no contexto da inteligência artificial, diante dos desafios representados pela opacidade algorítmica e pela assimetria informacional. Inserido no campo do Direito Negocial, o estudo emprega revisão bibliográfica e análise normativa para compreender a evolução da liberdade, da privacidade e da proteção de dados e o surgimento da sociedade informacional e demonstrar como a inteligência artificial amplifica os desafios informacionais. A pesquisa demonstra que a opacidade algorítmica e a assimetria informacional tornam insuficientes os modelos jurídicos clássicos, concebidos para contextos analógicos. Sustenta-se a hipótese de que a transparência e a *accountability* devem constituir princípios estruturantes da regulação da inteligência artificial e da proteção de dados, fornecendo inteligibilidade mínima, redistribuição institucional do conhecimento e mecanismos de responsabilização aptos a mitigar riscos informacionais. Conclui-se que tais princípios somente alcançam eficácia quando articulados em uma arquitetura multinível de governança, técnica, normativa, institucional, corporativa e social, capaz de transformar comandos abstratos em práticas verificáveis com potencial para enfrentar os desafios informacionais. Conclui-se também que a consolidação dessa governança depende de avanços normativos, capacitação técnica de operadores jurídicos, pluralidade metodológica e cooperação internacional, assegurando a efetiva tutela da proteção de dados nas relações informacionais.

Palavras-chave: Proteção de dados pessoais; Inteligência artificial; Transparência; *Accountability*; Opacidade algorítmica.

SILVA, Felipe Gomes. **Data Protection and Artificial Intelligence: transparency and *accountability* in mitigating algorithmic opacity and informational asymmetry**. 2025. 160 pp. Dissertation (Master's Degree in Business Law) – State University of Londrina, Center for Applied Social Sciences, Graduate Program in Business Law, Londrina, 2025.

ABSTRACT

Examines how to structure the effective legal protection of personal data in the context of artificial intelligence, in light of the challenges posed by algorithmic opacity and informational asymmetry. Framed within the field of Business Law, the study employs bibliographic review and normative analysis to examine the evolution of freedom, privacy, and data protection, as well as the emergence of the informational society, and to demonstrate how artificial intelligence amplifies informational challenges. The research shows that algorithmic opacity and informational asymmetry render classical legal models, designed for analog contexts, insufficient. It sustains the hypothesis that transparency and accountability must constitute structuring principles of AI regulation and data protection, providing minimum intelligibility, institutional redistribution of knowledge, and mechanisms of responsibility capable of mitigating informational risks. It concludes that these principles only become effective when articulated within a multilevel governance architecture, technical, normative, institutional, corporate, and social, capable of transforming abstract commands into verifiable practices with the potential to address informational challenges. It also concludes that consolidating such governance depends on normative advancements, technical capacity-building for legal operators, methodological pluralism, and international cooperation, ensuring the effective protection of personal data in informational relations.

Key-words: Data protection; Artificial intelligence; Transparency; *Accountability*; Algorithmic opacity.

SUMÁRIO

1 INTRODUÇÃO

A consolidação da sociedade informacional, marcada pela centralidade dos dados e pela crescente interconexão entre indivíduos, instituições e sistemas tecnológicos, redefiniu profundamente as dinâmicas sociais contemporâneas. Nesse contexto, a circulação massiva de informações e a dependência estrutural de infraestruturas digitais passaram a integrar a própria lógica de organização econômica, política e comunicacional. A transformação não se limita à intensificação dos fluxos de dados, mas envolve a reconfiguração dos espaços de sociabilidade e das formas de exercício do poder, evidenciando novas tensões entre eficiência tecnológica, autonomia individual e proteção de direitos fundamentais.

Paralelamente à expansão das tecnologias digitais, a emergência de mecanismos de vigilância difusa ampliou os riscos associados à exposição informacional. A coleta sistemática de dados, a personalização comportamental e a análise preditiva tornaram-se práticas amplamente disseminadas em ambientes privados e institucionais, dando origem a estruturas sofisticadas de monitoramento. O tratamento intensivo de informações pessoais passou a ser operacionalizado por agentes dotados de capacidade tecnológica e econômica significativamente superior à dos indivíduos, o que produz cenários de vulnerabilidade informacional e fragilização da autodeterminação individual. Essas transformações desafiam os contornos tradicionais dos direitos à liberdade e à privacidade, exigindo sua reinterpretação em um ambiente marcado pela assimetria entre titulares e controladores de dados.

No campo da liberdade, vem à tona uma dualidade inerente à essa nova forma de sociedade tecnológica, enquanto a tecnologia pode empoderar os indivíduos, oferecendo novas oportunidades de liberdade e autodeterminação, ela também pode ser utilizada para exercer controle e vigilância, restringindo essas mesmas liberdades. Essa dualidade é central para a compreensão dos desafios contemporâneos da liberdade na era digital.

É nesse cenário que a inteligência artificial (IA) assume papel central. A adoção de sistemas capazes de processar grandes volumes de dados, identificar padrões e tomar decisões autônomas amplia exponencialmente o potencial de atuação informacional dos agentes que os controlam. Embora essas tecnologias ofereçam benefícios relevantes, como automação, eficiência e otimização de processos, também introduzem riscos inéditos. A opacidade dos modelos algorítmicos, a

dificuldade de rastreabilidade das decisões automatizadas e a impossibilidade de compreender plenamente os mecanismos internos de funcionamento de sistemas complexos criam um ambiente de incerteza quanto à sua conformidade com parâmetros éticos e jurídicos. Da mesma forma, a crescente assimetria informacional entre indivíduos e instituições reforça desigualdades, limita a capacidade de contestação dos titulares e tensiona a efetividade das normas destinadas à proteção de dados pessoais.

No âmbito do Direito Negocial, tais desafios assumem contornos próprios. A utilização de algoritmos em processos decisórios contratuais, operações comerciais e modelos de negócios baseados em dados intensificou a interdependência entre autonomia privada e estruturas tecnológicas opacas. Relações negociais mediadas por plataformas digitais, contratos automatizados e práticas comerciais orientadas por perfis comportamentais tornam-se cada vez mais frequentes, ao mesmo tempo em que ampliam a complexidade informacional. A tensão entre liberdade contratual e proteção de direitos fundamentais se agrava quando a lógica da autorregulação privada se apoia em sistemas cujo funcionamento escapa ao conhecimento ou controle dos titulares. O resultado é a emergência de um ambiente negocial em que a assimetria entre as partes não decorre apenas de fatores econômicos, mas também da concentração informacional e da capacidade de exploração algorítmica.

A globalização dos fluxos de dados aprofunda essa problemática, uma vez que a circulação transnacional de informações e a atuação de agentes econômicos em múltiplas jurisdições desafiam a eficácia dos sistemas normativos nacionais. Diferentes regimes de proteção de dados, estruturas regulatórias assimétricas e métricas divergentes de responsabilização tornam-se obstáculos à proteção uniforme dos titulares. No plano das relações empresariais e internacionais, tal cenário aumenta a relevância de modelos normativos capazes de articular segurança jurídica, inovação tecnológica e tutela dos direitos fundamentais. É nesse contexto que se insere a presente pesquisa, vinculada à linha “Estado contemporâneo: relações empresariais e relações internacionais”, bem como ao projeto institucional voltado a analisar as incertezas e desafios

A partir desse panorama, emerge o problema central que orienta esta investigação: como estruturar a efetiva tutela jurídica da proteção de dados pessoais no contexto da inteligência artificial, diante dos riscos decorrentes da opacidade algorítmica e da assimetria informacional na sociedade informacional? A questão

envolve compreender as limitações dos modelos tradicionais de tutela, identificar os elementos que produzem vulnerabilidades informacionais e examinar as possibilidades de reorganização normativa aptas a lidar com os desafios trazidos pela tecnologia.

A hipótese adotada sustenta que a efetiva tutela jurídica da proteção de dados pessoais, no contexto da inteligência artificial, exige a centralidade dos princípios da transparência e da *accountability* como fundamentos estruturantes da regulação tecnológica. Esses princípios, quando adequadamente operacionalizados, permitem conferir inteligibilidade aos sistemas algoritmos, possibilitar sua fiscalização, reforçar a rastreabilidade das decisões automatizadas e promover a responsabilização dos agentes envolvidos. Assim, estabelecem o suporte dogmático necessário para enfrentar os desequilíbrios informacionais que caracterizam a sociedade contemporânea.

O objetivo geral da pesquisa consiste em demonstrar que a consolidação da transparência e da *accountability* como eixos estruturantes da regulação da inteligência artificial representa condição indispensável para a proteção de dados pessoais e para o equilíbrio das relações negociais e institucionais. Para alcançá-lo, são definidos três objetivos específicos: (i) examinar os fundamentos teóricos da liberdade, da privacidade e da proteção de dados na sociedade informacional; (ii) analisar os impactos da inteligência artificial sobre esses direitos, com ênfase nos fenômenos da opacidade algorítmica e da assimetria informacional; e (iii) investigar instrumentos jurídicos, técnicos e de governança capazes de concretizar os princípios da transparência e da *accountability*.

Em relação ao marco metodológico, opta-se pela realização de uma revisão sistemática de literatura associada a uma pesquisa documental, cujo objetivo é identificar e selecionar, de maneira criteriosa, os estudos mais relevantes sobre proteção de dados pessoais e os desafios jurídicos decorrentes da inteligência artificial, com foco nos riscos da opacidade algorítmica e da assimetria informacional, e nos princípios da transparência e da *accountability* como instrumentos jurídicos essenciais à tutela da proteção de dados no contexto da inteligência artificial.

O marco teórico organiza-se em quatro eixos: (i) autores que fundamentam a evolução da privacidade e da proteção de dados, como Warren e Brandeis, Westin e Solove; (ii) teóricos da sociedade informacional, como Lévy e Castells; (iii) autores que analisam vigilância e riscos informacionais, como Foucault, Rodotà e Bauman e (iv)

autores que enfrentam os desafios específicos da inteligência artificial, mobilizam-se ainda as contribuições de Virginia Dignum, Kate Crawford, Stuart Russell e Peter Norvig. Esses referenciais sustentam a posição adotada nesta pesquisa, segundo a qual a opacidade algorítmica e a assimetria informacional exigem a centralidade normativa da transparência e da *accountability*.

No que se refere à organização do trabalho, a dissertação está estruturada em quatro capítulos, organizados de forma a assegurar a progressão lógica da análise. O primeiro capítulo desenvolve uma reflexão sobre liberdade, privacidade e vigilância na sociedade da informação, abordando inicialmente os fundamentos filosóficos da liberdade na era digital, a evolução histórica da privacidade e seus fundamentos conceituais, para em seguida examinar a sociedade da informação e a sociedade em rede, delimitando seus contornos em relação à sociedade da vigilância e à transição da privacidade para a proteção de dados. O segundo capítulo aprofunda os desafios da inteligência artificial na sociedade informacional, expondo a evolução tecnológica e o papel do big data, apresentando definições, características e técnicas fundamentais da IA, bem como analisando os riscos jurídicos estruturantes derivados da opacidade algorítmica e da assimetria informacional. O terceiro capítulo dedica-se à análise dos princípios da transparência e da *accountability* como fundamentos normativos estruturantes da tutela jurídica da proteção de dados pessoais no contexto da inteligência artificial. Nesse ponto, examina-se a noção de transparência algorítmica e de *accountability* algorítmica, avançando para a discussão dos instrumentos técnicos e de governança voltados à efetivação desses princípios e dos desafios de sua aplicabilidade prática. Por fim, o quinto capítulo apresenta a conclusão, em que o problema de pesquisa é retomado, a hipótese é avaliada, os principais achados são sintetizados com suas implicações para o Direito Negocial, e são indicadas contribuições científicas, limitações e recomendações para pesquisas futuras.

2 LIBERDADE, PRIVACIDADE E VIGILÂNCIA NA SOCIEDADE DA INFORMAÇÃO

A compreensão adequada do problema investigado exige revisitar três categorias conceituais fundamentais, liberdade, privacidade e vigilância, que, embora tenham origens históricas distintas, convergem de modo decisivo no contexto da sociedade da informação. Apesar de originados em momentos históricos diversos, esses conceitos convergem no cenário contemporâneo, em que a circulação massiva de informações redefine relações sociais, econômicas e jurídicas. A transição para uma sociedade estruturada pela comunicação digital, pelo processamento de dados e pela interconexão em larga escala transformou radicalmente as condições de exercício da liberdade, as expectativas de privacidade e os mecanismos de vigilância. Essa transformação demanda revisitar os fundamentos filosóficos dessas categorias para compreender como se reconfiguram em um ambiente marcado pela centralidade dos fluxos informacionais.

É nesse contexto que se insere o presente capítulo. Seu objetivo é reconstruir as bases teóricas que vinculam liberdade, privacidade e vigilância, demonstrando como a evolução histórica desses conceitos prepara o terreno para compreender os desafios específicos da sociedade da informação. A partir dessa reconstrução, será possível analisar, nos itens seguintes, de que maneira a privacidade se tornou um instrumento de proteção da liberdade e como a vigilância passou a assumir novas configurações em uma sociedade organizada por redes, dados e informação.

2.1 LIBERDADE NA ERA DIGITAL

O campo de estudo da liberdade é um universo inteiro com inúmeras perspectivas filosóficas, políticas e jurídicas que atravessam a história. Diante dessa complexidade e da multiplicidade de abordagens possíveis, a presente investigação não pretende esgotar o tema, mas sim trazer conceitos bases que apoiem o desenvolvimento da ideia de privacidade como liberdade e fundamente as discussões que serão oportunamente apresentadas.

A liberdade, conceito central nas discussões filosóficas e jurídicas, revela distintas acepções históricas e teóricas, cada uma refletindo contextos políticos, sociais e culturais específicos. No contexto contemporâneo, profundamente marcado

pelo avanço das tecnologias digitais e da inteligência artificial (IA), torna-se imprescindível revisitar os fundamentos filosóficos e conceituais da liberdade, explorando suas diferentes vertentes e implicações normativas.

A reflexão sobre liberdade ocupa lugar central na tradição filosófica e jurídica, tendo assumido novas camadas de complexidade na era digital. O presente trabalho retoma esse debate a partir de uma perspectiva orientada pela proteção de dados, ampliando e revisitando argumentos previamente delineados em estudo anterior do autor sobre liberdade na era digital (SILVA; COSTA; MUNIZ, 2024).

Isaiah Berlin, ao distinguir os dois sentidos principais de liberdade, negativa e positiva, lançou as bases para uma abordagem que permanece atual frente aos desafios trazidos pela digitalização. A liberdade negativa, concebida como ausência de coerção externa, refere-se ao espaço de atuação em que o indivíduo pode agir sem interferência arbitrária. Já a liberdade positiva se relaciona à capacidade de autodeterminação, ou seja, ao domínio do sujeito sobre suas próprias escolhas e à sua aptidão de se autogovernar racionalmente (BERLIN, 2002, p. 229)

A liberdade negativa, na formulação clássica de Isaiah Berlin, corresponde ao espaço no qual o indivíduo pode agir sem ser submetido a impedimentos externos arbitrários. Trata-se da dimensão da não interferência, na qual a atuação de terceiros, seja do Estado, seja de particulares, não restringe escolhas que, em condições normais, estariam ao alcance da pessoa (Berlin, 2002, p. 229; 234). Esse entendimento delimita a liberdade como um território protegido contra intromissões injustificadas, enfatizando a importância de preservar uma zona mínima de ação individual.

Em contraste, a liberdade positiva remete à ideia de autodeterminação. Berlin (2002, p. 236) descreve esse sentido como o desejo humano de conduzir a própria vida segundo razões e finalidades escolhidas pelo próprio sujeito. Sob essa perspectiva, o indivíduo só é verdadeiramente livre quando possui capacidade efetiva de se autogovernar. Maria Ligia Elias (2014, p. 21) aprofunda essa leitura ao associar a liberdade positiva a um processo racional de autodomínio, no qual a autonomia deriva não apenas da ausência de obstáculos, mas da possibilidade de orientar a própria existência de modo refletido.

Embora representem dimensões distintas, os dois sentidos de liberdade não são antagônicos: há circunstâncias em que a ausência de interferência (liberdade negativa) funciona como condição necessária para que a autodeterminação (liberdade

positiva) se realize. Entretanto, como já discutido em estudo anterior deste autor (Silva; Costa; Muniz, 2024), essa dualidade não esgota a complexidade do debate contemporâneo sobre liberdade.

É nesse ponto que a crítica republicana se torna relevante. Quentin Skinner argumenta que não basta avaliar a liberdade pela ocorrência ou não de coerções concretas. Para ele, a liberdade é comprometida sempre que um indivíduo encontra-se vulnerável ao poder arbitrário de outro, ainda que esse poder não se manifeste ativamente (Skinner, 1998, p. 70-71). A mera possibilidade de interferência autoritária já configura uma forma de sujeição, pois coloca o sujeito em estado permanente de dependência. A liberdade, nessa chave, é ausência de dominação.

Essa reformulação altera significativamente o eixo do debate: ser livre não é apenas viver sem obstáculos visíveis, mas também estar protegido contra relações estruturais que permitam a qualquer agente, público ou privado, exercer poder arbitrário sobre a esfera individual. A concepção republicana, portanto, desloca o foco da não interferência para a necessidade de instituições, garantias e mecanismos de controle que impeçam a emergência de poderes assimétricos e não supervisionados.

Na mesma linha, Philip Pettit também entende como insuficiente a dicotomia entre liberdade positiva e negativa de Berlin, propondo uma terceira via: a liberdade como não dominação. Essa não exige apenas que o indivíduo esteja livre de obstáculos (como o conceito negativo), mas que esteja institucionalmente protegido contra qualquer forma de poder arbitrário. Ser livre, segundo Pettit, é não estar sujeito à vontade arbitrária de outrem, o que exige instituições democráticas, deliberação pública, e mecanismos de contestação e vigilância cívica (PETTIT, p. 61, 1997).

Nesse sentido, a leitura desenvolvida por Philip Pettit foi fundamental para a consolidação da teoria republicana contemporânea, pois, ao definir a liberdade como não dominação, desloca o foco da ausência de interferência concreta para a existência de estruturas de proteção contra a arbitrariedade. Como observa César Ramos (2007, p. 301-325), a dominação, nessa perspectiva, não exige a ocorrência efetiva de coerção, bastando que exista a possibilidade de um agente deter poder para interferir na esfera individual sem a devida justificação pública e controle institucional. A liberdade republicana, assim, exige condições objetivas de não sujeição, isto é, a ausência de subordinação à vontade unilateral de outros. Para que essa liberdade se realize, não basta garantir direitos formais ou ausência momentânea de coerção, é imprescindível assegurar que os indivíduos não estejam vulneráveis à

interferência arbitrária, ainda que esta não se concretize. A teoria de Pettit, portanto, não apenas revisita as bases do republicanismo clássico, como propõe um modelo normativo centrado na vigilância mútua, na transparência institucional e na responsabilização pública, condições indispensáveis para a promoção da liberdade como valor político e jurídico substantivo.

Essa concepção de liberdade como não dominação e a necessidade de estruturas que protejam os indivíduos contra isso será particularmente relevante para compreender a dimensão normativa da privacidade e da proteção de dados como instrumentos de defesa da liberdade no contexto da sociedade moderna.

Enquanto Pettit e Skinner desenvolvem a noção de liberdade como não dominação a partir de uma crítica ao poder institucional arbitrário, voltando-se para os mecanismos de controle estatal, John Stuart Mill (2011, p. 22) amplia esse debate ao identificar que as ameaças à liberdade não se restringem ao âmbito estatal. Mill adverte que a proteção contra a tirania do estado não é suficiente, sendo igualmente necessária a proteção contra “a tirania da opinião e do sentimento prevalente”, isto é, a imposição social de padrões morais e comportamentais que suprimem a individualidade e a diversidade de modos de vida. Com isso, Mill reafirma a centralidade da autonomia individual como condição para o bem-estar humano, sustentando que a diversidade de opiniões, valores e estilos de vida é essencial para o progresso da sociedade. Sua concepção de liberdade não se limita à ausência de coerção, mas exige um ambiente em que o indivíduo possa desenvolver-se livremente, sem ser coagido pelas expectativas homogeneizantes da maioria.

Essas diferentes perspectivas filosóficas permitem uma compreensão aprofundada e multidimensional do conceito de liberdade. Considerando a problemática deste estudo, é necessário relacionar esses conceitos com o avanço tecnológico.

O avanço tecnológico trouxe significativas contribuições para a ampliação das liberdades individuais sob diversas perspectivas filosóficas. Do ponto de vista da liberdade negativa, entendida como ausência de coerção externa, essa é expandida com as novas formas de comunicação e expressão digital que permitem que indivíduos comuniquem ideias e exerçam sua liberdade de associação em escala global, sem obstáculos imediatos, proporcionando um ambiente onde a censura e a restrição são minimizadas. Sob a ótica da liberdade positiva, associada à autodeterminação e ao autogoverno racional, as tecnologias digitais causam impacto

por meio de acesso ampliado à informação, com a disponibilidade dos saberes, o que traz maiores possibilidades à racionalidade do indivíduo, impactando em seu autogoverno racional. Isso permite que o homem possua mais ferramentas para ser o seu próprio senhor, trazendo liberdade de pensamento, enriquecendo a capacidade dos indivíduos de tomar decisões informadas. Ao facilitar a circulação de conhecimento e ideias, a tecnologia fortalece a autonomia pessoal e a capacidade de autogoverno, elementos centrais da liberdade positiva.

Já em relação à liberdade como não dominação, a tecnologia pode funcionar como ferramenta de emancipação, ao proporcionar um espaço de liberdade para que a sociedade possa manifestar suas opiniões, se organizar e lutar contra a tirania estatal. No aspecto social, John Stuart Mill (2011, p. 25) enfatiza que a liberdade está diretamente ligada ao livre desenvolvimento da individualidade e à possibilidade de o sujeito viver conforme suas convicções, mesmo quando contrariam os valores dominantes. A tecnologia, quando usada para ampliar o pluralismo e assegurar espaços de autonomia frente à opinião majoritária, torna-se, portanto, uma aliada da liberdade também sob essa perspectiva.

Apesar das potencialidades emancipatórias, o ambiente digital também introduz novas formas de restrição e controle que afetam diretamente as concepções contemporâneas de liberdade. No plano da liberdade negativa, a atuação de algoritmos opacos, práticas de vigilância persistente e censura indireta em plataformas digitais representam formas de interferência silenciosa, estatais ou privadas, que comprometem o livre exercício da ação individual. Já sob a perspectiva da liberdade positiva, a autonomia do sujeito é ameaçada pela manipulação de preferências, pela filtragem de informações e pela captura da atenção, que minam a capacidade de tomada de decisões conscientes e informadas. Em um cenário marcado pela coleta massiva de dados e pela predição comportamental, o sujeito torna-se objeto de programação algorítmica, e não mais agente livre.

Do ponto de vista da liberdade como não dominação, o risco se agrava ainda mais: indivíduos passam a viver sob a constante possibilidade de interferência por parte de entidades privadas que operam sem transparência, *accountability* ou controle democrático, o que configura uma forma estrutural de sujeição. A tirania da opinião, denunciada por Mill, ressurge amplificada pela arquitetura das redes sociais, onde a homogeneização de discursos, a vigilância mútua e o silenciamento por pressão social limitam o livre desenvolvimento da personalidade. Para Mill, a liberdade está

intrinsecamente ligada à diversidade de modos de vida e à proteção contra coerções morais e culturais da maioria, nesse sentido, o ambiente digital, quando regido por lógicas de uniformização, ranking e exclusão, compromete diretamente a individualidade, a autonomia e o progresso ético e intelectual da sociedade.

Isso ressalta a dualidade intrínseca à tecnologia e liberdade. Enquanto as tecnologias digitais promovem a liberdade e a democratização da informação, elas também podem ser instrumentalizadas para exercer controle e vigilância. Esse paradoxo destaca a necessidade de uma vigilância contínua sobre como essas tecnologias são usadas e regulamentadas.

Essas diferentes perspectivas filosóficas permitem uma compreensão aprofundada e multidimensional do conceito de liberdade. Juntas, destacam que uma visão completa da liberdade no contexto contemporâneo com todas suas complexidades.

Essa base conceitual será essencial para a continuidade da investigação, especialmente ao se abordar, nos capítulos seguintes, a evolução histórica da privacidade, o surgimento da sociedade da informação, sua transformação em sociedade de vigilância e os desafios contemporâneos trazidos pela inteligência artificial. Cada uma dessas dimensões será analisada com maior profundidade, de modo a evidenciar como a proteção da liberdade, em suas múltiplas expressões, exige respostas normativas compatíveis com as complexidades do mundo digital.

Em suas múltiplas expressões, no contexto contemporâneo e digital, a liberdade pode ser relacionada com a privacidade. A privacidade protege o espaço da não interferência (liberdade negativa), resguarda a autonomia decisional e o autogoverno informacional (liberdade positiva), e representa uma salvaguarda contra formas arbitrárias e invisíveis de controle social e tecnológico (liberdade como não dominação). Além disso, constitui condição indispensável para o livre desenvolvimento da personalidade, tal como propõe Mill, ao garantir que os indivíduos possam construir identidades plurais, desenvolver valores próprios e resistir às imposições homogeneizantes da sociedade digital. Assim, compreender a privacidade como um prolongamento da liberdade é fundamental para a construção de uma teoria jurídica que seja capaz de enfrentar os desafios contemporâneos. É com essa perspectiva que se avança, no capítulo seguinte, à análise da evolução histórica e dos fundamentos conceituais da privacidade, a fim de compreender sua natureza, função e relevância no mundo digital.

2.2 PRIVACIDADE NA ERA DIGITAL

O primeiro conceito de privacidade filosoficamente relevante foi trazido como “o direito a ficar só ou o direito de ser deixado em paz” por Warren e Brandeis na sua obra *The Right to Privacy* publicada em 1890. Nesta obra, os autores preocupados com as novas complexidades sociais como o aumento das interações comerciais, o crescimento da imprensa e do jornalismo sensacionalista, e o desenvolvimento de novas tecnologias como a fotografia instantânea, defenderam que o ordenamento jurídico precisava reconhecer e proteger a esfera íntima do indivíduo, isto é, a possibilidade de conduzir a própria existência sem interferências ou publicizações indevidas. Esse entendimento, em síntese, elevou o “direito de ficar só” (the right to be let alone) ao patamar de um atributo essencial da personalidade, análogo as proteções contra agressões físicas, mas voltado às dimensões afetivas e psíquicas do sujeito (WARREN; BRANDEIS, 2013).

Esta definição apesar de ser historicamente importante, não foi suficiente para definir privacidade ao longo do tempo pois foi construída em um momento histórico onde houve o surgimento das primeiras fotografias instantâneas, porém, o desenvolvimento de novas tecnologias não se limitou a tal invenção.

Assim, com o passar dos anos, diversos autores dedicaram-se ao estudo do tema, um dos estudos acadêmicos mais relevantes sobre o assunto foi a obra *Privacy and Freedom* de Alan Westin em 1967. Westin (1967, p. 22) que define a privacidade como o “reivindicar de indivíduos, grupos ou instituições o direito de decidir por si mesmos quando, como e até que ponto informações a seu respeito serão comunicadas a outros”. Vale ressaltar que essas informações, no período de escrita do texto, eram quase todas registradas de forma analógica, como cartas, formulário, relatórios, dossiês e microfilmes em casos de fotografias. Westin enfatiza que tal direito de controlar “quando, como e até que ponto” essas informações aparecem perante terceiros é crucial para o bem-estar psicológico, a liberdade de associação, a formação de opiniões autênticas e a preservação de relacionamentos de confiança

Este conceito introduziu a primeira relação entre privacidade e o contexto informacional, o que é especialmente relevante para o fenômeno da contemporaneidade da sociedade da informação e para o marco teórico-conceitual desta pesquisa.

Após o lançamento da obra de Westin surgiram mais pesquisas sobre o

assunto correlacionando privacidade e o contexto informacional, como *The assault on privacy* de A. R. Miller em 1970, *Invasion of privacy: a clarification of concepts* de L. Lusky em 1971 e *Elaboratori elettronici e controllo sociale* de Stéfano Rodotá em 1972. Para esta pesquisa, adota-se como melhor compreensão de privacidade o conceito trazido por Rodotà como “o direito de manter o controle sobre suas próprias informações e de determinar a maneira de construir sua própria esfera particular”. Também destaca que a autodeterminação informativa foi reconhecida pela primeira vez em 1983 pela Corte Constitucional Alemã (RODOTÀ, 2008, p. 15).

Assim, a privacidade como “direito de ficar só” perde lugar para uma privacidade com núcleo conceitual atrelado pela possibilidade de cada sujeito controlar suas informações. Esta definição trazida por Rodotà é essencialmente importante para o surgimento de um conceito extremamente importante ao tema que é o da “autodeterminação informativa” citada por este com seu primeiro reconhecimento oficial em 1983.

Ao examinar esta decisão mencionada acima, essa é relacionada a uma sentença proferida pelo Tribunal Constitucional da Alemanha e trata de questão referente ao recenseamento da população, das profissões, das residências e dos locais de trabalho. Tal decisão reconheceu, pela primeira vez, o direito fundamental à autodeterminação informativa, fundamentado nos artigos 2º, §1º (livre desenvolvimento da personalidade) e 1º, §1º (dignidade humana) da Lei Fundamental alemã. Nesse contexto, destacou-se que as novas tecnologias relacionadas ao processamento eletrônico de dados permitiriam a criação de perfis detalhados da personalidade dos indivíduos sem seu controle adequado, exigindo, portanto, que se garantisse ao cidadão o direito de decidir sobre o modo e a extensão da coleta e utilização de seus dados pessoais (MENDES, 2020, p. 11-12).

O recenseamento da população também pode ser denominado censo demográfico, no Brasil este estudo é feito pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e possui o seu primeiro registro em 1872 no estudo “Recenseamento do Brazil em 1872”, onde pessoas eram quantificadas e perfiladas (classificadas em perfis) de acordo com seus estados civis, raças, religião, nacionalidade e nível de instrução ainda sob o regime de Império do Brazil (IBGE, 1874). O recenseamento tem seus primeiros registros históricos no Egito e na antiga Roma e tinha como objetivo obter informações para fins de impostos, serviço militar e outros assuntos governamentais, porém, tem início como prática regular e estruturada

na Europa no século XIX. (MARTINS, 2014)

Examina-se que a coleta massiva de dados e elaboração de perfis dos indivíduos tem seu início muito tempo antes do surgimento do conceito oficial de autodeterminação informativa. Este momento histórico inaugura oficialmente a ideia de privacidade, em seu sentido amplo pela capacidade do sujeito de controlar suas informações, como liberdade. Essas mudanças históricas demonstram a dificuldade de se conceituar privacidade ao longo do tempo, pois essa definição haveria que acompanhar as constantes mudanças da sociedade.

Em estudos mais recentes, Daniel J. Solove, em sua obra *Understanding Privacy* publicada na Universidade de Havard, afirma que o conceito de privacidade ainda é extremamente amplo, complexo e difícil de delimitar com precisão. Segundo o autor, embora a privacidade seja reconhecida universalmente como um direito fundamental e essencial à liberdade e à democracia, seu conteúdo permanece incerto e sujeito a diversas interpretações, e destaca que a dificuldade conceitual é agravada pela revolução tecnológica e pela capacidade crescente de coleta, processamento e disseminação massiva de informações pessoais. Em razão disso, argumenta que é inadequado buscar uma definição singular e rígida para a privacidade, já que tal conceito engloba situações diversas e frequentemente desconectadas entre si. Assim, Solove sugere que a privacidade seja compreendida por meio de uma abordagem pluralista e pragmática, analisando-se contextos específicos e identificando-se claramente os tipos concretos de danos que resultam de violações à privacidade, para melhor orientar as decisões jurídicas e políticas públicas sobre o tema. Essa nova realidade tornou o debate sobre a privacidade ainda mais urgente e complexo, devido às variadas situações práticas envolvendo tecnologias de vigilância, meios de comunicação e processamento digital de dados (SOLOVE, 2008, p. 1-11).

No Brasil, essa visão é compartilhada por Danilo Doneda que descreve que a indefinição quanto ao conceito de privacidade é uma característica intrínseca dessa matéria e que não deve ser entendida como um defeito ou obstáculo. Doneda destaca que “Talvez uma “definição” do que seja a privacidade não seja propriamente a principal questão a ser enfrentada” (Doneda, 2020, p. 77). Contudo, este autor também entende que a definição trazida por Stefano Rodotà como “direito de manter controle sobre as próprias informações e de determinar as modalidades de construção da esfera privada” é a mais harmônica com o desenvolvimento da matéria, colocando a informação como elemento objetivo, trazendo a autodeterminação informativa, bem

como a construção da esfera privada como finalidade, como aspecto do livre desenvolvimento da personalidade (DONEDA, 2020, p. 96).

Diante da evolução conceitual da privacidade, desde a sua formulação inicial como o simples direito de ficar só até a proteção da autodeterminação informativa e como aspecto do livre desenvolvimento da personalidade proposta por Rodotà, percebe-se que o núcleo da questão se desloca para o controle efetivo que o indivíduo exerce sobre suas próprias informações, colocando a informação de forma central nesta discussão. Essa transição reflete diretamente os desafios contemporâneos impostos pela denominada sociedade da informação de Pierre Lévy ou da sociedade em rede da Manuel Castells, na qual os dados pessoais são coletados, tratados e compartilhados em larga escala e com rapidez sem precedentes, afetando profundamente a esfera de liberdade dos indivíduos. Nesse contexto, torna-se essencial examinar como a privacidade atua como pilar da liberdade frente aos desafios específicos impostos pela sociedade da informação.

2.3 Sociedade da Informação e Sociedade em Rede

Como identificado anteriormente, o primeiro conceito de privacidade surgiu de forma contemporânea ao surgimento da fotografia instantânea em meados de 1890, porém, o desenvolvimento de novas tecnologias não se limitou a tal invenção, e os estudos sobre o assunto se tornaram calorosos no final da década de 1960 e início da década de 1970.

Estas datas coincidem com a revolução informacional, que segundo Manuel Castells em sua obra A Sociedade em Rede (1942, p. 64-70) foi iniciada na década de 1970 onde a sociedade passou por uma transformação histórica através de um novo paradigma tecnológico, organizado com base na tecnologia da informação, com o advento da internet nos Estados Unidos da América. Segundo este, esta revolução da tecnologia da informação levou a um processo de reestruturação do sistema capitalista a partir da década de 80, concretizando um novo estilo de produção, comunicação, gerenciamento e vida, o que nos levou ao pós-industrialismo ou informacionalismo através da sociedade informacional. Sua característica informacional refere-se ao fato de que a produtividade e a competitividade agora não são relacionadas a produção de bens de consumo, mas sim quanto a capacidade de

gerar, processar e aplicar de forma inteligente as informações, o que é traduzida pela célebre frase “Dados são o novo petróleo” de Clive Humby. (Castells, 1942, p. 135).

Como pano teórico para esta pesquisa, além da Sociedade em Rede concebida por Manuel Castells, tem-se a Sociedade da Informação trazida por Pierre Lévy. Há consonâncias de datas entre tais autores, pois Lévy em sua obra Cibercultura (1999, p. 23-31) estabelece como uma data de virada fundamental na sociedade os anos 70, em razão do surgimento do microprocessador que possibilitou o avanço meteórico dos processos econômicos e sociais de grande amplitude, bem como do computador pessoal na década de 1980, que aumentaram a capacidade de comunicação das informações e dos indivíduos, com a emergência do “ciberespaço”.

A partir desse momento, o mundo, anteriormente analógico, assume também uma dimensão virtual, onde a informação, já intimamente relacionada ao conceito de privacidade, passa a ser digitalizada e processada em larga escala, em razão da aprimoração da informática que é a ciência que estuda o conjunto de informações e conhecimentos por meios digitais.

A informática reúne técnicas para digitalizar, armazenar, transportar, processar e tratar a informação. Os órgãos de tratamento de informação tornam-se processadores dentro de chips que executam essas tarefas em grande velocidade. Digitalizar uma informação consiste em transformá-la em números, imagens e sons também podem ser digitalizados, tudo se torna digital. A informação digitalizada é processada de forma automática, extremamente veloz, com alto grau de precisão e em larga escala. Nenhum outro processo conhecido pela humanidade reunia essas quatro qualidades, o que traz a revolução informacional com a emergência da cibercultura (LÉVY, 1999, p. 33-52).

Nesse cenário emergente, as informações tornam-se facilmente transformáveis em elementos quantificáveis – números, imagens e sons digitais –, potencializando significativamente a coleta, o armazenamento, o processamento e a difusão massiva dessas informações. Nesse contexto, surge uma nova forma de organização social, denominada por Manuel Castells de Sociedade em Rede ou por Pierre Lévy de Sociedade da Informação, cujos aspectos e implicações serão aprofundados nos tópicos seguintes deste estudo.

2.3.1 A Sociedade da Informação

Pierre Lévy, em sua obra Cibercultura (1999), analisa como as tecnologias digitais da informática transformam a comunicação e a expressão cultural, enfatizando o impacto social e cultural da tecnologia digital, observando como ela molda novas formas de sociedade e interação. Assim, define este movimento não somente como uma revolução tecnológica, mas um movimento social capaz de alterar toda dinâmica da sociedade.

O início deste movimento possui relação direta com o surgimento das redes de computadores, conhecida como internet, na década de 1970, onde essas redes de computadores começaram a se conectar entre si, juntando umas às outras enquanto o número de pessoas e de computadores conectados nelas começou a crescer de forma exponencial. Nesse contexto emerge o ciberespaço, com “novo espaço de comunicação, de sociabilidade, de organização e de transação, mas também novo mercado da informação e do conhecimento” (LÉVY, 1999, p. 32).

Dentro deste contexto, observa-se que não se trata meramente de uma evolução tecnológica, mas uma nova forma de organização social que impacta as relações sociais, negociais e econômicas, denominada como cibercultura.

Levy (1999, p. 125-129) destaca que a cibercultura pode ser entendida como um “movimento social”, onde seu grupo líder é a juventude metropolitana escolarizada. O programa da cibercultura é liderado por suas palavras de ordem: interconexão, comunidades virtuais e inteligência coletiva, ao qual estabelece que são os três princípios que orientaram o surgimento do ciberespaço.

A interconexão possui conexão intima ao conceito de internet, sendo o horizonte técnico da cibercultura estabelecendo a comunicação universal, estabelecendo que cada computador, aparelho ou máquina, da torradeira ao automóvel, deve estar possuir conexão com a internet. Lévy (1999, p. 129) afirma que este é o imperativo categórico da cibercultura, apontando para uma civilização interconectada de forma global, sem fronteiras com presença generalizada.

Por trás da interconexão há a técnica, a qual foi objeto de estudo por Pierre Lévy (1993, p. 43), onde analisou o papel das tecnologias da informação na constituição das culturas, e também destacou como fenômeno indissociável da cibercultura e da interconexão o surgimento dos computadores pessoais, destacando o papel do Silicon Valley neste movimento, formado por uma grande comunidade de

jovens californianos reunidos ao redor da universidade de Stanford com extrema abundância e variedade de tecnologias da informação. A técnica estudada durante toda obra passa desde o primeiro computador, o Eniac dos anos 40 exclusivamente para usos de programação e processamento até a formação do computador pessoal como hoje é conhecido, associado ao uso de uma tela, mouse e teclado que simbolizam a própria máquina (Lévy, 1993, p. 102). Alerta o autor que não há identidade estável na informática, pois essa é imprevisível, o que pode transformar radicalmente seu significado e uso, a definição e conceito neste caso sempre é o mais recente, a última evolução tecnológica (LÉVY, 1993, p. 103).

Na obra examinada extrai-se que o autor possui uma condicionante futura em suas afirmações ao falar da internet e da conexão generalizada de aparelhos como “se este programa se concretizar”, pois foi redigida no início dessa tecnologia. Atualmente não é exagerado dizer que até mesmo as torradeiras estão conectadas nesse tecido universal. Ao tratar da definição e conceito da tecnologia da informação ou da informática, importante é sua contribuição de que esse deve ser analisado sempre de forma contemporânea, analisando as últimas invenções tecnológicas, pois essas podem radicalmente transformar seu uso e significado, é neste contexto que essa pesquisa se recortar no contexto da inteligência artificial, a qual será objeto aprofundado de estudo nos próximos capítulos, pois essa se revela como a última grande inovação da tecnologia.

A segunda palavra de ordem da cibercultura são as comunidades virtuais. Essa deriva intrinsicamente do primeiro princípio da interconexão pois funda-se neste. Para LÉVY (1999, p. 130):

> Uma comunidade virtual é construída sobre as afinidades de interesses, de conhecimentos, sobre projetos mútuos, em um processo de cooperação ou de troca, tudo isso independentemente das proximidades geográficas e das filiações institucionais.
> Para aqueles que não as praticaram, esclarecemos que, longe de serem frias, as relações on-line não excluem as emoções fortes. Além disso, nem a responsabilidade nem a opinião pública e seu julgamento desaparecem no ciberespaço. Enfim, é raro que a comunicação por meio de redes de computadores substitua pura e simplesmente os encontros físicos: na maior parte do tempo, é um complemento ou um adicional.

Para compreender as comunidades virtuais de forma aprofundada, deve-se examinar o contido em sua obra O que é virtual? de 1996 dedicada a explicar tal fenômeno da virtualização. Neste estudo, o autor elucida a falsa premissa antagônica

entre o real e o virtual. Explica a virtualização não como uma maneira de ser, existir ou não existir, mas como uma dinâmica social com o deslocamento de seu centro de gravidade ontológico do físico para o imaterial. Cita como exemplo a virtualização de uma empresa, onde há a substituição da presença física de seus empregados pela participação numa rede de comunicação eletrônica, o centro de gravidade desta empresa não é mais um conjunto de departamentos, postos de trabalhos e livros de pontos, mas um processo virtual de coordenação e comunicação que redistribui as coordenadas da coletividade de trabalho e das exigências impostas a cada um de seus membros conectados. Isso demonstra uma das principais modalidades da virtualização, o desprendimento do aqui e do agora, o virtual com muita frequência não está presente, porém, não quer dizer que não exista (LÉVY, 1996, p. 18-19).

A virtualização é uma dinâmica multifacetada. Há a virtualização do corpo, em razão das técnicas de comunicação e da telepresença, estamos ao mesmo tempo aqui e no virtual, como se houvesse dois corpos, o físico e o virtual. Por sua vez, a virtualização do texto é o fenômeno explicado pela digitalização, ao qual traz uma nova dinâmica ao texto, ao invés de estático e imutável ao ser produzido em papel físico, este se torna fluido, reconfigurável, não linear, circulando na rede mundial de computadores onde cada participante é um autor e um editor potencial. No campo da econômica, essa dinâmica traz uma economia da desterritorialização, a moeda, base das finanças, tornou-se virtual, acarretando uma desterritorialização da força de trabalho, da transação comercial e do consumo, que por muito tempo existiam somente localmente. Isto criou um mercado novo, o mercado *online*, ao qual não conhece distancias geográficas (LÉVY, 1996, p. 15-67).

Melhor exemplo para compreensão das comunidades virtuais na atualidade são as redes sociais, na qual se pode observar todos os fenômenos da virtualização, tornando nossos corpos virtuais em interações em tempo real com comunidades que não estão necessariamente próximas geograficamente, comunicando com pessoas que não estão presentes em nossa realidade física, permitindo o consumo de conteúdos descentralizados e desterritorializados. Por vezes, estas comunidades virtuais levam as pessoas a viver em mundo desconectado com o que realmente vivem em seus dias, o perfeito exemplo do desprendimento daqui e do agora, contudo, não é correto afirmar que esta não exista, pois certamente é real e composta por pessoas reais, e pode em alguns momentos e para algumas pessoas serem mais reais do que aquilo que se é vivido no mundo físico.

O último princípio e palavra de ordem da cibercultura é a inteligência coletiva. Lévy, em sua obra dedicada ao tema A Inteligência Coletiva: por uma antropologia do ciberespaço, a descreve de maneira breve e sintética como “(...) uma inteligência distribuída por toda a parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências” (Lévy, 1994, p. 29). A inteligência coletiva também pode ser mais facilmente compreendida como “um novo meio de comunicação, de pensamento e de trabalho para as sociedades humanas” (LÉVY, 1994, p. 11).

Nesse caso, inteligência não pode ser interpretada como um conceito exclusivamente cognitivo como a inteligência humana, mas compreendida como trabalhar em comum acordo. Com a criação da internet e dos computadores pessoais e a emergência do ciberespaço, um meio de comunicação em tempo real e permanente, que permite o compartilhamento do saber e moldando uma nova forma de pensamento e possibilita formas, cargos e modalidades de trabalho nunca imaginadas. Com isso, a mobilização efetiva das competências ganha destaque no ciberespaço, pois para mobilizar competências é necessário identificá-las, sendo essa tarefa mais fácil com toda a informação e todo compartilhamento de saber disponível neste novo espaço antropológico, permitindo a valorização técnica da sociedade. (LÉVY, 1994, p. 31).

Assim, a inteligência coletiva proposta por Lévy se concretiza hoje em inúmeras experiências de colaboração e construção conjunta de conhecimento, potencializadas pelas redes digitais e pela permanente troca de informações em tempo real. Plataformas colaborativas através de conteúdos em texto, imagens, áudio, vídeo, e *streamings*, comunidades virtuais de aprendizagem e espaços de inovação aberta mostram de forma prática como o compartilhamento descentralizado de saberes pode gerar soluções ágeis e criativas para questões complexas, bem como novas modalidades de trabalho e organização social. Nessa perspectiva, o ciberespaço não apenas valoriza as habilidades individuais ao congregá-las em um ambiente de interação contínua, mas também incentiva a emergência de novas competências, solidificando o princípio de que, ao trabalharem em comum acordo por meio das tecnologias digitais, grupos e comunidades tornam-se capazes de realizar empreendimentos e projetos que superam o que cada indivíduo poderia alcançar isoladamente.

De forma mais suscinta, a interconexão na cibercultura aponta para uma

civilização totalmente interconectada por meio da internet, sem fronteiras. As comunidades virtuais, por sua vez, exploram novas formas expressar a opinião pública, que se encontra intimamente ligada ao conceito da democracia moderna. A inteligência coletiva, por sua vez, é o fim último da cibercultura, colocando em sinergia e a disponibilidade os saberes, as imaginações e as energias espirituais daqueles conectados. (LEVY, 1999, p. 124-131).

Compreendidos esses conceitos-chave da cibercultura — interconexão, comunidades virtuais e inteligência coletiva — torna-se evidente que o ciberespaço não é mero instrumento tecnológico, mas um novo espaço de sociabilidade e produção intelectual, capaz de moldar a forma como indivíduos e coletividades se organizam, aprendem, compartilham conhecimento e estruturam relações de poder. Na prática, a noção de cibercultura transcende a simples adoção de dispositivos digitais, pois envolve a articulação contínua entre práticas sociais, culturais e econômicas que emergem em cada inovação e redefinem nossa experiência do tempo, do espaço e do próprio sentido de comunidade. Desse modo, ao expandir as possibilidades de interação e criação conjunta, o movimento social descrito por Lévy reforça a relevância de examinar como essas novas modalidades de colaboração e comunicação influenciam fenômenos mais amplos — tais como a liberdade, a privacidade, a criatividade e as relações empresariais —, preparando terreno para a análise das transformações atuais suscitadas pela Sociedade Informacional e pela intensa mobilização de dados pessoais em escala global.

2.3.2 A Sociedade em Rede

Em paralelo à concepção de Sociedade da Informação proposta por Pierre Lévy, estão as lições trazidas por Manuel Castells e o seu conceito de Sociedade em Rede, sendo que ambas tratam sobre a sociedade informacional. Este conceito será explorado introduzindo as bases conceituais que sustentam o pensamento de Castells, identificando pontos de contato ou divergência em relação à perspectiva de Lévy.

O primeiro ponto de contato é inicialmente observado pelo momento e o contexto histórico do surgimento da sociedade informacional. Para Castells (1942, p. 64-70), a sociedade passou por uma transformação histórica através de um novo

paradigma tecnológico em 1970, organizado com base na tecnologia da informação, com o advento da internet nos Estados Unidos da América. Esta revolução da tecnologia da informação levou a um processo de reestruturação do sistema capitalista a partir da década de 80, concretizando um novo estilo de produção, comunicação, gerenciamento e vida, o que nos levou ao pós-industrialismo ou informacionalismo através da sociedade informacional.

Essa nova era informacional afetou a forma de relacionamento do ser com a sociedade, as novas tecnologias da informação integraram o mundo em redes globais de instrumentalidade, a comunicação através dos computadores e da internet criou diversas “comunidades virtuais”. Isso levou ao surgimento de um novo paradigma social entre a identidade do ser frente as redes, com uma defesa da personalidade, identidade e cultura do ser contra a lógica dos aparatos e mercados impostos pelas comunidades virtuais e pela memória coletiva, criando uma dicotomia entre a Rede e o Ser, gerando uma crise da identidade do ser (CASTELLS, 1942, p. 77-78).

A teoria apresentada nos demonstra a existência de uma verdadeira revolução a partir do surgimento da internet, com impactos diretos da tecnologia da informação na estrutura social e econômica no mundo a partir da década de 80, comparada com uma mudança social com impactos semelhantes ao da Revolução Industrial.

Essa revolução ainda não está finalizada, a história da revolução da tecnologia da informação acontece de forma intensa, sendo que qualquer relato histórico se torna obsoleto em uma velocidade inimaginável, entre o momento de escrita deste artigo e o momento de sua publicação, provavelmente, novas aplicações para a tecnologia e inteligência artificial terão sido lançadas e impactarão a nossa sociedade. Isso é explicado através da Lei de Moore, onde se verifica uma tendencia de aumento de desempenho dos microchips e processadores de forma exponencial, ou seja, quanto mais avanço tecnológico se tem, mais rápido é o avanço tecnológico futuro (CASTELLS, 1942, p. 95).

Transitando do aspecto tecnológico para o aspecto econômico, a sociedade informacional levou ao surgimento de uma nova economia em escala global, que tem como características principais ser informacional, global e em rede. Sua característica informacional refere-se ao fato de que a produtividade e a competitividade agora não são relacionadas a produção de bens de consumo, mas sim quanto a capacidade de gerar, processar e aplicar de forma inteligente as informações, o que é traduzida pela célebre frase “Dados são o novo petróleo” de Clive Humby. A característica global se

deve ao fato de que essas estruturas informacionais estão organizadas em escalas globais, devida a facilidade destes dados e informações transitarem de forma rápida e irrestrita por diferentes países e regiões. E essa economia é em rede pois a concorrência e interação entre as empresas acontece em uma rede global de interação (CASTELLS, 1942, p. 135).

Essa nova economia criada pela sociedade informacional impacta diretamente o tema e o problema enfrentado neste artigo, pois as tecnologias da informação, como a inteligência artificial, são operadas em um novo contexto econômico e social, portanto, as ferramentas de controle e solução utilizadas pela antiga sociedade industrial já não são mais eficazes neste novo contexto.

Um grande exemplo disto é a internacionalização da produção, com o surgimento de grupos empresariais multinacionais, transnacionais e redes internacionais de produção acelerados a partir da década de 90. Isso não significa que a sociedade informacional favorece diretamente essas empresas, até porque verificamos uma grande tendencia das empresas de pequeno e médio porte se destacarem nesta nova economia por serem bem adaptadas ao sistema produtivo flexível da economia informal, porém, observamos que essas multinacionais permanecem no centro de estrutura do poder econômica nesta nova economia e passam a ser mais poderosas diante da globalização e do enfraquecimento dos Estados, aproveitando-se das redes internacionais (CASTELLS, 1942, p. 217-221).

Como conclusão de sua teoria, Manuel Castells (1942, p. 553) expõe que redes constituem a nova forma de organização da nossa sociedade e uma forma de organização capitalista, onde as empresas organizam-se em redes internacionais e os capitalistas não são mais os donos dos meios de produção, em verdade, não existe uma classe capitalista global, mas uma rede integrada de capital global, cujos interesses movimentam as economias e influenciam as sociedades, essa rede não é comandada por indivíduos, mas por uma entidade capitalista despersonificada, distribuídos de forma aleatórias e comandados um fluxo financeiro global constante. Em seu aspecto social, a sociedade ingressou em um novo estágio em que superou o domínio da Natureza e até mesmo em partes a necessidade do trabalho, alcançando o nível de conhecimento e organização social que lhe permite viver em um mundo predominantemente social. Porém, isso não é necessariamente animador, porque os humanos agora tem que lidar finalmente sozinhos com a natureza humana, assim, novos desafios serão impostos a sociedade e pode ser o início de uma nova batalha

social histórica.

A nova era informacional, como exposto por Castells, implicará em novos desafios para a sociedade, e a solução para esses problemas será desafiadora pois não poderemos recorrer aos velhos institutos jurídicos e sociais conhecidos, pois esses serão ineficazes frente a uma nova economia e estrutura social existente, teremos que explorar opções que coadunem com a nova realidade informacional, global e em rede.

2.4 LIMITES ENTRE A SOCIEDADE INFORMACIONAL E A SOCIEDADE DA VIGILÂNCIA

Após a análise das perspectivas de Pierre Lévy e Manuel Castells sobre a Sociedade da Informação e a Sociedade em Rede, importa agora refletir sobre as como essa revolução informacional impacta o conceito de liberdade e privacidade ao possibilita a modernização de mecanismos de controle e vigilância, contribuindo para a expansão e consolidação da Sociedade da Vigilância.

Pierre Lévy privilegia em sua obra os impactos positivos trazidos pela cibercultura, trata também da crítica da dominação, destacando críticas de teóricos como Arthur Kroker e Michael A. Weinstein, que apontam para a possibilidade do domínio de uma nova classe virtual, com o domínio daqueles que tem o controle da indústria dos sonhos (cinema, televisão, videogames) junto com os que dominam os programas, a eletrônica e a telecomunicações. Esses teóricos apontam as mesmas tecnologias que proporcionam liberdade de expressão e acesso também podem criar tipos de controle e vigilância, limitando, assim, a liberdade (LEVY, 1999, p. 222).

A relação entre privacidade e vigilância e o impacto da sociedade informacional é abordada por Stefano Rodotà (2008. p. 126) que defende que na sociedade informacional, onde a informação é coletada, tratada e compartilhada em larga escala e com rapidez sem precedentes, essa também se torna alimento indispensável para funcionamento deste sistema, por um preço demasiadamente caro à privacidade e a tutela plena das informações pessoais, sendo esta a fronteira entre a sociedade da informação e a sociedade da vigilância.

Para a compreensão do tema da vigilância, deve ser entendido que este é objeto de estudos por cientistas sociais antes da sociedade informacional. Michael Foucault (2014, p. 172-197demonstrou como a vigilância se tornou um operador

econômico decisivo no capitalismo, classificando-a como uma ferramenta do poder disciplinar. O autor sugere como mecanismos de vigilância são utilizados como forma de controle dos indivíduos em seus comportamentos e posicionamentos sociais. A utilização da vigilância como controle permite que não seja necessário recorrer à força para “obrigar o condenado ao bom comportamento, o louco à calma, o operário ao trabalho, o escolar à aplicação, o doente à observância das receitas”, basta os vigiar e punir.

A teoria do panóptico de Foucault oferece uma analogia poderosa para compreender como a vigilância pode ser utilizada para exercer controle social. O Panóptico de Bentham é uma figura arquitetural imaginada para uma prisão, composta um espaço com uma torre em seu centro e uma construção em anel ao seu redor. Na torre ao centro, ficaria localizado o vigia, com janelas para todos os lados, e na construção periférica os loucos, doentes, condenados, operários ou escolares, todos aqueles submetidos ao poder disciplinar. Essa construção permitiria que todos os espaços de controle poderiam ser vistos sem parar e reconhecidos imediatamente. O efeito mais importante disso é induzir no individuo controlado um estado consciente e permanente de visibilidade, tornando a vigilância permanente em seus efeitos mesmo que descontinuada em sua ação. Para seu pleno funcionamento, é necessária a observação de dois princípios, de que esse poder deve ser visível, na medida de que todo sujeito controlado deve ter a desconfiança de que sempre pode ser observado, e inverificável, para nunca saber se realmente está sendo observado (FOUCAULT, 2014, p. 194-195).

A teoria do panóptico e a forma de sua apresentação através do Panóptico de Bentham devem ser lidas como uma alegoria sobre todo o poder disciplinar e de que se estendem por toda a sociedade. Mais do que uma arquitetura concreta, o panóptico simboliza a maneira como o controle e a vigilância são internalizados pelos indivíduos, criando uma sensação permanente de observação e conformidade. Tal alegoria traduz a forma como regras, padrões de comportamento e hierarquias de poder se articulam para moldar a conduta de cada pessoa. Nesse sentido, o panóptico funciona menos como um edifício e mais como um paradigma que esclarece os mecanismos invisíveis, mas efetivos, de controle social.

Com a emergência da sociedade informacional, o modelo panóptico, enquanto alegoria do poder disciplinar e da normalização, adquire novos contornos. Se no arranjo originário a vigilância se materializava em um espaço arquitetônico — o

Panóptico de Bentham —, na sociedade informacional ela se expande para redes digitais, bancos de dados e algoritmos capazes de rastrear e analisar comportamentos em tempo real.

Rodotà (2008, p. 28) ao tratar da sociedade informacional e o enorme aumento das quantidades de informações pessoais coletadas, tratadas e compartilhadas neste novo meio de organização social, destaca que essa possui também como objetivo a gestão de programas de intervenção social, por instituições públicas e privadas, bem como o controle do comportamento político, objetivos esses que estão alinhados ao ideal do poder disciplinar e do poder de normalização tratado por Michael Foucault, que ao tratar sobre o poder político, estabelece que para que esse poder seja exercido "(...) deve adquirir o instrumento para uma vigilância permanente, exaustiva, onipresente, capaz de tornar tudo visível, mas com a condição de se tornar ela mesma invisível" (FOUCAULT, 2014, p. 207).

O aparelho disciplinar perfeito para FOUCAULT (2014, p. 170) seria:

> O aparelho disciplinar perfeito capacitaria um único olhar que tudo vê permanentemente. Um ponto central seria ao mesmo tempo fonte de luz que iluminasse todas as coisas, e lugar de convergência para tudo o que deve ser sabido: olho perfeito a que nada escapa e centro em direção ao qual todos os olhares convergem.

O "aparelho disciplinar perfeito" descrito por Foucault revela-se amplificado pela sociedade informacional. Com o advento da internet, dos computadores pessoais e do ciberespaço, a arquitetura panóptica, antes restrita à vigilância física dos encarcerados ou à disciplina em instituições físicas, estende-se, na sociedade informacional, aos circuitos digitais que interligam cada indivíduo. Os mecanismos que, a um só tempo, podem oferecer acesso à informação e inclusão digital, também engendram espaços de controle e normalização, fazendo com que o projeto foucaultiano do poder disciplinar encontre na tecnologia o palco perfeito para sua difusão e sofisticação na era da informação.

A coleta de dados pessoais, o monitoramento *online* e a presença de algoritmos, embora expandam as oportunidades de comunicação e acesso à informação, também impõem barreiras invisíveis à liberdade de pensamento e ação, restringindo a autonomia individual em um grau anteriormente inimaginável, trazendo essa dualidade intrínseca da cibercultura entre a expansão e a restrição da liberdade.

De acordo com Bauman (1999), os sistemas de controle virtuais, como o Big

Data, facilitam uma melhor circulação na internet. Ele sugere que, à medida que somos monitorados constantemente pelos algoritmos poderosos do Facebook e mecanismos semelhantes, nossa liberdade de movimentação no ciberespaço aumenta. Esses controles, ao invés de restringir, possibilitam maior mobilidade. Assim, os logins sociais e o conteúdo acessível na internet nos permitem acessar uma vasta gama de informações e interagir com diversas pessoas em diferentes locais. Bauman argumenta que, quanto mais informações um banco de dados possui sobre um indivíduo, maior é sua liberdade de navegação. Portanto, a liberdade de navegação está diretamente relacionada ao nível de vigilância e controle exercido, funcionando como uma oferta de "liberdade de navegação" através da maximização da vigilância, demonstrando a dualidade que a cibercultura implica a expansão e a restrição da liberdade.

Essa busca por informação é atrelada a própria vigilância. Foucault (2014, p. 178) já abordava a necessidade de quantificação e qualificação pelo poder disciplinar, ao qual hierarquizam os indivíduos em bons e maus sob a ótica do controle. Em perspectiva da sociedade de vigilância, Rodotà (2008, p. 111) aborda em suas obras a sociedade da classificação. Na sociedade informacional, a maioria das relações entre usuários e fornecedores são administradas por programas que geram informações sobre essas relações e que dão direito pela relação contratual que esses fornecedores adquiram automaticamente uma série de informações sobre seus usuários como sua identificação, horário e local de utilização dos serviços, suas preferências de consumo e outras que permitem identificar e qualificar esses indivíduos ao traçar perfis dos usuários. Assim, para melhor experiência do usuário no uso dessas tecnologias é incentivado que esse cada vez mais compartilhe suas informações na rede.

Em obra dedicada ao tema da Vigilância Líquida, o pensador Zygmunt Bauman dialoga com o sociólogo David Lyon, construindo uma teoria sobre a vigilância líquida como pós-pan-óptico. Ao ser questionado sobre o fim do pan-óptico, Bauman (2013, p. 58) argumenta que esse está vivo e mais forte do que nunca, eletronicamente reforçado, de forma tão poderosa que Bentham ou Foucault sequer poderiam imaginar. Lyon (2013, p. 68) destaca que o pós-pan-óptico não possui somente o objetivo de repressão do poder disciplinar como anteriormente, apontando seu uso atrelado à vigilância do consumidor. Nesse caso, as técnicas de identificação e classificação são combinadas com técnicas de marketing, classificando os indivíduos

em grupos de acordo com seu comportamento de consumo e direcionando estratégias de marketing segmentadas, processo pelo qual os consumidores e seus desejos são manipulados sem a sua ciência.

Os pensamentos de Bauman e Lyon são amparados em estudos desenvolvidos por Oscar Gandy em seu *The Panoptic Sort: A Political Economy of Personal Information* que traz reflexões sobre a "economia política das informações pessoais". Gandy enfatiza que a vigilância não é mera coleta passiva de dados, mas parte de uma estratégia ativa e contínua de moldar e prever comportamentos. Esse cenário, na visão dele, representa um avanço de mecanismos de "controle digital" — ou "digital enclosure" — em que não apenas o espaço online é cercado por sistemas de rastreamento, mas a própria liberdade de ação e decisão dos usuários é gradualmente condicionada pela arquitetura de vigilância e análise de dados. A coleta intensa de informações permite os vigilantes saber e até mesmo antecipar quais informações cada um de nós pode precisar, mesmo que ainda não tenhamos consciência disso, permitindo a estes influenciar ações individuais, adequando ofertas e mensagens de acordo com o perfil de cada usuário. Assim, o usuário participa ativamente de um sistema no qual suas respostas alimentam a próxima onda de conteúdo ou recomendações (GANDY, 1993, p. 33-34).

Essa é a terceira fase da revolução da sociedade consumista. A primeira revolução foi marcada pela transição entre atender necessidades dos consumidores para a criação da própria necessidade através da tentação, sedução e estímulo do desejo de consumo. Nessa nova onda, os fornecedores podem dirigir ofertas a pessoas pré-dispostas a aceitá-las com entusiasmo. Isso é refletido pelo advento da Amazon, Facebook e Google que demonstram o atual estado da arte. Como exemplo, ao acessar o site da Amazon, nessa nova fase da sociedade consumista, o indivíduo é recepcionado por produtos especialmente selecionado para ele, de acordo com seu perfil de consumo traçado por sua própria cooperação com os servidores da Amazon, gerando uma alta probabilidade de tentação. Assim é feito o monitoramento, classificação e manipulação dos consumidores através do consumismo e das novas tecnologias (Bauman; Lyon, 2013, p. 113-117). A vigilância a favor da sociedade consumista é diretamente relacionada a crítica da dominação de Arthur Kroker e Michael A. Weinstein exposta por Pierre Lévy (1999, p. 222) configurada pela dominação de uma nova classe virtual, os detentores das tecnologias.

Ao observar as múltiplas perspectivas que envolvem a vigilância na sociedade

informacional, verifica-se que os mesmos dispositivos tecnológicos que expandem a liberdade de expressão e o acesso à informação também impulsionam o controle social e a restrição de direitos. A teoria foucaultiana do panóptico, reelaborada pelos debates sobre a sociedade da informação e sociedade em rede de Lévy e Castells e analisada por Rodotá ao tratar do crescimento da chamada “sociedade da vigilância”, evidencia a tendência de se estruturar práticas contínuas de observação e classificação dos indivíduos. Consequentemente, a privacidade e a autonomia de cada indivíduo passam a ser permeados por uma sensação permanente de ameaça.

Tal cenário, demanda reflexões éticas e jurídicas acerca da liberdade, da privacidade e da autonomia. A busca de um equilíbrio efetivo entre os benefícios da revolução informacional e o respeito aos direitos individuais revela-se fundamental para manter a dignidade humana diante dessas tecnologias que, em seu avanço contínuo, arriscam o espaço de autodeterminação e liberdade que constitui o centro de uma sociedade democrática.

2.5 DA PRIVACIDADE À PROTEÇÃO DE DADOS

Neste sentido, se observa que a própria concepção de privacidade se transfigurou ao longo das últimas décadas para se fixar em torno da informação e, em especial, dos dados pessoais, revelando a transição de uma dimensão meramente espacial (o “direito de estar só”) para uma perspectiva fortemente centrada nos fluxos informacionais. Sob essa ótica, o debate passa a englobar a chamada *informational privacy* dos Estados Unidos (com a preocupação voltada ao acesso a dados em órgãos públicos e às atividades de proteção de crédito), bem como a autodeterminação informativa reconhecida pelo Tribunal Constitucional alemão e as regulamentações da União Europeia, especialmente a partir da Diretiva 95/46/CE. Em todos esses marcos, persiste uma referência objetiva à disciplina dos dados pessoais como herdeira e ao mesmo tempo atualizadora do tradicional direito à privacidade, impondo características e exigências próprias para lidar com os desafios impostos pela sociedade informacional (Doneda, 2020, p. 164). Essa visão também é compartilhada por Newton de Lucca e Renata Mota Maciel Madeira Dezem (2018, p. 258): “Como se vê, o reconhecimento do direito à privacidade pode ser considerando o primeiro desdobramento do que na sequência passou-se a denominar “direito à

proteção dos dados pessoais”(...)”.

Com isso, evidencia-se que o tratamento dos dados pessoais assumiu um lugar central na configuração da sociedade contemporânea, seja na vertente de seu potencial emancipatório e inclusivo, seja na lógica de controle e manipulação de preferências individuais. As tensões entre liberdade, privacidade e vigilância, analisadas sob a ótica da evolução histórica desses conceitos e do advento da sociedade informacional, realçam a urgência de se estabelecer um arcabouço jurídico capaz de garantir transparência, segurança e respeito aos direitos fundamentais. Esse cenário introduz o tema da proteção de dados como um eixo fundamental para compreender as novas dinâmicas tecnológicas no contexto da inteligência artificial. No próximo capítulo, portanto, serão abordados os instrumentos jurídicos, os marcos legais e a aplicação prática da tecnologia na tutela dos dados pessoais, ampliando a discussão sobre como conciliar avanços tecnológicos e o resguardo efetivo da liberdade e autonomia de cada cidadão.

Ao abordar a consolidação dos direitos à privacidade e à proteção de dados como aspectos fundamentais da personalidade, observa-se uma evolução legislativa em sucessivas gerações, refletindo as demandas históricas e tecnológicas de cada período. Na década de 1970, por exemplo, o Land de Hesse (Alemanha) editou a primeira lei local de proteção de dados, ainda voltada à coleta estatal de informações. Nessa mesma época, nos Estados Unidos, o Fair Credit Reporting Act (FCRA) de 1970 regulou cadastros de crédito, indicando uma preocupação inicial quanto ao uso empresarial de dados pessoais. Em 1973, a Suécia promulgou sua Data Legen nº 289, primeira lei nacional no tema, criando o Dataispektionen para fiscalizar o manuseio de informações (DONEDA, 2020, p. 166-167).

No percurso seguinte, surge o Privacy Act (1974) norte-americano, estabelecendo normas de coleta e tratamento de dados em agências federais. Ainda na Alemanha, a Bundesdatenschutzgesetz, de 1977, consolidou em âmbito federal os princípios antes circunscritos às legislações locais. Com a lei francesa Informatique et Libertés, de 1978, surge a Comissão Nacional de Informática e Liberdades (CNIL), marcando a segunda geração das leis de proteção de dados e reforçando a ideia de uma liberdade negativa do indivíduo ao decidir sobre o uso de suas informações (DONEDA, 2020, p. 167-168).

Em 1981, a Convenção nº 108 do Conselho da Europa torna-se o primeiro instrumento internacional juridicamente vinculante, estendendo a outros países a

obrigação de tratar dados pessoais em conformidade com direitos fundamentais. Já a década de 1990 testemunha o fortalecimento do debate europeu, graças à Diretiva 95/46/CE, que uniformizou a proteção de dados nos Estados-Membros da União Europeia, assegurando ao indivíduo o controle sobre suas informações e impulsionando a livre circulação de dados no bloco (Lucca; Dezem, 2018). Em 2000, a Carta de Direitos Fundamentais da União Europeia reconheceu, de forma autônoma, a proteção de dados como um direito fundamental, aproximando-se da noção de autodeterminação informativa (Rodotà, 2008). Esse reconhecimento institucional, aliado à expansão das comunicações eletrônicas reguladas pela e-Privacy Directive (2002), ajudou a preparar o terreno para regulações mais abrangentes (LUCCA; DEZEM, 2018)

Nesse mesmo período, a relação entre a Europa e os Estados Unidos em matéria de transferência de dados era marcada por um profundo contraste: enquanto a tradição europeia sedimentava normas rigorosas e mantinha autoridades de fiscalização, o modelo norte-americano prevaleceu um pensamento liberal fundado na autorregulação setorial e na livre transmissão de informações, muitas vezes em detrimento do direito à privacidade, o que resultou em leis pontuais – como a HIPAA (1996), para dados de saúde. Foi exatamente nessa contraposição que surgiu o Acordo Safe Harbor, em 2000, visando compatibilizar as exigências mais rígidas da União Europeia com a abordagem liberal dos EUA – mas acabou sendo considerado insuficiente pela própria União Europeia, que o invalidou em 2015 e posteriormente o substituiu pelo Privacy Shield, a fim de equilibrar transferências de dados pessoais entre Europa e Estados Unidos (LUCCA; DEZEM, 2018).

Apesar dos avanços promovidos pela legislação europeia no tema, com destaque à Diretiva 95/46/CE e a Carta de Direitos Fundamentais da União Europeia, com o transcurso do tempo e o avanço tecnológico com velocidade inimaginável, essas já não apresentavam respostas adequadas para os desafios propostos pelo tema. Assim, após anos de discussões sobre o assunto, a consolidação da matéria de proteção de dados na União Europeia se efetivou com o Regulamento Geral de Proteção de Dados (RGPD) ou *General Data Protection Regulation (GDPR)* aprovado em 2016. O RGPD unificou, modernizou e trouxe regras ainda mais abrangentes sobre proteção de dados, com dispositivos aptos a garantir a proteção aos usuários de internet, assegurando aos titulares de dados maior controle sobre seus dados, impondo obrigações rigorosas às instituições controladoras e operadoras de dados.

Essa norma europeia também fortaleceu o papel das autoridades nacionais de proteção de dados (LUCCA; DEZEM, 2018).

Como consequência das ações regulatórias da UE, surge o "efeito Bruxelas", provavelmente não intencional, no qual a União Europeia exerce uma influência significativa na competição tecnológica global, determinando as regras que as empresas seguem internacionalmente (Floridi *et al.,* 2024, p. 5), assim, a legislação europeia se tornou referência para diversos países e blocos econômicos, que passaram a adotar ou adaptar legislações inspiradas nesse modelo.

No Brasil, o tema da proteção de dados era respaldado somente em dispositivos constitucionais voltados à tutela da privacidade e à garantia da informação. A Constituição Federal contemplava o problema da informação inicialmente por meio das garantias à liberdade de expressão e ao direito à informação, ao mesmo tempo em que considerava invioláveis a vida privada e a intimidade (art. 5º, X). Além disso, previa a ação de habeas data (art. 5º, LXXII), assegurando o acesso e a retificação de dados pessoais, e proibia a interceptação de comunicações sem autorização legal, abrangendo aspectos telefônicos, telegráficos e de dados (art. 5º, XII). Em paralelo, protegia determinados aspectos específicos da privacidade, como a inviolabilidade de domicílio (art. 5º, XI) e de correspondência (art. 5º, XII). Caso seja compreendido que proteção de dados é um desdobramento direto da privacidade, poderia se concluir que essa seria protegida por nossa Constituição Federal como um direito fundamental (DONEDA, 2020, p. 266).

Mais precisamente, o livre desenvolvimento da personalidade seria o direito fundamental que melhor fundamentaria a interpretação de um direito fundamental à proteção de dados. Diante dessas correlações, em 2020, o Supremo Tribunal Federal (STF) reconheceu o direito fundamental à proteção de dados. Em vista da necessidade de tratar o direito à proteção de dados com autonomia, houve a Proposta de Emenda à Constituição (PEC) nº 17/2019, ao qual propôs a inclusão do direito fundamental à proteção de dados pessoais no artigo 5º e artigo 22 da Constituição Federal, sendo esta proposta aprovada em fevereiro de 2022 e publicada a Emenda Constitucional 115/2022. Neste período, houve discussão sobre a real necessidade desta alteração constitucional, porém, essa deve ser compreendida como um importante passo na busca pela efetiva proteção de dados no Brasil, em razão de seu aspecto técnico, pois deve ser reconhecida como um direito autônomo, e seu aspecto cultural, como forma de conscientização nacional sobre a relevância do tema

(QUEIROZ, 2022, p. 53).

No âmbito infraconstitucional, a trajetória de debates sobre proteção de dados pessoais que se iniciou por volta de 2010, culminou na edição da Lei n. 13.709/2018 (Lei Geral de Proteção de Dados – LGPD). Inspirada em parte pelo modelo europeu (especialmente pelo Regulamento Geral de Proteção de Dados – GDPR), a LGPD introduziu no ordenamento brasileiro princípios como finalidade, consentimento, transparência e segurança, além de determinar a criação de uma Autoridade Nacional de Proteção de Dados (ANPD) para fiscalizar a aplicação e promover a observância das regras. Assim, o Brasil passou a contar com arcabouço legal específico para o tratamento de dados pessoais (DONEDA, 2020, p. 310-312).

Durante as discussões para aprovação da LGPD, também surgiu o questionamento sobre a necessidade de tal regulamentação normativa. Em análise sobre o tema, Newton de Lucca e Renata Mota Maciel (2021) abordam a necessidade mercadológica brasileira de uma disciplina normativa sobre proteção de dados para o país, em razão de que empresas e capital estrangeiro encontravam resistência em realizar investimentos no Brasil em razão da ausência de segurança jurídica no tema de proteção de dados, ao qual já possuía sólida robustez em outras jurisdições.

Em conjunto com a Emenda Constitucional 115/2022, que elevou a proteção de dados a direito fundamental autônomo, a LGPD representa um marco na consolidação do tema no país, adequando a legislação interna aos padrões globais e atendendo às novas demandas sociais e econômicas do ambiente digital.

Dessa forma, ao longo das últimas décadas, a proteção de dados pessoais consolidou-se em sucessivas etapas, partindo das primeiras leis nos anos 1970, passando pelos marcos europeus como a Diretiva 95/46/CE e o GDPR, e pelos modelos mais liberais adotados pelos Estados Unidos. Esse processo histórico propiciou um alinhamento progressivo de legislações e princípios, refletindo a importância de resguardar direitos fundamentais em uma sociedade cada vez mais alicerçada na troca de informações. O Brasil, ao editar sua Lei Geral de Proteção de Dados e, posteriormente, elevar a proteção de dados ao status de direito fundamental, integra-se definitivamente a essa evolução global, garantindo maior segurança jurídica e consolidando um arcabouço normativo capaz de enfrentar os desafios impostos pela intensa circulação de dados na era digital.

Diante dessa consolidação normativa, torna-se evidente que a proteção de dados não se esgota em sua dimensão jurídico-formal. A própria razão de sua

emergência, a crescente circulação de informações e a intensificação das práticas de tratamento em escala massiva, indica que o debate jurídico só pode ser plenamente compreendido quando articulado ao processo histórico de evolução tecnológica que marcou as últimas décadas. Assim, compreender a trajetória da proteção de dados exige avançar para a análise da evolução tecnológica, da explosão dos dados e do surgimento da inteligência artificial.

2.6 Evolução Tecnológica, Big Data e Inteligência Artificial

Para compreensão da evolução tecnológica, há que ser compreendido o conceito de tecnologia. Conforme interpretação de Álvaro Vieira Pinto, a tecnologia provém da técnica, sendo que a técnica se manifesta como uma faculdade imanente ao ser humano, a única espécie capaz de criar e reinventar soluções artificiais para os problemas que enfrenta em seu cotidiano. Por sua vez, a tecnologia consiste na ciência da técnica, fruto de uma etapa mais avançada do desenvolvimento humano e que se consolida à medida que a sociedade dispõe dos instrumentos lógicos e materiais necessários à aplicação de conhecimentos científicos. Sob tal perspectiva, não há que se confundir a condição natural do homem de produzir e operar artefatos – própria da técnica – com o uso e evolução sistêmicos desses recursos em benefício de exigências sociais específicas, o que passa a caracterizar a verdadeira noção de tecnologia (PINTO, 2005, p. 49 e 221).

Por essa razão, Álvaro Vieira Pinto (2005, p. 284) repudia a expressão comumente utilizada para os tempos atuais como “era da tecnologia” ou “era tecnológica” pois, segundo sua visão, a humanidade vive sempre em uma era tecnológica que reflete as necessidades sociais de cada momento histórico.

É claro que, nesse processo evolutivo, algumas tecnologias e inovações tiveram mais impacto e velocidade que outras. Nos tempos atuais, essa evolução acontece de forma meteórica, como trata PASTORE:

> Há momentos em que a história corre mais depressa. Estamos em um deles. As mudanças tecnológicas têm sido meteóricas. Na década de 70, uma inovação industrial durava, em média, dois anos. Na década de 80 passou a durar apenas um ano. Depois disso tornava-se obsoleta ou era apropriada por grande parte dos concorrentes. Na década de 90, a duração passou para apenas seis meses. (PASTORE, 2005, p. 1)

Assim, o intervalo temporal mencionado por Pastore, em que as inovações industriais passavam por ciclos cada vez mais curtos entre as décadas de 1970 e 1990, coincide precisamente com a periodização proposta por Manuel Castells e Pierre Lévy. Enquanto Pastore destaca a “velocidade meteórica” das mudanças tecnológicas nesses anos, Castells identifica a década de 1970 como palco da revolução informacional e Lévy assinala o mesmo período como a “data de virada” que potencializou a formação do ciberespaço. Dessa forma, fica evidente que o auge da evolução tecnológica que transformou profundamente os processos sociais e econômicos insere-se no intervalo histórico a que esses autores se referem, reforçando a ideia de que tal aceleração tecnológica esteve no cerne do surgimento da sociedade em rede ou sociedade informacional.

A partir desta revolução digital, os processos de inovação têm se tornado cada vez mais céleres, impulsionando a “destruição criativa”, termo cunhado por Joseph Alois Schumpeter, que descreve que o surgimento de novas tecnologias está vinculado a ondas, que implicam na criação de bens e serviços nunca antes imaginados, de forma revolucionária, disruptiva e extremamente rápida, destruindo as formas tradicionais (SCHUMPETER, p. 94, 1939).

Esse processo de evolução tecnológica e inovação, introduzem produtos e serviços cada vez mais disruptivos e expõe uma disparidade crescente entre a velocidade do desenvolvimento tecnológico e a capacidade regulatória de acompanhar tais mudanças. As normas jurídicas tendem a operar segundo marcos jurídicos construídos sob paradigmas anteriores, resultando, muitas vezes, em cenários de defasagem regulatória.

Nesse contexto surgimento da sociedade informacional, insere-se uma nova mudança de paradigma, a inteligência artificial (IA), uma nova forma de tecnologia que “busca fazer simulações de processos específicos da inteligência humana por meio de recursos computacionais” (Peixoto; Silva, 2019, p. 20), que desponta como uma nova virada de paradigma tecnológico, projetando a evolução tecnológica a níveis de autonomia e complexidade até então inexplorados.

O primeiro estudo sobre inteligência artificial é da década de 1950, onde Alan Turing, um dos grandes precursores da tecnologia da informação e responsável por grande parte dessas invenções, levantou a hipótese de que esses computadores e internet, além de impactar as informações e a comunicação, poderiam também pensar

e serem inteligentes através de seu trabalho *Computer machinery and intelligence* (TURING, 1950).

Neste estudo, Turing dá o passo fundamental de questionar se “as máquinas podem pensar?” e, para abordar essa questão, propõe substituir o debate metafísico por um experimento prático, o “Teste de Turing”, que também é conhecido como o “Jogo da Imitação”. Nesse teste, uma pessoa, conversando via mensagens escritas, tenta distinguir se a outra parte é um ser humano ou um computador. Se o interrogador não consegue perceber a diferença, diz-se que a máquina “passou” no teste. Turing argumenta, portanto, que a questão da consciência ou do “pensamento genuíno” é melhor compreendida por meio de um critério funcional, baseado na capacidade de a máquina exibir respostas adequadas em um diálogo. Com isso, ele inaugura não apenas uma discussão epistemológica sobre o que seria considerado “inteligência”, mas também estabelece uma meta para a pesquisa em IA: desenvolver algoritmos capazes de, por meio de manipulações simbólicas e lógicas, imitar ou simular os processos cognitivos humano (TURING, 1950).

Após o conceito de IA ser inaugurado por Turing, esse evoluiu significativamente. Em 1956, por iniciativa de John McCarthy, vários pesquisadores de temas relacionados a tecnologia da informação organizaram a conferência de Dartmouth, a qual é reconhecida como marco inaugural do campo da inteligência artificial. O evento consistiu em um encontro acadêmico de dois meses que reuniu dez pesquisadores norte-americanos interessados em teoria dos autômatos, redes neurais e estudo da inteligência humana. Nesse contexto, McCarthy, junto com Marvin Minsky, Claude Shannon e Nathaniel Rochester, buscou investigar a hipótese de que todos os aspectos da inteligência poderiam ser descritos com tanta precisão a ponto de serem simulados por máquinas. O encontro em si não resultou em avanços científicos imediatos, mas promoveu um intercâmbio crucial entre pesquisadores que dominaram o campo pelos próximos vinte anos, contribuindo significativamente para a consolidação da inteligência artificial como uma disciplina autônoma dentro das ciências da computação, voltada para reproduzir capacidades humanas como criatividade, autoaperfeiçoamento e uso da linguagem (RUSSEL; NORVIG, 2021, p. 17-18).

Após a conferência de Dartmouth, o campo da inteligência artificial vivenciou um período inicial de grandes expectativas, caracterizado por avanços em programas pioneiros que simulavam habilidades cognitivas para resolução de problemas

específicos, através do *machine learning* ou aprendizado de máquina*,* termo introduzido por Arthur Samuel, um pesquisador da IBM que desenvolveu um programa pioneiro capaz de jogar damas e melhorar seu desempenho com base nas partidas anteriores. Assim, o aprendizado de máquina é uma área que estuda algoritmos capazes de aprender com os dados sem precisar ser explicitamente programados para realizar tarefas específicas. Ou seja, esses algoritmos conseguem "aprender" padrões e generalizar resultados a partir dos exemplos aos quais são expostos. (RUSSEL; NORVIG, 2021, p. 18-23).

Contudo, nesse período não houve um avanço expressivo no sentido de revolucionar as tecnologias da informação existentes, assim, apesar das expectativas iniciais, muitos desses sistemas mostraram-se incapazes de lidar com problemas mais amplos e complexos, evidenciando dificuldades práticas não previstas pelos pesquisadores e causando um período posterior de desencanto, ocasionando o primeiro "inverno da IA", no início da década de 1970. Entre meados dos anos 1970 e o final dos anos 1980, a Inteligência Artificial viveu uma fase de recuperação e expansão baseada em uma nova abordagem: os chamados sistemas especialistas. Ao invés dos métodos anteriores, que tentavam resolver problemas por meio de técnicas gerais e genéricas, surgiram programas que utilizavam conhecimentos específicos extraídos diretamente de especialistas humanos, permitindo resolver problemas complexos em áreas muito especializadas como a medicina. No entanto, apesar desses sucessos pontuais, no final dos anos 1980 ocorreu um novo período de crise, o chamado segundo inverno da IA, devido ao fato de muitas expectativas comerciais exageradas não terem sido atendidas, levando ao fracasso de diversas empresas que apostaram alto em aplicações práticas, mas pouco flexíveis. Este segundo inverno foi caracterizado por ceticismo renovado e redução significativa de financiamento público e privado na área (RUSSEL; NORVIG, 2021, p. 23-28).

De acordo com Goodfellow, Bengio e Courville (2016, p. 12-26), foi somente com o advento das redes neurais e do aprendizado profundo, conhecida como *deep learning*, a qual é uma subárea específica dentro do *machine learning,* que trouxe uma nova estação de verão para a IA, emergindo como um marco histórico que verdadeiramente revolucionou o campo a partir de 1988. O desenvolvimento histórico das redes neurais ocorreu em ondas sucessivas, começando com os primeiros modelos conexionistas das décadas de 1940 a 1960, avançando para as redes neurais dos anos 1980 a 1990, até atingir uma popularidade a partir de 2006. Essa

última fase destacou-se pela capacidade significativamente aprimorada dos algoritmos de aprendizado profundo em reconhecer padrões complexos, graças ao aumento do poder computacional e à disponibilidade de grandes volumes de dados.

Os modelos conexionistas representam a ideia geral de que é possível simular a forma como o cérebro humano funciona utilizando unidades de processamento simples conectadas entre si, conhecidas como neurônios artificiais. Esses modelos enfatizam que o conhecimento está distribuído nas conexões entre essas unidades, permitindo que aprendam através de exemplos e experiências, sem que seja necessário programar explicitamente cada comportamento desejado. A partir deste modelo, surgiu o aprendizado profundo que é uma técnica dentro da inteligência artificial que usa redes neurais artificiais inspiradas no cérebro humano, organizadas em diversas camadas interligadas que funcionam como filtros capazes de reconhecer padrões detalhados em grandes quantidades de informação. Por exemplo, ao identificar uma imagem, as primeiras camadas captam formas simples como linhas e cores, enquanto camadas mais profundas combinam esses elementos para reconhecer objetos ou rostos (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 12-26).

Diferentemente das abordagens anteriores em inteligência artificial, que exigiam a programação manual e detalhada de cada passo ou regra lógica específica para resolver problemas, os modelos conexionistas e o aprendizado profundo caracterizam-se pela capacidade de aprender diretamente dos dados, sem necessidade de regras explícitas pré-definidas. Enquanto sistemas especialistas e métodos tradicionais da IA dependiam do conhecimento detalhado e codificado por especialistas humanos, as redes neurais conexionistas são capazes de aprender padrões complexos de maneira autônoma a partir de exemplos apresentados, adaptando-se continuamente a novas informações.

Esses modelos têm sido fundamentais para o avanço de tecnologias como carros autônomos, assistentes virtuais e sistemas de recomendação, entre outros. Quanto mais utilizada essa tecnologia, mais dados são expostos para as redes neurais que podem melhorar continuamente. A combinação dessas técnicas modernas com o poder computacional crescente e a disponibilidade de grandes volumes de dados - *big data* - criou um ambiente propício para o crescimento exponencial da IA (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 20-22).

Sendo o *big data* resultado do aumento do poder computacional e da

disponibilidade de grandes volumes de dados, esse compartilha suas raízes com o desenvolvimento do *deep learning*, que trouxe a IA para o verão novamente. Todos esses acontecimentos são frutos da arvore da sociedade informacional, a qual já foi compreendida que essa possui suas raízes no advento da internet, do computador e do celular pessoal.

O *big data* foi um termo cunhado por cientistas como da astronomia e genômica que vivenciaram um avanço significativo nos anos 2000, amparados pela enorme quantidade de informações disponíveis para desenvolverem seus estudos. A quantidade de informações cresceu tanto que já não podia ser examinada através de computadores tradicionais, levando ao desenvolvimento de tecnologias de processamento como a MapReduce da Google e a Hadoop do Yahoo, que permitiu a manipulação e análise de informações em escala muito maior. *Big data* não é somente a existência de grande volume de informações, mas “se refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneiras que alterem os mercados, as organizações, a relação entre cidadãos e governos etc” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 204).

O *big data* tem forte relação com previsões, e pode até mesmo causar confusão com a própria área de estudo da inteligência artificial e do *machine learning*, mas deve ser diferenciado destes, pois este conceito trata da aplicação de formulações matemáticas pré-definidas a enorme quantidade de dados para prever possibilidades. Quanto maior o volume de dados alimentados, maior é a precisão dessas previsões. Muitos aspectos do mundo que eram analisados através de lentes humanas, agora poderão ser complementados ou substituídos por sistemas computadorizados. Assim como a internet emancipou a sociedade informacional, o *big data* revolucionou aspectos fundamentais dessa sociedade, ao lhe dar um poder de dimensão quantitativa de forma estratosférica (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 205-300)

Phil Simon em sua obra *Too Big to Ignore: The Business Case for Big Data* (2013) destaca a magnitude impressionante do fenômeno do Big Data, enfatizando sua expansão constante e exponencial. Para dar uma ideia mais concreta dessa dimensão, estima-se que, já em 2009, o volume total de informações existentes na *web* ultrapassava 500 exabytes, o equivalente à capacidade somada de armazenamento de aproximadamente 500 milhões de computadores domésticos,

considerando que cada computador tradicional possua um disco rígido com capacidade de 1 terabyte (TB). No mesmo sentido, Simon aponta que, somente em 2008, o consumo de dados nos Estados Unidos atingiu 3,6 zettabytes, uma escala ainda maior, equivalente à capacidade combinada de armazenamento de cerca de 3,6 bilhões de computadores tradicionais com discos rígidos de 1 TB cada. Além disso, o autor exemplifica com situações do cotidiano digital: apenas em 2011, cerca de 48 horas de vídeo eram carregadas no YouTube a cada minuto, acumulando mais de 1 trilhão de visualizações naquele ano. Esses números ilustram não apenas a imensidão dos dados produzidos diariamente, mas também o ritmo vertiginoso com que continuam crescendo, demonstrando por que o Big Data é considerado um fenômeno de proporções inéditas e continuamente expansivas (SIMON, 2013, p. 74-76).

O *big data* permitiu o desenvolvimento da inteligência artificial generativa (IA Generativa). Segundo Goodfellow, Bengio e Courville (2016, p. 651-720), a IA generativa constitui uma classe específica de algoritmos dentro do aprendizado profundo, capaz não apenas de analisar e classificar informações, mas também de criar conteúdos originais e coerentes, como textos, imagens e áudios, a partir do aprendizado prévio sobre grandes volumes de dados. Tal avanço abriu caminho para novos modelos, entre eles os modelos de linguagem em larga escala (LLMs).

O *Large Language Model (LLM)* ou modelo de linguagem de larga escala é um novo modelo de linguagem dentro do *deep learning* e da IA generativa. Esses modelos são sistemas baseados em aprendizado profundo capazes de compreender, gerar e manipular textos de maneira extremamente avançada e coerente. Impulsionados pela ampla disponibilidade de dados gerados pelo fenômeno do Big Data e pelo significativo aumento do poder computacional, os LLMs popularizaram-se rapidamente e passaram a ocupar um papel central na disseminação e aplicação prática da inteligência artificial em diversas áreas. O lançamento do ChatGPT pela OpenAI, por exemplo, marcou um ponto de inflexão ao demonstrar publicamente, e em larga escala, as potencialidades impressionantes desses sistemas, popularizando e transformando a inteligência artificial em assunto frequente nas mídias, no debate acadêmico e nas discussões sociais cotidianas em escala global (AI NOW, 2023).

Nesse cenário histórico da tecnologia, é possível compreender que há uma combinação sinérgica entre as inovações em inteligência artificial, especialmente com o advento dos modelos conexionistas, do aprendizado profundo, da IA generativa e do modelo de linguagem de larga escala, a massificação dos dados promovida pelo

fenômeno do Big Data e a consolidação da sociedade informacional descrita por Castells e Lévy. Diante desse cenário, é fundamental aprofundar a compreensão sobre a inteligência artificial, explorando suas definições conceituais, características essenciais, classificações e principais aplicações, objeto de estudo detalhado a seguir.

3 DESAFIOS DA INTELIGÊNCIA ARTIFICIAL NA SOCIEDADE INFORMACIONAL

O capítulo anterior apresentou a forma pela qual a sociedade informacional redefiniu a dinâmica entre liberdade, privacidade e proteção de dados, ressaltando o impacto do tratamento de dados pessoais sobre a liberdade, e aprofundou-se tal reflexão ao compreender como a proteção de dados se posicionou como um direito fundamental e ao investigar de que modo as inovações tecnológicas expandem o âmbito dessa discussão. Evidenciou também que a evolução tecnológica, impulsionada pelo Big Data, pelo aumento do poder computacional e pelo avanço dos modelos contemporâneos de inteligência artificial, ampliou exponencialmente a escala, a velocidade e a profundidade do tratamento de dados, criando novas dinâmicas informacionais que excedem os contornos tradicionais da tutela jurídica.

É precisamente nesse ponto que se insere o presente capítulo. Se o capítulo 2 examinou a formação histórica e conceitual do ambiente informacional, bem como as bases tecnológicas que viabilizaram o surgimento da IA contemporânea, este capítulo volta-se aos desafios específicos que a inteligência artificial introduz nesse ecossistema. A IA não apenas intensifica problemas já conhecidos da sociedade informacional, ela os reconfigura qualitativamente. Assim, a discussão desloca-se da mera coleta e circulação de dados para a análise de como sistemas algorítmicos processam esses dados, geram significados, moldam comportamentos e influenciam relações sociais, econômicas e jurídicas.

Diante disso, torna-se indispensável compreender de que modo a inteligência artificial produz novas formas de risco, especialmente no que diz respeito à opacidade algorítmica e à assimetria informacional, desafios que constituem o núcleo deste capítulo. Busca-se examinar como a complexidade técnica dos modelos contemporâneos, a concentração de infraestrutura computacional, o acúmulo de dados em grandes plataformas e a distância cognitiva entre desenvolvedores, reguladores e usuários produzem um ambiente no qual decisões automatizadas impactam indivíduos sem que estes disponham de meios adequados para compreendê-las, contestá-las ou influenciá-las.

Assim, este capítulo aprofunda as definições e características essenciais da IA, analisa seus métodos e classificações relevantes para o debate jurídico, e investiga os riscos informacionais que justificam a necessidade de instrumentos normativos e técnicos específicos. Mais do que descrever a tecnologia, busca-se situá-la dentro das

tensões estruturais da sociedade informacional.

3.1 INTELIGÊNCIA ARTIFICIAL

Em razão do caráter altamente disruptivo da inteligência artificial e em decorrência de seu potencial para processar e gerar decisões sobre grande volume de informações, a IA eleva a novos patamares as preocupações com privacidade e vigilância, pois algoritmos cada vez mais sofisticados podem não apenas identificar padrões de comportamento, mas também direcionar políticas e produtos de maneira quase invisível aos indivíduos. Nesse processo, o respeito aos princípios de proteção de dados torna-se ainda mais urgente e complexo, sob pena de amplificar desigualdades e cercear a autonomia. Porém, para regular essa tecnologia ou adequá-la dentro das regulações existentes, é necessário um conhecimento técnico mínimo de seu funcionamento. Desse modo, esse tópico volta-se especificamente ao tema da IA, em que serão examinadas suas definições, características e aplicações.

3.1.1 O conceito de IA

A compreensão do conceito de Inteligência Artificial (IA) revela-se fundamental para delimitar com precisão o seu objeto de estudo, suas aplicações e seus impactos sociais e jurídicos. No entanto, tal conceito não possui uma definição única e estática, variando significativamente conforme a abordagem teórica adotada, seja ela filosófica, técnica ou contextual. Essa multiplicidade de interpretações decorre do caráter interdisciplinar da IA, que perpassa campos como a ciência da computação, a matemática, a linguística, a neurociência e a filosofia da mente. Nesse sentido, identificar as definições mais relevantes e os autores que contribuíram de forma significativa para sua formulação é um passo indispensável para compreender os desafios normativos a serem enfrentados em decorrência dessa tecnologia.

John McCarthy, um dos fundadores da área e responsável por cunhar o termo “inteligência artificial” na conferência de Dartmouth em 1956, em texto posterior, intitulado *What is Artificial Intelligence?* define a IA como:

> É a ciência e a engenharia de construir máquinas inteligentes, especialmente programas de computador inteligentes. Está relacionada à tarefa semelhante de usar computadores para compreender a inteligência humana, mas a IA não precisa se restringir a métodos que sejam biologicamente observáveis (McCarthy, 2007, p. 2).

Essa definição enfatiza não apenas o aspecto técnico da construção de sistemas inteligentes, mas também uma distinção importante, apesar de seu surgimento ser baseado na reprodução do sistema da inteligência humana, essa não se limita no que podemos biologicamente observar ou reproduzir, podendo tornar-se mais avançada do que processos neurais humanos, o que amplia seu campo de atuação e suas implicações sociais.

Para uma definição mais contemporânea sobre inteligência artificial, Russell e Norvig (2021, p. 1-5) dividem quatro possíveis definições para IA: a) Sistemas que pensam como humanos, definindo-as como modelos cognitivos que reproduzem a inteligência humana; b) Sistemas que agem como humanos, sendo máquinas capazes de executar tarefas com desempenho comparável ou superior ao humano; c) Sistemas que pensam racionalmente, definindo como máquinas que conseguem pensar através da aplicação da lógica de raciocínio e d) Sistemas que agem racionalmente, que considera que IA é um conjunto de agentes inteligentes capazes de tomar decisões ótimas ou próximas do ideal, maximizando chances de sucesso, mesmo diante de incertezas ou limitações. Russell e Norvig preferem a última abordagem por permitir uma mensuração clara de desempenho e não estar restrita à reprodução de processos cognitivos humanos, valorizando mais a eficácia prática das máquinas inteligentes.

Pelo seu caráter amplo e flexível, o conceito de inteligência artificial possui diversos pontos de vista, de maneira mais genérica, pode-se dizer que é uma tecnologia que “busca fazer simulações de processos específicos da inteligência humana por meio de recursos computacionais” (Peixoto; Silva, 2019, p. 20). De forma parecida, Teixeira e Cheliga (2020, p. 16) fazem a conceituação como “sistema computacional criado para simular racionalmente as tomadas de decisão dos seres humanos, tentando traduzir em algoritmos o funcionamento do cérebro humano”.

Ambos os conceitos trazem uma ideia mais ampla de inteligência artificial como um sistema ou recurso computacional que imita em partes a inteligência humana. Para Peter Stone (2016, p. 3), é importante que seja feita uma conceituação ampla sobre o tema para não limitar pesquisas sobre o assunto, entendendo que a inteligência

artificial é um termo guarda-chuva que está dentro de diversas áreas como: ciências da computação, linguística, matemática, filosofia, probabilística, neurociência e teoria da decisão.

É importante destacar que há uma possível interpretação epistemológica errônea sobre o termo inteligência artificial, pois pode induzir que essa é dotada de autonomia em relação a interferência humana, porém, esta depende da criação humana de sua estrutura e aprendizado (Barreto Júnior; Venturi Junior, 2020, p. 340). A importância desta compreensão impacta diretamente nos estudos sobre a regulação do assunto, ao entender que a inteligência artificial age através de vieses algorítmicos, e não possui autonomia de vontade própria.

Neste ponto, insere-se a discussão entre IA fraca e a IA forte. Stuart Russel e Peter Norvig apresentam as distinções entre esses conceitos a partir de uma perspectiva filosófica. A IA fraca refere-se aos sistemas projetados para executar tarefas específicas com alto grau de eficiência, mas sem qualquer consciência, entendimento ou inteligência geral comparável à humana. Esse tipo de IA não possui intenções, sentimentos ou compreensão genuína, operando com base em algoritmos para resolver problemas delimitados. De outra forma, poderiam ser descritas como máquinas que realizam tarefas inteligentes, mas sem realmente “entenderem” ou replicarem a cognição humana. A IA forte, por sua vez, é descrita como máquinas com consciência genuína, entendimento profundo, autonomia plena e capacidade de raciocínio geral, ou seja, uma inteligência equivalente à humana ou superior em todos os aspectos cognitivos (RUSSEL; NORVIG, 2021, p. 1020-1033).

No presente estado da arte de IA, a IA forte representa um conceito teórico e ainda não alcançado na prática. Apesar de avanços impressionantes em modelos como o ChatGPT e outros modelos atuais que podem manter diálogos complexos, gerar textos criativos ou resolver problemas matemáticos, essas capacidades derivam do processamento estatístico de grandes volumes de dados e padrões linguísticos, não de compreensão real, consciência ou intencionalidade. Esses sistemas não possuem mente, subjetividade, nem a capacidade de entender o mundo como os humanos compreendem, portanto, ainda são classificados como IA fraca. Porém, fica evidente que há uma busca, quase utópica e ideológica, para a criação da IA forte.

Na obra *Responsible Artificial Intelligence* (2019), Virginia Dignum reconhece que não há uma definição única para inteligência artificial (IA), dada a amplitude do campo e as diversas abordagens envolvidas. Para compreensão deste conceito, ela

apresenta duas diferentes perspectivas. A primeira perspectiva é da ciência da computação e da engenharia, sendo essa uma visão técnica, que enfatiza o aspecto funcional da IA de construir sistemas computacionais capazes de executar tarefas que normalmente requereriam inteligência humana, como reconhecimento de fala, visão, planejamento ou tomada de decisão. A segunda perspectiva é filosófica, compreendendo a IA como um campo de investigação sobre a própria natureza da inteligência, questionando se é possível reproduzir ou simular o pensamento humano em máquinas (DIGNUM, 2019, p. 9-16).

Em uma perspectiva sociológica, através de lentes críticas aos conceitos técnicos de IA, Kate Crawford, na obra *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021, p. 7-9), compreende que a inteligência artificial não é nem “artificial” nem “inteligente”, por não ser uma entidade autônoma e puramente técnica, mas uma formação sociotécnica e material profundamente enraizada em estruturas de poder. Para a autora, a IA é resultado de complexas cadeias produtivas que envolvem extração de recursos naturais, exploração de trabalho humano e investimentos massivos de capital, refletindo e reproduzindo desigualdades sociais, econômicas e políticas. Nesse sentido, mais do que algoritmos ou redes neurais, a inteligência artificial deve ser entendida como um arranjo industrial que corporifica interesses hegemônicos e está inextricavelmente ligada a dinâmicas históricas, culturais e institucionais que moldam sua produção e aplicação.

Apesar das abordagens filosóficas e sociológicas que enriquecem o debate conceitual sobre a inteligência artificial, é necessário reconhecer que, para o campo jurídico e, especialmente, para fins regulatórios, torna-se imprescindível a adoção de uma definição mais pragmática e operacional. De forma mais técnica, na proposta de regulação da inteligência artificial no âmbito da União Européia, denominada *Artificial Intelligence Act*, foi sugerido o seguinte conceito, em seu artigo 3º, item 1, traduzido:

> 'sistema de inteligência artificial' (sistema IA) significa um software que é desenvolvido com uma ou mais das técnicas e abordagens listadas no Anexo I e pode, para um conjunto definido de objetivos humanos, gerar saídas como conteúdo, previsões, recomendações ou decisões que influenciam os ambientes com os quais interagem; (Comissão Europeia, 2021)

No citado Anexo I, são citados três tipos de técnicas e abordagens da inteligência artificial: a) Abordagens de aprendizado de máquina (machine learning), incluindo o aprendizado supervisionado, não supervisionado, por reforço, e de

aprendizado profundo (deep learning); b) Abordagens baseadas em lógica e conhecimento; c) Abordagens estatísticas, incluindo métodos de buscas e otimização (COMISSÃO EUROPEIA, 2021).

No Brasil, atualmente, tramita o PL nº 2338/2023, apresentado pelo Senador Rodrigo Pacheco. O projeto de lei está sob análise da Comissão Temporária sobre Inteligência Artificial no Brasil. Neste projeto, foi proposta a seguinte conceituação sobre inteligência artificial:

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

Diante da pluralidade de abordagens apresentadas, que vão desde as concepções filosóficas e sociológicas até definições técnico-funcionais da ciência da computação e das iniciativas regulatórias, constata-se que o conceito de inteligência artificial se encontra em constante construção e adaptação. Cada perspectiva analisada contribui para o delineamento de um entendimento mais robusto e contextualizado do fenômeno. Reconhecer a complexidade inerente à definição da IA não significa renunciar a uma delimitação conceitual, mas sim compreender que esta deve ser suficientemente abrangente para abarcar seus múltiplos aspectos, sem perder de vista a necessidade de precisão e aplicabilidade, especialmente em contextos jurídicos e regulatórios.

3.1.2 Técnicas e Métodos Principais

A compreensão das principais técnicas e métodos subjacentes à Inteligência Artificial é essencial para delimitar os contornos de sua aplicação prática, bem como para subsidiar uma regulação eficaz e proporcional aos riscos envolvidos. Dentre os instrumentos tecnológicos mais relevantes na atualidade, destacam-se o *machine learning*, as técnicas de aprendizado automatizado, o *deep learning*, a inteligência artificial generativa e os modelos de linguagem de larga escala (*Large Language*

Models – LLMs).

A escolha de técnicas e métodos é inevitável diante da amplitude deste campo de estudo, mas precisamente a escolha destas trazidas anteriormente decorre da centralidade que essas abordagens adquiriram no desenvolvimento contemporâneo da Inteligência Artificial (IA). No prefácio de sua 4ª edição da obra *Artificial Intelligence: A Modern Approach* de 2021, Stuart Russel e Peter Norvig destacam que ao contrário da 3ª edição de 2010, esta nova versão prioriza o *machine learning* em detrimento da engenharia de conhecimento manual, justificada pela ampla disponibilidade de dados, maior poder computacional e o avanço dos algoritmos. Além disso, o *deep learning* ganhou capítulos próprios devido ao seu impacto crescente (Russel, Norvig, 2021, p. 5). A opção por tratar de *Large Language Models* (LLMs) e *IA Generativa* se justifica, ainda, pelo destaque que essas tecnologias como ChatGPT da OpenAI, Gemini da Google, Claude da Anthropic, LLaMa da Meta (Facebook) e Deepseek-R1 da startup chinesa Deepseek, vêm recebendo, tanto no meio acadêmico quanto nos ambientes regulatórios, empresariais e sociais.

Cada uma dessas abordagens representa não apenas um avanço técnico, mas também implica novos desafios, em virtude de sua crescente capacidade de autonomia operacional e impacto social. Assim, este capítulo se propõe a examinar as referidas técnicas em seus aspectos conceituais, funcionais e normativos, de forma a construir um arcabouço teórico capaz de dialogar criticamente com as inovações tecnológicas contemporâneas e seus desdobramentos na esfera jurídica.

O ponto de partida para a análise de todas técnicas e métodos propostos é o *machine learning* ou aprendizado de máquina, pois é a base fundadora para todas demais. A capacidade de aprender é o ponto fundamental que torna a inteligência artificial fascinante atualmente.

O aprendizado, de maneira geral, acontece quando um agente aprimora seu desempenho a partir de observações feitas sobre o mundo. Se esse agente é uma máquina, denominamos isso como aprendizado de máquina ou *machine learning*. Essa técnica é definida por Russel e Norvig (2021, p. 693) como: “um computador observa alguns dados, constrói um modelo com base nesses dados e usa esse modelo tanto como uma hipótese sobre o mundo quanto como um software que pode resolver problemas”. Em linhas gerais, isso pode ser descrito como uma aprendizagem a partir de exemplos ou experiências.

A definição de aprendizado de máquina evoluiu ao longo das décadas,

refletindo tanto avanços técnicos quanto diferentes perspectivas epistemológicas. Em sua formulação pioneira, Arthur Samuel (1959) descreveu-o como o “campo de estudo que dá aos computadores a habilidade de aprender sem ser explicitamente programado” (Géron, 2019, p. 4). Posteriormente, Tom Mitchell (1997) propôs um conceito mais formalizado e voltado à prática científica, ao afirmar que “diz-se que um programa de computador aprende pela experiência E em relação a algum tipo de tarefa T e alguma medida de desempenho P se o seu desempenho em T, conforme medido por P, melhora com a experiência E” (Géron, 2019, p. 4). Na obra contemporânea de Aurélien Géron (2019), essas abordagens são sintetizadas na seguinte definição: “aprendizado de máquina é a ciência (e a arte) da programação de computadores para que eles possam aprender com os dados” (GÉRON, 2019, p. 4).

O estudo do aprendizado de máquina está atrelado inevitavelmente ao campo de estudo da inteligência artificial, porém, o *machine learning* se alimenta da fonte de outros campos de estudos como matemática, lógica, probabilidade e estatística, teoria da computação, neurociência, teoria da informação e outros (Faceli et al, 2011, p. 3).

Segundo Mitchell (2019, p. 37-40), o campo da inteligência artificial passou por uma ruptura significativa com o surgimento do aprendizado de máquina, que se afastou propositalmente da abordagem simbólica tradicional, frequentemente chamada de *GOFAI* (*Good Old-Fashioned Artificial Intelligence*), e passou a se consolidar como eixo central da pesquisa e inovação tecnológica. A partir da década de 2000, uma sucessão de conquistas tecnológicas, como sistemas de tradução automática, assistentes virtuais, reconhecimento facial e veículos autônomos, evidenciou o impacto disruptivo dessas técnicas, colocando a IA no centro do interesse científico, comercial e midiático, como também reformulou o mercado de startups: atualmente, grande parte dos novos modelos de negócio gira em torno da lógica de "pegue X e adicione IA", o que demonstra a centralidade dessa tecnologia na economia digital contemporânea.

Mas por qual motivo queremos que as máquinas aprendam? Para Russell e Norvig (2021, p. 694) há duas razões centrais para se desejar que as máquinas aprendam. A primeira reside na impossibilidade de os programadores anteciparem todas as situações futuras que um sistema pode enfrentar, como no caso de robôs que navegam por labirintos ou programas que analisam o mercado financeiro e precisam se adaptar a cenários variáveis. A segunda razão decorre da própria

limitação humana, pois muitas tarefas que os indivíduos realizam de forma intuitiva, como o reconhecimento facial, não são facilmente traduzíveis em instruções escritas, o que torna os algoritmos de aprendizado de máquina indispensáveis para solucionar problemas cuja lógica é de difícil formalização.

Em todas as definições abordadas, compreende-se que aprendizado de máquina ou *machine learning* possui um fundamento nuclear na possibilidade de uma máquina aprender a partir de experiências e observações que são alimentadas através de dados. Em razão de sua capacidade de permitir que máquinas extraiam padrões, adaptem-se a contextos mutáveis e resolvam problemas sem depender de instruções explícitas, é possível afirmar que o aprendizado de máquina se consolidou como o núcleo dinâmico da inteligência artificial contemporânea. Compreendido que as máquinas podem aprender, impõe-se uma pergunta fundamental: como, afinal, as máquinas aprendem? A resposta a essa indagação exige a análise das diferentes formas de aprendizado que estruturam esse campo.

As formas de aprendizado das máquinas são comumente classificadas de acordo com o modelo de supervisão que estão submetidas em seu treinamento e categorizadas em sistema supervisionado, não supervisionado, semissupervisionado e aprendizado por reforço (Géron, 2019, p. 8). Essa classificação também pode ser dividida entre tarefas preditivas, sendo essas através de sistemas supervisionados, ou tarefas descritivas, sendo essas através de sistemas não supervisionados. As tarefas preditivas são de previsão, objetivando encontrar um modelo ou hipótese, a partir dos dados de treinamento para prever novas situações, e as tarefas descritivas são de descrição, onde o objetivo é explorar um conjunto de dados (FACELI *et al,* 2011, p. 8).

O aprendizado supervisionado pode ser conceituado tecnicamente como:

> O termo supervisionado vem da simulação da presença de um “supervisor externo”, que conhece a saída (rótulo) desejada para cada exemplo (conjunto de valores para os atributos de entrada). Com isso, o supervisor externo pode avaliar a capacidade da hipótese induzida de predizer o valor de saída para novos exemplos (FACELI *et al,* 2011, p. 6).

Também de forma técnica, o aprendizado supervisionado ocorre quando o agente é exposto a pares de entrada e saída e aprende uma função que mapeia as entradas para as saídas. As entradas são os dados brutos ou processados que alimentam o sistema de aprendizado, as saídas são chamadas de rótulos e servem

para classificar cada entrada. Por exemplo, as entradas podem ser imagens captadas por uma câmera, e a saída objetiva identificar e classificar entre “ônibus” ou “pedestre” (RUSSEL; NORVIG, 2021, p. 701).

De forma didática, o aprendizado supervisionado pode ser compreendido a partir de uma analogia simples: ensinar uma criança a reconhecer frutas. Nesse processo, a criança observa diversas imagens acompanhadas de seus respectivos nomes, por exemplo, “isto é uma maçã”, “isto é uma banana”, “isto é uma laranja”. Ou seja, cada informação visual (entrada) vem acompanhada de uma resposta correta (rótulo), o que permite à criança associar padrões visuais aos nomes das frutas. Da mesma forma, no aprendizado supervisionado, a inteligência artificial é treinada com um conjunto de dados rotulados previamente, aprendendo a partir dessa relação explícita entre entradas e saídas. Após esse treinamento, a IA torna-se capaz de generalizar o conhecimento adquirido e realizar previsões ou classificações corretas quando confrontado com novos dados.

O aprendizado não supervisionado, por sua vez, “os dados de treinamento não são rotulados, o sistema tenta aprender sem um professor” (Géron, 2019, p. 10). Nesse caso, a inteligência artificial busca identificar padrões e estruturas nos dados de entrada sem receber qualquer exemplificação de saída ou rótulo prévio. A tarefa mais comum nesse tipo de abordagem é o agrupamento (*clustering*), no qual o sistema organiza os dados em grupos com características semelhantes. Por exemplo, ao analisar um grande conjunto de imagens obtidas da internet, um sistema de visão computacional pode formar automaticamente um grupo com imagens semelhantes entre si, as quais um ser humano reconheceria como “gatos” (RUSSELL; NORVIG, 2021, p. 702).

A partir do mesmo exemplo da criança e das frutas utilizado anteriormente, para tornar mais didática a compreensão do aprendizado não supervisionado, esse pode ser ilustrado pela situação em que uma criança recebe um cesto com diversas frutas, mas sem que ninguém lhe diga o nome de cada uma delas. Diante disso, ela começa a observar cores, formatos e tamanhos, agrupando as frutas por similaridade. Ainda que não conheça os nomes corretos, ela identifica padrões e organiza os objetos com base em suas características percebidas. De maneira análoga, no aprendizado não supervisionado, o sistema computacional é alimentado apenas com dados brutos, sem rótulos ou respostas prévias, e deve ser capaz de detectar padrões ocultos nesses dados.

O aprendizado semissupervisionado é uma mescla entre os dois sistemas compreendidos acima, lidando com “uma grande quantidade de dados não rotulados e um pouco de dados rotulados” (Géron, 2019, p. 13). Um exemplo prático do aprendizado não supervisionado pode ser observado em serviços de organização de imagens, como o Google Fotos, que são capazes de identificar automaticamente padrões visuais semelhantes em diferentes fotos, agrupando rostos que pertencem à mesma pessoa sem que o usuário precise informá-lo previamente. Após esse agrupamento, basta a atribuição de um único rótulo para que o sistema reconheça e nomeie todas as ocorrências daquela pessoa, facilitando a busca e a organização das imagens (GÉRON, 2019, p. 13).

Por fim, no aprendizado por reforço, o agente não recebe instruções explícitas sobre quais ações tomar, mas aprende por meio da interação com o ambiente, recebendo sinais de recompensa ou punição após determinadas ações. Com base nesses resultados, ele ajusta seu comportamento para maximizar recompensas futuras, identificando quais decisões anteriores contribuíram para o sucesso ou fracasso de uma tarefa, como ocorre em jogos como o xadrez, onde a informação sobre vitória ou derrota serve como estímulo para aprendizado adaptativo (Russell; Norvig, 2021, p. 703). De forma simplificada, “a meta é reforçar ou compensar uma ação considerada positiva e punir uma ação considerada negativa” (Faceli et al, 2011, p. 7). O ajuste de seu comportamento nesse caso é feito com o aprendizado da melhor estratégia para obter o maior número de recompensas ao longo do tempo, sendo isso denominado como política (GÉRON, 2019, p. 14).

Retomando o exemplo utilizado da criança e o cesto de frutas, no caso do aprendizado por reforço, uma criança, sem saber previamente quais frutas são doces ou azedas, decide prová-las uma a uma. A cada mordida, ela percebe que algumas têm um sabor agradável (recompensa), enquanto outras são desagradáveis (punição). Com o tempo, ela passa a reconhecer quais frutas prefere e escolhe aquelas que lhe proporcionam experiências positivas.

Outro aspecto relevante na configuração de sistemas de aprendizado de máquina diz respeito à forma como os dados são adquiridos durante o treinamento do modelo, podendo ocorrer de maneira em lotes (*batch learning*) ou online (*online learning*). No aprendizado em lotes, o sistema é treinado com um conjunto de dados alimentado de uma só vez, sendo ideal para cenários em que o volume de dados é estático. Já no aprendizado online, o modelo é atualizado de forma constante, à

medida que os dados são recebidos em fluxo contínuo, o que permite sua adaptação em tempo real (GÉRON, 2019, p. 9).

Diante das diferentes abordagens apresentadas, observa-se que os modelos de aprendizado supervisionado, não supervisionado, semissupervisionado e por reforço, bem como o aprendizado em lotes e o aprendizado online oferecem distintos caminhos pelos quais as máquinas podem adquirir conhecimento a partir de dados e experiências. Cada técnica possui características próprias quanto ao grau de intervenção humana, à presença de rótulos e à forma de adaptação ao ambiente, permitindo que os sistemas se tornem progressivamente mais autônomos e eficazes na resolução de tarefas complexas. Essas estratégias de aprendizado, embora distintas em sua lógica e aplicação, compartilham o mesmo objetivo: possibilitar que máquinas compreendam padrões, tomem decisões e evoluam com base em interações com o mundo real.

Consolidado esse panorama, impõe-se o aprofundamento em uma vertente que tem se destacado de maneira expressiva nos últimos anos em razão de seu desempenho em tarefas que envolvem linguagem, imagem e tomada de decisão, o aprendizado profundo ou *deep learning*, técnica que não é uma categoria distinta de aprendizado, mas sim uma subárea ou técnica que se insere dentro das categorias tradicionais.

O surgimento do *deep learning*, como vertente autônoma e central no campo da inteligência artificial, remonta ao ano de 2006, quando Geoffrey Hinton e colaboradores publicaram um artigo demonstrando ser possível treinar redes neurais profundas capazes de reconhecer dígitos manuscritos com acurácia superior a 98%. Até então, prevalecia entre os pesquisadores a crença de que redes profundas seriam inviáveis de treinar, razão pela qual a abordagem havia sido praticamente abandonada desde os anos 1990. O êxito alcançado não apenas reabilitou o interesse científico pelo tema, mas também inaugurou uma nova era de pesquisas, marcada por avanços substanciais que superavam os limites das técnicas tradicionais de aprendizado de máquina, especialmente quando associadas a grandes volumes de dados e poder computacional elevado (GÉRON, 2019, p. 14).

O aprendizado profundo, ou *deep learning*, constitui uma das mais importantes abordagens contemporâneas no campo da inteligência artificial, destacando-se pela utilização de redes neurais artificiais com múltiplas camadas hierárquicas que permitem a extração de padrões complexos a partir de grandes volumes de dados. O

aprendizado profundo se destaca por sua capacidade de processar informações de maneira hierárquica, construindo representações complexas a partir de elementos simples. Quando, por exemplo, uma imagem é fornecida a uma rede neural profunda, o sistema não tenta reconhecer diretamente o objeto final. Em vez disso, a informação percorre diversas camadas: inicialmente, são identificadas bordas e contrastes básicos entre pixels; em seguida, essas bordas são combinadas para formar contornos, cantos e formas mais estruturadas; posteriormente, a rede passa a reconhecer partes específicas de objetos, como rostos, membros ou componentes visuais característicos; por fim, essas partes são integradas para que o modelo seja capaz de classificar o conteúdo da imagem em categorias gerais, como “pessoa”, “carro” ou “animal”. Cada camada contribui com um nível mais abstrato de interpretação, permitindo que a rede aprenda padrões complexos de forma progressiva e autônoma. Esse encadeamento de transformações sucessivas é o que torna o *deep learning* especialmente eficaz em tarefas que envolvem visão computacional, reconhecimento de fala e outras aplicações sensoriais. (GOODFELLOW *et al*., 2016, p. 20).

Como observa Mitchell (2019, p. 70), é essencial compreender que o termo *deep* no conceito de *deep learning* não se refere à sofisticação do conteúdo aprendido, mas sim à quantidade de camadas ocultas presentes na rede neural treinada. Essa distinção conceitual é relevante, pois a profundidade da rede está diretamente relacionada à sua capacidade de abstração hierárquica. Além disso, a autora destaca que as redes neurais profundas mais eficazes são aquelas cuja estrutura se inspira diretamente em descobertas da neurociência, especialmente no funcionamento do sistema visual humano. Diferentemente das redes tradicionais com múltiplas camadas, que apenas buscavam inspiração geral no cérebro humano, as arquiteturas modernas de *deep learning* mimetizam de forma mais fiel as estruturas cerebrais envolvidas na percepção visual, o que tem contribuído significativamente para seu êxito em tarefas como classificação de imagens e reconhecimento de padrões.

Esses modelos computacionais inspirados no funcionamento do cérebro humano são denominados como redes neurais artificiais (RNAs) e são compostos por unidades simples interconectadas, denominadas neurônios, que se organizam em camadas. Nesse contexto, o aprendizado profundo (deep learning) configura-se como uma subcategoria das RNAs, caracterizada pelo uso de múltiplas camadas ocultas entre as camadas de entrada e de saída. A profundidade da rede, ou seja, o número

de camadas, é o que define essa abordagem, e não necessariamente a complexidade da tarefa realizada. Redes profundas são capazes de extrair representações altamente abstratas dos dados, a partir de uma arquitetura hierárquica que aprende padrões em diferentes níveis (DIGNUM, 2019, p. 27-28).

O aprendizado profundo se firmou como uma das principais formas de construir sistemas de inteligência artificial mais eficientes e capazes. Por meio de suas várias camadas, ele permite que as máquinas aprendam de forma progressiva, identificando padrões simples e, a partir deles, construindo interpretações mais complexas dos dados. Essa forma de aprendizado tem impulsionado grandes avanços em áreas como reconhecimento de imagens, tradução automática, diagnósticos médicos e, principalmente, na criação de sistemas capazes de gerar textos, imagens e sons com alto grau de realismo. É justamente nesse ponto que surge a chamada inteligência artificial generativa, um dos desdobramentos mais marcantes do uso do *deep learning* nos últimos anos.

A inteligência artificial generativa ocupa a posição mais avançada dentro do campo do aprendizado profundo, sendo considerada um dos desdobramentos mais sofisticados dessa tecnologia. Por muitos anos, os modelos de redes neurais enfrentavam limitações estruturais e computacionais que inviabilizavam sua aplicação prática em tarefas criativas. Esse cenário começou a se transformar com o desenvolvimento de arquiteturas mais profundas e complexas, associadas ao aumento exponencial da capacidade de processamento e à disponibilidade de grandes volumes de dados. (GOODFELLOW *et al.,* 2016, p. 12-20).

O ponto de inflexão para a consolidação da IA generativa ocorreu com o desenvolvimento e a disseminação dos modelos de linguagem baseados na arquitetura *transformer*. Os modelos transformer são um tipo de sistema utilizado na inteligência artificial para entender e produzir linguagem, como textos, falas e comandos. Eles foram criados para superar limitações de modelos anteriores, que analisavam as palavras de um texto em sequência, como quem lê palavra por palavra. O diferencial do transformer é que ele consegue analisar todas as palavras ao mesmo tempo, entendendo o contexto completo de uma frase ou parágrafo. Isso permite ao sistema "compreender" que palavras como “juiz” e “tribunal” aparecem em contextos semelhantes, mesmo sem uma definição formal. (RUSSELL E NORVIG, 2021, p. 1606).

Esse cenário se transformou totalmente com o surgimento dos *Large Language*

Models (LLMs). Os Modelos de Linguagem de Grande Escala (LLMs), conforme explica Kaplan (2023, p. 33-34), são sistemas de inteligência artificial generativa que produzem respostas em linguagem natural a partir de perguntas ou comandos. Com a utilização da arquitetura *transformer* treinadas com enormes conjuntos de textos disponíveis publicamente, como livros, sites e artigos. Os LLMs operam não a partir de regras linguísticas explícitas, mas sim por meio da detecção de padrões estatísticos que emergem dos dados analisados.

A partir de 2018, modelos como GPT-2 e RoBERTa demonstraram capacidades surpreendentes em tarefas de linguagem natural. Esses modelos treinados com enormes volumes de textos retirados da internet, tornando possível que esses sistemas gerem novos conteúdos, como pareceres, petições, relatórios ou notícias, com base nos padrões que aprenderam. Esses modelos, originalmente projetados para prever a próxima palavra em um texto, passaram a ser utilizados em traduções, respostas a perguntas, compreensão textual e geração criativa de conteúdo (RUSSELL E NORVIG, 2021, p. 1607-1608).

A IA generativa, assim, deixou de ser um experimento laboratorial e passou a integrar sistemas de produção textual, visual e sonora, modificando profundamente os fluxos de criação, personalização e circulação de informação. Após o domínio das redes neurais profundas, as pesquisas avançam para modelos capazes não apenas de reconhecer ou classificar dados, mas de gerar novas informações de maneira autônoma. Em vez de apenas identificar ou responder, esses sistemas passaram a simular a criatividade humana, reproduzindo estilos, estruturas e conteúdos que se assemelham ao que um ser humano produziria. Trata-se de um novo patamar da inteligência artificial, no qual os algoritmos deixam de apenas processar o mundo e passam também a construí-lo.

Embora celebrados por seu potencial técnico, os LLMs também geram importantes implicações político-econômicas, especialmente no que se refere à concentração de poder. O AI Now Institute (2023, p. 15-18) argumenta que o desenvolvimento e a operação desses modelos exigem infraestruturas computacionais massivas, conjuntos de dados bilionários e equipes altamente especializadas, recursos que hoje estão acessíveis apenas a um grupo restrito de empresas, como OpenAI, Google, Amazon, Meta e Microsoft. Ao dominarem a produção, distribuição e o treinamento desses sistemas, tais corporações reforçam sua posição estratégica no ecossistema digital, dificultando a regulação democrática

e a concorrência justa. Além disso, os LLMs contribuem para a ampliação da vigilância de dados, para a disseminação de desinformação e para a opacidade dos processos decisórios automatizados, aprofundando desigualdades informacionais e jurídicas.

Diante do exposto, observa-se que as técnicas e métodos centrais da inteligência artificial podem ser organizados em uma sequência de complexidade crescente, partindo do aprendizado de máquina (machine learning), que permite que sistemas identifiquem padrões e façam previsões com base em dados, passando pelo aprendizado profundo (deep learning), que se vale de redes neurais com múltiplas camadas para representar informações de forma hierárquica, até alcançar os atuais modelos de inteligência artificial generativa, capazes não apenas de reconhecer ou classificar dados, mas de produzir novos textos, imagens e sons. Essa trajetória revela um salto qualitativo na autonomia e na capacidade de abstração dos sistemas de IA, que hoje já desempenham funções antes consideradas exclusivamente humanas. As diferentes modalidades de aprendizado supervisionado, não supervisionado, por reforço ou semissupervisionado representam estratégias para organizar esse processo de aprendizagem, sendo complementadas por formas específicas de treinamento, como o aprendizado por lotes ou online.

Ao centro dessa transformação tecnológica estão os Large Language Models (LLMs), que integram o estado da arte da IA contemporânea. Esses modelos demonstram como o avanço técnico, viabilizado pela arquitetura transformer, permitiu que máquinas compreendessem e manipulassem a linguagem com altíssimo grau de sofisticação. Entretanto, para além das inovações, essas tecnologias impõem desafios normativos significativos: afetam a liberdade, a privacidade, a proteção de dados, o livre desenvolvimento da personalidade, alteram a dinâmica do trabalho e concentram poder econômico e informacional nas mãos de poucas empresas. Por essa razão, o exame das técnicas e métodos da IA não se limita ao seu funcionamento interno, mas exige uma análise crítica de suas implicações jurídicas, sociais e políticas. Com isso, o presente capítulo busca não apenas esclarecer tecnicamente o leitor, mas oferecer uma base sólida para reflexões regulatórias nas seções seguintes.

3.2 Desafios da Sociedade Informacional

A crescente sofisticação dos sistemas de inteligência artificial, impulsionada

pelas técnicas de aprendizado profundo e pelos modelos de linguagem de larga escala (LLMs), trouxe consigo não apenas avanços expressivos em termos de desempenho técnico, mas também um conjunto complexo de novos desafios jurídicos e sociais. Entre esses desafios, destaca-se de forma recorrente a chamada opacidade algorítmica, fenômeno que diz respeito à dificuldade ou até mesmo impossibilidade de compreender, auditar ou explicar de maneira clara o funcionamento interno dos modelos computacionais que processam e produzem decisões com impactos significativos na vida das pessoas. Essa opacidade não decorre apenas de negligência ou má-fé, mas também de fatores técnicos e estruturais, devido ao alto grau de complexidade matemática, da natureza estatística do aprendizado de máquina e da ausência de mecanismos técnicos de interpretação legíveis para humanos que extrapolam a própria condição humana.

Essa condição técnica da IA moderna se articula diretamente com um segundo problema de natureza política e normativa: a assimetria informacional entre os atores que desenvolvem e operam os sistemas algorítmicos, geralmente grandes corporações privadas ou governos, e os indivíduos, organizações e até instituições públicas que são impactados por suas decisões. A falta de transparência sobre como os algoritmos funcionam, somada à concentração do conhecimento técnico, da infraestrutura computacional e dos dados em poder de poucos agentes econômicos, amplia a assimetria de poder e fragiliza os mecanismos tradicionais de responsabilização. Diante disso, este item propõe uma análise integrada da opacidade algorítmica, da assimetria informacional e dos riscos jurídicos estruturantes que emergem desse cenário, com vistas a fundamentar, nos capítulos seguintes, a centralidade dos princípios da transparência e da *accountability* como respostas normativas indispensáveis na regulação da inteligência artificial.

3.2.1 Opacidade Algorítmica

A compreensão da opacidade algorítmica como um fenômeno jurídico e técnico relevante exige, antes de tudo, a identificação de suas formas constitutivas. Embora o termo tenha ganhado destaque no debate público e regulatório, sua natureza é multifacetada e demanda uma abordagem conceitual mais precisa. Nesse sentido, a tipologia desenvolvida por Jenna Burrell (2016) oferece um marco teórico robusto para

sistematizar as origens e implicações da opacidade nos sistemas de inteligência artificial, distinguindo entre intencionais, técnicas e estruturais. Tal estrutura permite compreender como a opacidade se manifesta não apenas como uma consequência do avanço tecnológico, mas também como resultado de escolhas econômicas, barreiras cognitivas e desigualdades institucionais. A autora defende que a opacidade não é somente técnico, mas sim um problema político e ético quando aplicado em decisões de impacto social direto, como algoritmos de classificação e ranqueamento de pessoas, exemplificados por esta em sistemas de concessão de crédito, análise de candidatos, investigações policiais e pontuações de consumidores ou cidadãos (BURRELL, 2016, p.1).

A opacidade intencional decorre da atuação deliberada de agentes que optam por restringir o acesso a informações sobre os sistemas algorítmicos com o objetivo de proteger segredos comerciais, estratégias de mercado ou propriedade intelectual. Essa forma de opacidade é frequentemente justificada pela lógica concorrencial, mas implica sérios entraves à supervisão externa, ao escrutínio público e à responsabilização jurídica de sistemas que impactam diretamente a esfera de direitos dos indivíduos (BURRELL, 2016, p. 3).

A opacidade técnica emerge da complexidade intrínseca dos modelos de aprendizado profundo (*deep learning*), pois suas múltiplas camadas de processamento dificultam a rastreabilidade lógica das decisões produzidas. Ainda que o código-fonte esteja disponível, a operação interna desses sistemas permanece indecifrável, inclusive para seus próprios desenvolvedores, em razão da natureza não-linear, probabilística e adaptativa dos algoritmos envolvidos (Burrell, 2016, p. 2-3).

Por fim, a opacidade estrutural refere-se ao descompasso entre os conhecimentos especializados exigidos para compreender os sistemas algorítmicos e o nível de compreensão dos usuários, reguladores ou cidadãos afetados por suas decisões. Esse tipo de opacidade não é intencional nem necessariamente técnica, mas resulta da assimetria informacional e cognitiva entre desenvolvedores e o público em geral, evidenciando uma barreira epistemológica que fragiliza a capacidade de contestação, fiscalização e participação democrática (BURRELL, 2016, p. 4).

A abordagem proposta por Burrell revela, portanto, que a opacidade algorítmica não é um fenômeno uniforme, mas sim multifacetado, exigindo soluções jurídicas e regulatórias diferenciadas conforme sua origem e natureza. Tal compreensão é fundamental para a formulação de mecanismos normativos eficazes de mitigação da

opacidade.

Ao aprofundar a problemática da opacidade intencional, Pasquale (2015) introduz a noção de sociedade da caixa-preta (*black box society*), caracterizada pela crescente invisibilidade dos processos decisórios automatizados que regulam a vida econômica, política e cultural dos indivíduos. Segundo o autor, algoritmos opacos operam como mecanismos de classificação invisíveis, moldando comportamentos, determinando acessos e estruturando oportunidades com base em lógicas que escapam ao controle democrático e à fiscalização pública. Trata-se de uma infraestrutura de poder informacional que atua de modo silencioso, mas com efeitos profundos sobre os direitos fundamentais, em especial a privacidade, a igualdade e a liberdade de expressão.

Para Pasquale (2015, p. 8-12), a opacidade não é apenas uma limitação técnica ou uma externalidade regulável, mas um recurso estratégico mobilizado por corporações e governos para manter posições de domínio, ocultar práticas discriminatórias e dificultar a responsabilização. O funcionamento das plataformas digitais, dos sistemas de pontuação de crédito, dos algoritmos preditivos no setor penal e das ferramentas de análise de perfil de consumidores é intencionalmente mantido fora do alcance do público e mesmo das autoridades reguladoras. Nesses contextos, o desconhecimento não é acidental, mas funcional: ele permite o exercício de um poder classificatório assimétrico, que reproduz desigualdades e neutraliza o exercício de direitos. O autor recorre à metáfora do “espelho unidirecional” (one-way mirror), como aqueles utilizados em departamentos policiais, para ilustrar a assimetria informacional que estrutura a economia digital: enquanto as corporações e os sistemas automatizados têm acesso contínuo, profundo e detalhado aos dados, decisões e preferências dos usuários, estes são mantidos em ignorância quanto aos critérios que os classificam, segmentam ou excluem.

Essa forma de opacidade, classificada por Burrell (2016) como intencional, representa uma escolha estratégica dos agentes econômicos que operam sistemas algorítmicos, os quais deliberadamente restringem o acesso à lógica de funcionamento de seus modelos com o objetivo de proteger segredos comerciais, manter vantagens competitivas e evitar responsabilização. Trata-se de uma forma de invisibilidade construída, que visa não apenas ocultar os processos internos da decisão automatizada, mas também deslocar o ônus da transparência para os indivíduos afetados

No “capitalismo de vigilância” de Shoshana Zuboff (2019), a matéria prima é o comportamento humano que é utilizado como meio para gerar lucro e controle de mercado através de tecnologias para monitorar, prever e influenciar. Ao formular sua teoria, a autora destaca que essa lógica, embora difícil de detectar, em razão da opacidade, é altamente eficaz, pois explora as necessidades humanas de maneira direta, fornecendo informações ilimitadas e antecipando desejos de forma simples, transformando a vigilância e os padrões comportamentais em lucro.

Stefano Rodotà (2008, p. 111) ao tratar da sociedade de classificação expõe que um dos principais usos dessa ferramenta é político para controle dos cidadãos e que essa sociedade, segundo o autor, torna-se qualificada como autoritária ou ditatorial. No contexto de governos, essas tecnologias podem ser utilizadas em áreas sensíveis como segurança pública, fiscalização tributária, programas sociais e controle migratório. Como demonstram Brauneis e Goodman (2018, p. 108-109), mesmo em regimes democráticos e dotados de legislações de acesso à informação, a utilização de algoritmos por governos frequentemente escapa ao escrutínio público, com base em alegações de sigilo contratual ou segredo comercial. Tal prática contribui para consolidar a opacidade institucionalizada, impedindo que a sociedade e os órgãos de controle exerçam qualquer forma de fiscalização ou responsabilização sobre os critérios decisórios adotados.

Kaminski (2020, p. 51-55) reforça esse diagnóstico ao demonstrar como as estruturas jurídicas existentes, especialmente em regimes orientados pela proteção de dados pessoais, se mostram insuficientes para garantir *accountability* privada. A autora argumenta que a resistência das empresas em fornecer informações significativas sobre seus sistemas sob o pretexto de sigilo comercial neutraliza os mecanismos clássicos de responsabilização, como o direito à explicação ou o devido processo. Assim, a *accountability* torna-se condicional à vontade da própria corporação, o que subverte a lógica da regulação jurídica.

Essa dificuldade é também enfrentada por jornalistas investigativos e atores da sociedade civil, conforme aponta Diakopoulos (2013, p. 3-6). Em seu estudo sobre *Algorithmic Accountability Reporting*, o autor destaca os obstáculos enfrentados por profissionais que tentam investigar algoritmos com impacto público, uma vez que empresas impõem bloqueios ao acesso a dados, código e documentação. Essa obstrução deliberada inviabiliza a produção de informações compreensíveis e enfraquece o papel das mídias e das organizações sociais como instrumentos de

controle extrajurídico.

No campo da governança corporativa, Martin e Freeman (2020, p. 13-18) demonstram que a opacidade também está associada a decisões internas de governança, conformando uma ética empresarial pautada pela maximização de controle informacional e pela minimização de riscos legais e reputacionais. Os autores defendem que a responsabilidade social corporativa deve incluir a adoção voluntária de mecanismos de transparência algorítmica e a construção de canais de contestação acessíveis, como forma de responder às exigências éticas da era digital.

O AI Now Institute (2023, p. 15-20) reforça esse cenário ao mostrar que a concentração do poder computacional e da expertise técnica nas mãos de um número reduzido de empresas, como OpenAI, Google, Meta, Amazon e Microsoft, resulta na consolidação de uma infraestrutura opaca e resistente à regulação. Segundo o relatório, essas corporações não apenas controlam o desenvolvimento e a distribuição dos modelos mais avançados, como também impõem cláusulas contratuais e estruturas proprietárias que impedem a realização de auditorias externas e o acesso público a informações essenciais. Essa arquitetura institucional sustenta uma forma de opacidade não acidental, mas estruturada como mecanismo de controle e proteção do capital informacional.

Segundo Burrell (2016, p. 2-3), os algoritmos de aprendizado de máquina, especialmente os que operam em arquitetura de redes neurais profundas, tornam-se opacos não porque seu código-fonte esteja inacessível, mas porque sua lógica de operação é intrinsecamente difícil de ser rastreada e interpretada mesmo por especialistas. Trata-se de um paradoxo moderno: quanto mais precisos e eficientes se tornam os sistemas algorítmicos, mais difícil se torna explicar, com clareza e completude, o caminho causal que levou a uma determinada decisão.

Ainda que legislações como o GDPR e a LGPD prevejam direitos à explicação, à revisão de decisões automatizadas e ao acesso a informações pessoais processadas por sistemas automatizados, tais dispositivos jurídicos esbarram na baixa interpretabilidade dos modelos contemporâneos de IA. Na prática, o que se oferece ao titular de dados é uma descrição genérica do tipo de dado utilizado ou do objetivo do tratamento, o que dificilmente satisfaz os requisitos de transparência substancial e *accountability* (KAMINSKI, 2020, p. 23-24).

Assim, a opacidade técnica dos sistemas de inteligência artificial baseados em aprendizado de máquina decorre, em grande medida, da própria complexidade

estrutural desses modelos. Durante o processo de desenvolvimento, os engenheiros definem uma estrutura matemática com milhões de parâmetros ajustáveis cujo comportamento coletivo não pode ser facilmente interpretado por observação direta. A esses parâmetros se aplicam funções que orientam o ajuste automático das respostas do sistema com base na comparação entre os resultados esperados e os efetivamente produzidos. Ainda que cada componente isolado do algoritmo possa ser simples, o volume e a interação entre eles tornam o sistema como um todo virtualmente inacessível à análise compreensiva, justificando o uso da metáfora da "caixa-preta" para descrever essa forma de opacidade algorítmica inerente ao aprendizado profundo (DIGNUM, p. 23-24, 2019).

À medida que os modelos aumentam em tamanho, dados e custo computacional, seu desempenho melhora de forma previsível, mas com custos proporcionais em opacidade técnica. A relação entre escala e acurácia reforça a impenetrabilidade técnica desses modelos, que passam a funcionar como "caixas-pretas". Esse ganho técnico vem à custa da compreensibilidade humana. A maneira como os modelos ajustam bilhões de parâmetros em redes profundas os torna, na prática, irredutíveis a explicações causais claras, uma marca típica da opacidade técnica (HESTNESS *et al.,* 2017).

Para Pasquale (2015, p. 108), a opacidade técnica também pode ser considerada intencional. A arquitetura das plataformas e dos sistemas de inteligência artificial é projetada de forma a dificultar a auditabilidade, fragmentar a cadeia de responsabilização e deslocar o ônus da transparência para os indivíduos afetados, que, por sua vez, carecem dos meios técnicos e informacionais para exercer qualquer tipo de controle efetivo.

A complexidade técnica da inteligência artificial está diretamente relacionada às arquiteturas e métodos abordados anteriormente. Modelos de aprendizado profundo, como as redes neurais artificiais e os Large Language Models (LLMs), são treinados sobre grandes volumes de dados e operam por meio do ajuste de milhões ou até bilhões de parâmetros internos. Ainda que as técnicas aplicadas sejam matematicamente bem definidas, sua aplicação em larga escala, com alta densidade de camadas e parâmetros, torna o comportamento final do sistema difícil de prever e compreender. Essa estrutura técnica, voltada à eficiência e à acurácia, acaba limitando a capacidade de explicação de como uma decisão foi gerada.

Por isso, fala-se em opacidade técnica, os sistemas funcionam como

verdadeiras “caixas-pretas”, nas quais é possível observar entradas e saídas, mas não o caminho que liga uma à outra. Esse tipo de opacidade não é causado por sigilo ou má-fé, mas por limitações reais de inteligibilidade humana diante da complexidade dos modelos. Ainda que legislações como a LGPD e o GDPR garantam o direito à explicação, na prática esse direito esbarra na baixa interpretabilidade das decisões automatizadas. O que se oferece ao cidadão é, muitas vezes, uma descrição genérica dos dados utilizados, sem permitir a compreensão real da lógica decisória.

A última dimensão da opacidade é a estrutural, a qual é um problema social e político, que decorre da concentração do conhecimento técnico, e é classificado como estrutural porque está enraizado na desigualdade de nossa sociedade. Um cidadão ou até mesmo um agente público ou da sociedade civil, em geral, não possui tecnologia e conhecimento especializado necessário para compreender os modelos mais avançados de algoritmos (Burrell, 2016, p. 4-5). A opacidade estrutural deriva da incapacidade da sociedade de auditar ou fiscalizar esses algoritmos mesmo que tenham todos os elementos possíveis para isso.

Essa opacidade decorre da escala de aplicação desses algoritmos, na era do *big data*, bilhões ou trilhões de dados alimentam o aprendizado de máquina. A lógica desses sistemas é alterada conforme ele aprende com esses dados e esse número só aumenta. Essas técnicas enfrentam limites de recursos computacionais e podem aumentar sua eficiência à custa de sua transparência (BURRELL, 2016, p. 5-6).

A maior parte da sociedade, incluindo operadores do Direito, agentes públicos e membros da sociedade civil, não dispõe dos recursos tecnológicos, formativos e computacionais indispensáveis para auditar ou fiscalizar esses sistemas, mesmo que, em tese, os dados estejam disponíveis. Esse cenário se agrava com a expansão massiva da escala de aplicação dos algoritmos na era do big data. À medida que essas tecnologias são treinadas com volumes colossais e heterogêneos de dados, sua lógica interna se torna cada vez mais complexa, dificultando o rastreamento e a explicação dos critérios decisórios. Além disso, o desenvolvimento e a operação desses sistemas exigem infraestrutura computacional de alto desempenho, o que acaba concentrando sua produção e domínio em grandes centros corporativos ou institucionais. O resultado é um cenário em que o controle sobre os algoritmos torna-se inacessível para a maioria dos atores sociais, comprometendo a possibilidade de supervisão democrática e a efetividade de mecanismos de responsabilização.

Ao analisar as três dimensões da opacidade algorítmica, é possível

compreender que essa é multifacetada e possui raízes em diferentes aspectos tecnológicos, sociais e políticos, o que torna esse problema ainda mais desafiador. Compreendidas suas diferentes facetas e formas de manifestação, é necessário investigar as consequências que essa opacidade produz em nossa sociedade.

Sylvia Lu (2023, p. 5-6) aponta que a opacidade algorítmica constitui uma ameaça estrutural à efetividade dos direitos humanos, uma vez que compromete o acesso à informação sobre o tratamento de dados pessoais e impede o controle público e jurídico sobre decisões automatizadas potencialmente discriminatórias ou abusivas como decidir quem deve receber tratamento médico, quem deve conseguir um emprego ou quem deve permanecer encarcerado.

Os algoritmos também podem gerar a perpetuação de preconceitos e discriminações, pois esses atuam de forma amoral e se alimentam de dados da sociedade. Em tese, quanto mais acentuado for determinado preconceito, mais isso refletirá no padrão de aprendizado do algoritmo que tende a replicar este padrão de preconceito. Isso pode ser combatido através de técnicas de treinamentos adequadas e monitoradas, no entanto, essa prática é contraintuitiva a lógica do *big data* e da eficiência algorítmica, o que não agrada quem as controla.

A opacidade decisória dos sistemas de inteligência artificial não apenas limita a auditabilidade técnica e a compreensão dos processos internos, mas também produz efeitos psicológicos e normativos relevantes. Como observa Dignum (2019, p. 108), a percepção generalizada de que as decisões automatizadas são inacessíveis ou incontroláveis tende a gerar um sentimento de impotência perante as máquinas ao mesmo tempo em que dilui a noção de responsabilidade individual e institucional sobre os usos e impactos da IA.

Conforme apontado por Samek et al. (2019, p. 5-15), embora os modelos baseados em aprendizado profundo operem com estruturas altamente complexas e de difícil interpretação, diversas técnicas têm sido desenvolvidas com o objetivo de reduzir sua opacidade e possibilitar a explicação de suas decisões. Métodos especializados buscam identificar, com base em dados de entrada, quais elementos mais contribuíram para a saída gerada pelo modelo, permitindo certa reconstrução lógica do processo decisório. Ainda que essas abordagens não ofereçam transparência absoluta, elas representam um avanço significativo no esforço de tornar sistemas algorítmicos mais auditáveis e compreensíveis, sobretudo em contextos que exigem responsabilização e controle público.

A literatura técnica sobre explicabilidade demonstra que as técnicas de interpretação de modelos de aprendizado profundo vêm sendo aplicadas em diferentes domínios de alto impacto social, como saúde, segurança e sistemas de informação. Métodos explicativos têm sido utilizados para aprimorar a transparência em diagnósticos médicos assistidos por IA, identificar padrões relevantes em sistemas de reconhecimento facial, tornar mais compreensível a análise automatizada de imagens e vídeos, bem como esclarecer decisões tomadas por modelos de processamento de linguagem natural. Essas experiências revelam que, embora os modelos permaneçam tecnicamente complexos, é possível desenvolver abordagens que promovam algum grau de inteligibilidade funcional, especialmente quando vinculadas a contextos regulatórios ou éticos que exigem prestação de contas (SAMEK *et al.,* 2019, p. 285-379).

Contudo, a adoção efetiva dessas técnicas enfrenta barreiras significativas de natureza corporativa e política. Em muitos casos, empresas detentoras dos sistemas algorítmicos optam por não implementar mecanismos de explicabilidade robustos, seja para proteger segredos comerciais, seja para evitar a exposição de vieses estruturais e riscos jurídicos associados a decisões automatizadas. Do ponto de vista institucional, a explicabilidade plena pode comprometer estratégias comerciais, ao mesmo tempo em que demanda mudanças regulatórias e técnicas que confrontam interesses consolidados. Assim, embora tecnicamente viáveis, os métodos de explicação ainda dependem de incentivos normativos e pressões sociais para serem integrados de forma sistemática às práticas de desenvolvimento e operação de sistemas de inteligência artificial.

Diante das múltiplas dimensões da opacidade algorítmica intencional, técnica e estrutural, torna-se evidente que a dificuldade de compreensão, fiscalização e contestação das decisões automatizadas não decorre apenas de limitações tecnológicas, mas está profundamente ligada a estratégias institucionais de ocultamento, desigualdades informacionais e assimetrias de poder. A opacidade algorítmica, ao encobrir os critérios que regem decisões com impactos concretos sobre a vida dos indivíduos, gera um ambiente de incerteza que fragiliza os mecanismos tradicionais de responsabilização jurídica e democrática. No entanto, seus efeitos não se limitam ao desconhecimento sobre o funcionamento dos sistemas. Eles também produzem e aprofundam um fenômeno correlato: a assimetria informacional entre os desenvolvedores e operadores desses sistemas, de um lado,

e os cidadãos, usuários e autoridades públicas, de outro. Compreender essa assimetria é essencial para avaliar os riscos jurídicos e sociais associados ao uso da inteligência artificial na sociedade informacional.

3.2.2 Assimetria Informacional

A assimetria informacional é um conceito originado na teoria econômica e revela-se como um dos principais obstáculos à regulação eficiente em contextos de monopólio, especialmente nas indústrias de rede e setores intensivos em tecnologia. Jean Tirole (2020, p. 529) explica que as estruturas que dominam os monopólios tendem a utilizar estrategicamente o conhecimento que detêm, revelando dados apenas quando isso favorece seus interesses e ocultando-os quando a transparência ameaça sua rentabilidade. Nesse cenário, surge um desafio para regulação eficiente desses mercados.

Esse conceito adquire nova dimensão no contexto da inteligência artificial onde não apenas compromete a eficiência econômica e a concorrência leal, mas também fragiliza o controle democrático e o exercício de direitos fundamentais como a liberdade, privacidade e proteção de dados.

Ao tratar sobre a concentração da tecnologia e do conhecimento técnico pelas *big techs* de inteligência artificial, Hestness et al. (2017, p. 5) demonstram que o avanço no desempenho de modelos de aprendizado profundo está acompanhado de um aumento exponencial na complexidade dos sistemas, o que implica que poucos laboratórios corporativos, com acesso privilegiado a infraestrutura e expertise especializada, conseguem desenvolver, treinar e compreender plenamente tais modelos. Isso ocasiona a assimetria informacional neste setor uma vez que a compreensão efetiva do funcionamento algorítmico passa a depender de recursos institucionais e cognitivos altamente concentrados, tornando tais sistemas inacessíveis à fiscalização por atores externos, como reguladores, juristas ou a sociedade civil.

O relatório *AI Now 2023 Landscape Report* trouxe duas seções relacionadas ao assunto. Em *Too Big to Fail* (AI Now, 2025, p. 24), o relatório destaca que o atual ecossistema de inteligência artificial é marcado pela extrema concentração de poder técnico, econômico e político em torno de poucas empresas de tecnologia. Essa

centralização, sustentada por vantagens acumuladas em infraestrutura, dados e parcerias institucionais, torna tais corporações praticamente inatingíveis por mecanismos tradicionais de regulação. Essa condição de "grandes demais para falhar" não apenas compromete a competitividade do setor, como também reduz a capacidade de fiscalização externa sobre o funcionamento de seus sistemas algorítmicos. Em *AI Arms Race 2.0* (AI Now, 2025, p. 28) expõe que a corrida geopolítica e corporativa pela supremacia em inteligência artificial tem aprofundado a concentração de recursos estratégicos, como dados, poder computacional e talentos especializados, em mãos de poucos atores privados. Essa lógica de competição intensificada favorece um modelo de desenvolvimento tecnológico orientado à maximização de escala e lucro, em detrimento de valores democráticos e direitos fundamentais. O resultado é a consolidação de uma estrutura assimétrica de informação e poder.

A assimetria informacional constitui um desdobramento direto e estrutural da opacidade algorítmica, na medida em que aprofunda o desequilíbrio entre os agentes que desenvolvem, operam e controlam sistemas de inteligência artificial e aqueles que são por eles impactados. Se a opacidade torna os sistemas indecifráveis ou inacessíveis do ponto de vista técnico, jurídico ou institucional, a assimetria informacional expressa a consequência desse cenário com a concentração de conhecimento, dados e capacidade interpretativa nas mãos de poucos, em detrimento da ampla maioria dos atores sociais. Em contextos nos quais algoritmos operam com alto grau de complexidade e ausência de explicabilidade, a assimetria informacional limita o exercício efetivo de direitos, tornando o controle social e jurídico uma promessa formalmente reconhecida, mas materialmente inviável.

Embora relacionada à opacidade algorítmica estrutural, a assimetria informacional constitui um fenômeno conceitualmente distinto e mais abrangente. A opacidade estrutural, nos termos propostos por Jenna Burrell, refere-se à incapacidade técnica e cognitiva de compreender os sistemas algorítmicos, decorrente da complexidade dos modelos e da concentração de expertise. A assimetria informacional, por sua vez, é um fenômeno político e social decorrente dessa opacidade algorítmica que produz ou aprofunda desequilíbrios de poder e conhecimento, sendo essa a base do capitalismo de vigilância:

> O capitalismo de vigilância age por meio de assimetrias nunca antes vistas

> referentes ao conhecimento e ao poder que dele resulta. Ele sabe tudo sobre nós, ao passo que suas operações são programadas para não serem conhecidas por nós. Elas acumulam vastos domínios de um conhecimento novo proveniente de nós, mas que não é para nós. Elas predizem nosso futuro a fim de gerar ganhos para os outros, não para nós. Enquanto o capitalismo de vigilância e seus mercados futuros comportamentais tiverem permissão de prosperar, a propriedade desses novos meios de modificação comportamental irá ofuscar a propriedade dos meios de produção como o manancial da riqueza e do poder capitalistas no século XXI (ZUBOFF, 2021, p. 26)

Para Zuboff (2021, p. 35) esse nível de concentração de poder e conhecimento não possui precedentes na história humana. Nesse cenário, essa se estrutura como uma assimetria epistêmica, pois as grandes corporações não apenas sabem mais, mas definem os próprios parâmetros do que pode ser conhecido. A inteligência artificial, ao agravar essa assimetria, consolida esse novo modelo de dominação, a assimetria informacional, nesse sentido, passa a ser não apenas um sintoma, mas o próprio motor do capitalismo de vigilância.

Para Potters e van Winden (1992, p. 271-272), a assimetria informacional, no contexto da formulação de políticas públicas, consiste na vantagem estratégica detida por grupos de interesse que possuem informações técnicas ou específicas não acessíveis aos decisores públicos. Essa assimetria é explorada por meio do *lobbying* informacional, que não depende apenas do conteúdo transmitido, mas das características institucionais, dos incentivos e da credibilidade percebida desses grupos. A atuação desses agentes, munidos de expertise seletiva, pode assim distorcer a racionalidade do processo decisório, restringindo a neutralidade técnica e o interesse público na elaboração normativa.

A análise das múltiplas formas de opacidade e de seus desdobramentos em assimetria informacional revela que os sistemas de inteligência artificial não apenas desafiam a auditabilidade técnica e a compreensão social, mas também instauram novos regimes de poder baseados no controle exclusivo da informação. A invisibilidade dos critérios decisórios, somada à concentração de infraestrutura tecnológica e expertise em poucos polos econômicos e governamentais, compromete a efetividade de direitos fundamentais e fragiliza os mecanismos clássicos de responsabilização. Esse quadro projeta um cenário no qual a desigualdade informacional converte-se em desigualdade jurídica, tornando incerta a própria capacidade do direito em assegurar proteção adequada aos indivíduos diante de decisões automatizadas que impactam diretamente sua esfera de liberdade e

privacidade.

Diante disso, torna-se imperioso reconhecer que a opacidade algorítmica e a assimetria informacional não constituem apenas desafios técnicos, mas verdadeiros dilemas jurídicos e políticos da sociedade informacional. Ao mesmo tempo em que tais fenômenos evidenciam os limites dos instrumentos regulatórios tradicionais, também abrem espaço para a centralidade de princípios estruturantes capazes de reequilibrar as relações entre indivíduos, empresas e Estados. É nesse contexto que a seguir se dará o exame da transparência e da *accountability*, a qual se compreende dentro da hipótese estabelecida como fundamentos normativos indispensáveis à tutela jurídica da proteção de dados pessoais na era da inteligência artificial, oferecendo alternativas para mitigar a opacidade e reduzir as assimetrias que comprometem a legitimidade democrática e a segurança jurídica.

3.3 Princípios Para o Enfrentamento Dos Desafios Informacionais

A análise realizada até aqui permite compreender que a inteligência artificial não apenas aprofunda, mas transforma qualitativamente os desafios característicos da sociedade informacional. A combinação entre opacidade algorítmica e assimetria informacional produz um ambiente decisório em que titulares de dados, consumidores e mesmo autoridades regulatórias possuem reduzida capacidade de escrutínio, contestação e controle dos sistemas automatizados. O resultado é a consolidação de estruturas de poder informacional que fragilizam a autonomia individual e ampliam a dependência de agentes públicos e privados que detêm os recursos técnicos e econômicos necessários para operar tais tecnologias. Diante desse cenário, torna-se evidente que a tutela jurídica tradicional, centrada apenas em proibições, consentimentos ou deveres formais, revela-se insuficiente para responder aos desafios substantivos da IA.

É nesse ponto que emergem os princípios como eixo normativo indispensável para orientar a regulação da inteligência artificial e reequilibrar as relações informacionais. A atuação do direito, portanto, só pode ser eficaz se apoiada em princípios estruturantes capazes de orientar interpretações, fundamentar deveres, preencher lacunas regulatórias e guiar a atuação das instituições. Esses princípios assumem função constitutiva: não apenas orientam a aplicação das normas

existentes, mas indicam a direção para a construção de instrumentos técnicos necessários para enfrentar riscos que não podem ser eliminados, mas apenas mitigados.

Nesse sentido, dois princípios se destacam como pilares normativos aptos a enfrentar os desafios informacionais analisados: a transparência algorítmica e a *accountability*. Ambos respondem diretamente à dupla estrutura de risco apresentada no item anterior. A transparência atua sobre o déficit de inteligibilidade, permitindo que decisões automatizadas deixem de operar como caixas-pretas impermeáveis ao controle público. A *accountability*, por sua vez, enfrenta a assimetria informacional ao criar mecanismos efetivos de atribuição de responsabilidades, supervisão e sanção. Juntos, esses princípios formam o núcleo normativo capaz de permitir que a sociedade informacional seja governada com base em critérios de justiça, autonomia e proteção de direitos fundamentais. É a partir dessa fundamentação que se desenvolvem, nos itens seguintes, as dimensões teóricas, técnicas e normativas da transparência e da *accountability* aplicadas à inteligência artificial.

3.3.1 Transparência Algorítmica

O ideal de transparência em sua forma moderna surge historicamente no contexto do Iluminismo, associado ao projeto de racionalidade e ao anseio por verdades universais, com uma lógica epistemológica que vincula observação e verdade, pressupondo que quanto mais fatos são revelados, maior a possibilidade de conhecer a verdade. Nesse período, práticas jurídicas e administrativas começaram a institucionalizar a transparência como forma de governança, seja pela garantia do acesso a documentos públicos, seja pela criação de mecanismos de visibilidade voltados à disciplina social. A partir do século XX, esse ideal foi reforçado por legislações de acesso à informação e por políticas regulatórias de divulgação obrigatória, transformando a transparência em instrumento central de *accountability*, eficiência e legitimidade institucional. A compreensão moderna da transparência pode ser compreendida como valor público voltado a combater a corrupção, como mecanismo de abertura dos processos decisórios estatais e como ferramenta de governança destinada a promover maior eficiência administrativa (ANANNY; CRAWFORD, 2018, p. 974-975).

Esse ideal iluminista de transparência, projeta-se hoje no debate sobre a transparência algorítmica. A lógica permanece semelhante: parte-se da ideia de que abrir códigos, mecanismos de design ou mesmo a lógica interna do software seria suficiente para compreender e, assim, governar sistemas computacionais complexos. Pesquisadores em ciência da computação têm buscado, desde os anos 1960, desenvolver técnicas de visualização e explicação que tornem inteligíveis as decisões algorítmicas, evoluindo para estudos contemporâneos que associam a transparência à prevenção de discriminações e à construção de mecanismos de governança. Essa perspectiva, porém, ainda se ancora no pressuposto de que ver equivale a saber, e de que a observação dos processos internos da máquina pode, por si só, garantir *accountability* e legitimidade. Porém, o pressuposto de que "ver é saber" é limitado e pode ser enganoso, pois a simples abertura de códigos ou a disponibilização de visualizações técnicas não garante compreensão efetiva, sobretudo para públicos não especializados. Além disso, mesmo engenheiros podem não conseguir explicar integralmente sistemas de aprendizado profundo, o que demonstra que a complexidade técnica escapa à lógica clássica da transparência. Os autores enfatizam que essa confiança excessiva na visibilidade como sinônimo de conhecimento reduz a *accountability* a um processo meramente informacional, quando, na verdade, a responsabilização exige considerar relações de poder, contextos sociais e impactos materiais que vão muito além da possibilidade de observar o interior da "caixa-preta". (ANANNY; CRAWFORD, 2018, p. 976-977).

Essa trajetória evidencia que a transparência, desde sua gênese iluminista, sempre esteve associada à ideia de controle social e de legitimação institucional por meio da visibilidade. No entanto, à medida que a sociedade passou a depender cada vez mais de sistemas técnicos e informacionais, o conceito deixou de se restringir ao acesso a documentos estatais ou à publicidade de atos administrativos, sendo progressivamente transposto para o domínio tecnológico. Essa transposição mostra que, assim como antes se buscava vigiar o funcionamento do Estado e das instituições públicas, hoje emerge a necessidade de abrir à observação e ao escrutínio os processos decisórios mediados por algoritmos, controlados por grandes corporações, de modo a enfrentar a opacidade que caracteriza os sistemas de inteligência artificial.

Por essa razão, Luciano Floridi (2009, p. 107-109) sustenta que a transparência deve ser entendida como uma condição pró-ética, ou seja, um meio destinado a tornar efetivos outros princípios normativos e não como um valor absoluto em si mesma. Seu

real valor ético reside na capacidade de possibilitar que fundamentos como a responsabilidade, a segurança, a privacidade e o consentimento informado se concretizem de maneira adequada. Para tanto, o autor enfatiza que não basta que organizações e sistemas digitais simplesmente publiquem informações, é indispensável que revelem como essas informações foram produzidas, processadas e estruturadas, bem como quais valores e critérios éticos orientaram esse processo. A transparência só cumpre seu papel ético quando acompanhada da disponibilização de informação verdadeira, significativa e contextualizada, distinta da simples exposição de dados brutos, sob pena de gerar distorções e comprometer sua finalidade. Assim, no contexto da sociedade digital, a transparência só alcança efetividade quando articulada ao conceito de ethics by design, isto é, a incorporação estrutural da ética no funcionamento tecnológico, assegurando que os sistemas de informação sejam capazes de promover *accountability*, confiança social e legitimidade democrática.

Ao destacar a transparência como condição pró-ética, Floridi chama atenção para o risco de tratá-la como um fim autossuficiente. Essa advertência é especialmente relevante no campo da inteligência artificial, em que a simples abertura de informações técnicas não garante, por si só, maior proteção aos indivíduos. A experiência regulatória e o debate acadêmico revelam que a transparência só cumpre sua função quando articulada a finalidades normativas concretas, como a proteção da privacidade, a prevenção de abusos e a criação de mecanismos eficazes de responsabilização. Nesse sentido, a teoria de Floridi pode ser lida como um ponto de partida que ajuda a deslocar o foco do “quanto” de informação é divulgado para o “como” e “para quê” essa informação é disponibilizada, perspectiva que se conecta diretamente às preocupações práticas levantadas por Cheong sobre legitimidade democrática e contestação social.

Para Cheong, no campo da IA, a transparência deve ser entendida como um requisito estruturante para a proteção do bem-estar individual e coletivo na sociedade contemporânea. Não se trata apenas de um ideal normativo, mas de uma condição prática que garante que os cidadãos, as autoridades regulatórias e a sociedade em geral possam compreender de que modo os sistemas algorítmicos processam informações e produzem decisões que afetam diretamente a vida das pessoas. Ao assegurar essa visibilidade, a transparência confere aos indivíduos não apenas a possibilidade de conhecer os fundamentos de tais decisões, mas também de contestá-

las e exigir justificativas consistentes, instituindo mecanismos de correção e de responsabilização. Dessa forma, o princípio reforça a legitimidade democrática das tecnologias de IA, funcionando como contrapeso à opacidade algorítmica e como instrumento de preservação da autonomia e da confiança social (CHEONG, 2024, p. 1-2).

Como princípio jurídico, o princípio da transparência possui raízes históricas na proteção de dados pessoais, constituindo-se como um dos pilares normativos desde as primeiras legislações internacionais sobre o tema. Conforme Danilo Doneda (2022, p. 171), esse princípio era o primeiro a ser listado no chamado “núcleo comum” das legislações de proteção de dados, a transparência se manifestava como exigência de publicidade acerca da existência de bancos de dados pessoais, seja mediante autorização prévia de funcionamento, seja por meio de notificações a autoridades competentes ou da divulgação periódica de relatórios.

A presença da transparência nos primeiros marcos regulatórios de proteção de dados mostra que esse princípio foi pensado desde o início como condição mínima para equilibrar assimetrias informacionais entre cidadãos e instituições. Se antes bastava exigir publicidade sobre a existência de bancos de dados, a crescente sofisticação tecnológica ampliou a demanda por instrumentos mais robustos de abertura, capazes de revelar não apenas a coleta, mas também a lógica de uso e processamento dos dados.

Assim, o princípio da transparência consolidou-se como eixo normativo fundamental nas legislações contemporâneas de proteção de dados e de regulação da inteligência artificial. No âmbito europeu, o GDPR (União Europeia, 2016) estabelece, já em seu artigo 5º, que os dados pessoais devem ser tratados de forma lícita, leal e transparente em relação ao titular, inserindo a transparência como um dos pilares do tratamento de dados. No Brasil, a LGPD (Brasil, 2018) positivou a transparência como princípio específico (art. 6º, VI), assegurando aos titulares o direito de receber informações claras, precisas e acessíveis sobre o tratamento realizado e os respectivos agentes envolvidos, ainda que resguardados segredos comerciais e industriais.

Nas legislações trazidas sobre proteção de dados, a transparência algorítmica já tem recebido atenção. O Regulamento Geral de Proteção de Dados da União Europeia (GDPR) prevê o chamado “direito à explicação” que se refere à possibilidade de um indivíduo compreender os fundamentos de uma decisão que lhe afete de forma

significativa quando ela é tomada exclusivamente por meios automatizados, como ocorre em sistemas de inteligência artificial. A origem desse debate está no artigo 22 do GDPR, que prevê que toda pessoa tem o direito de não ficar sujeita a decisões baseadas unicamente em tratamento automatizado que produzam efeitos legais ou afetem de modo relevante a sua esfera jurídica. Além disso, os artigos 13, 14 e 15 estabelecem deveres de informação por parte do controlador, impondo que sejam fornecidas “informações significativas sobre a lógica envolvida” no tratamento, bem como sobre a importância e as consequências previstas desse processamento (União Europeia, 2016).

Porém, a efetividade deste direito vem sendo questionada. Tanto Edwards e Veale (2017) quanto Wachter, Mittelstadt e Floridi (2017) convergem na crítica à leitura segundo a qual o Regulamento Geral de Proteção de Dados teria criado um direito robusto à explicação de decisões automatizadas. Os primeiros ressaltam que, mesmo que tal direito fosse reconhecido, sua efetividade prática seria reduzida diante de barreiras técnicas, da complexidade dos modelos de aprendizado de máquina e de resistências corporativas pautadas em segredos comerciais. Já Wachter, Mittelstadt e Floridi demonstram, em análise dogmática, que o GDPR não prevê um direito pleno à explicação, mas apenas um direito limitado à informação sobre a lógica geral dos sistemas, aplicável em hipóteses restritas e com redação ambígua. Ambos os estudos, portanto, caminham no mesmo sentido: a transparência regulatória, tal como delineada no GDPR, revela-se insuficiente para assegurar *accountability* efetiva frente à opacidade algorítmica.

Wachter, Mittelstadt e Floridi (2017, p. 78-79) concluem que para que a norma pudesse assegurar *accountability* efetiva diante da opacidade algorítmica, os autores defendem que seria necessário reforçar o artigo 22 com a inclusão expressa de um direito de explicação, esclarecer termos vagos como “lógica envolvida” e “consequências previstas”, eliminar exceções que reduzem o alcance da regra, como a exclusão de decisões que não sejam “exclusivamente automatizadas” e, sobretudo, criar mecanismos externos de auditoria de algoritmos, de modo a equilibrar a exigência de transparência com a proteção de segredos comerciais.

Selbst e Powles (2017) apresentam uma leitura diferente daquela sustentada pelos autores citados acima. Enquanto esses últimos afirmam que o GDPR não prevê um direito robusto à explicação, mas apenas deveres informacionais genéricos, Selbst e Powles entendem que os artigos 13, 14 e 15 já estabelecem, sim, uma base

normativa para tal direito. Para eles, a exigência de fornecer “informações significativas sobre a lógica envolvida” deve ser interpretada de forma funcional e prática: a explicação deve permitir que o titular compreenda como seus dados foram tratados, possa contestar a decisão automatizada e identifique eventuais discriminações. Nessa perspectiva, o GDPR contém um direito à explicação implícito, que, embora não esteja redigido de forma expressa, deve ser aplicado de maneira flexível e orientada à efetividade (SELBST; POWLES, 2017, p. 236-238).

As divergências interpretativas em torno do GDPR evidenciam tanto o avanço quanto as insuficiências da regulação europeia. Se, por um lado, o regulamento colocou a transparência no centro da proteção de dados e abriu espaço para discutir o direito à explicação, por outro, mostrou-se limitado diante da complexidade técnica e dos interesses econômicos que permeiam os sistemas algorítmicos. Esse cenário impulsionou a formulação de novos marcos regulatórios voltados especificamente à inteligência artificial, em que a transparência deixa de ser apenas um subproduto da proteção de dados pessoais e passa a constituir dever jurídico próprio, com obrigações mais detalhadas nas regulações de inteligência artificial.

O AI Act europeu, aprovado em 2024, estabelece obrigações específicas para sistemas de alto risco, impondo a divulgação de informações sobre dados de treinamento, arquitetura de modelos e métricas de desempenho. Já nos Estados Unidos, a proposta de Algorithmic *Accountability* Act de 2023 avança no sentido de tornar obrigatórios relatórios de impacto em setores sensíveis, reforçando o esforço de criar instrumentos legais que promovam maior transparência no uso de sistemas algorítmicos (CHEONG, 2024, p. 3-4).

A transparência, portanto, passa ocupar posição central na regulação da inteligência artificial. O AI Act europeu (União Europeia, 2024) menciona o termo em mais de quarenta passagens, prevendo desde regras harmonizadas de transparência (art. 1º, d) até obrigações específicas para sistemas de alto risco, que devem ser desenvolvidos de modo a permitir a interpretação adequada de seus resultados (art. 13), bem como a obrigação de informar indivíduos quando interagem com sistemas de IA (art. 50). Em paralelo, a proposta de regulação brasileira de IA, através do Projeto de Lei nº 2.338 de 2023, também incorpora expressamente a transparência como princípio, associando-a à explicabilidade, inteligibilidade e auditabilidade (art. 3º, VI) (Senado Federal do Brasil, 2023). Em conjunto, esses dispositivos demonstram que a transparência deixou de ser apenas um ideal ético e se consolidou como um

dever jurídico transversal, presente tanto na proteção de dados pessoais quanto na disciplina normativa da inteligência artificial.

Essas mudanças se devem ao fato de que a responsabilidade na inteligência artificial somente pode ser efetivada se acompanhada de transparência. Para que se possa identificar responsáveis e avaliar a legitimidade das decisões automatizadas, é necessário compreender de que modo os sistemas foram projetados, quais critérios nortearam sua construção e quais escolhas foram feitas no decorrer de seu desenvolvimento. Em sua visão, a transparência constitui, portanto, uma condição indispensável para que haja *accountability* real no campo da IA (DIGNUM, 2020, p. 218-220).

Diakopoulos distingue quatro dimensões fundamentais da transparência em sistemas de inteligência artificial: a primeira refere-se aos dados, enfatizando a necessidade de clareza quanto à sua origem, qualidade e representatividade; a segunda incide sobre os modelos, destacando a importância de compreender a lógica e os parâmetros que estruturam o funcionamento algorítmico; a terceira envolve as decisões, apontando para a obrigação de oferecer explicações compreensíveis sobre os resultados produzidos; e, por fim, a quarta se relaciona à governança processual, abrangendo normas institucionais, políticas de auditoria e mecanismos de responsabilização que garantem legitimidade ao uso da tecnologia (DIAKOPOULOS, 2020, p. 199-202).

De modo similar, ao tratar da transparência, Virginia Dignum destaca que não basta verificar apenas os resultados fornecidos pelos sistemas de inteligência artificial. É igualmente essencial abrir o processo que os antecede, revelando como foram escolhidos os dados de treinamento, definidos os objetivos a serem alcançados e ajustados os parâmetros que guiam o desempenho algorítmico. Essa ênfase no percurso de design e implementação evidencia que a transparência deve alcançar toda a cadeia de desenvolvimento, funcionando como instrumento de responsabilidade (DIGNUM, 2020, p. 221-223).

A transparência cumpre múltiplas funções: de um lado, permite o escrutínio público e fortalece a *accountability* democrática, de outro, possui uma dimensão ética, ao possibilitar que os indivíduos compreendam como são afetados por decisões automatizadas. Além disso, exerce papel instrumental ao fomentar a confiança, reduzir vieses e aprimorar a qualidade dos sistemas de IA (Diakopoulos, 2020, p. 202-205).

As funções atribuídas à transparência revelam seu potencial de fortalecer a *accountability* democrática, fomentar confiança e aprimorar a qualidade técnica dos sistemas de inteligência artificial. Contudo, assumir tais funções de forma acrítica pode levar a uma idealização excessiva, como se a simples abertura de informações fosse suficiente para garantir justiça e legitimidade. A literatura mostra que, paralelamente a seus benefícios, a transparência está sujeita a importantes restrições técnicas, jurídicas e sociais. Reconhecer essas limitações é essencial para evitar que o princípio seja tratado como solução universal e para compreender que sua efetividade depende de ser calibrada conforme o contexto, os públicos envolvidos e os riscos de cada aplicação.

Diakopoulos (2020, p. 206-209) adverte que a transparência apresenta limitações importantes no contexto da inteligência artificial. Em alguns casos, pode gerar sobrecarga de informações sem efetiva compreensão pelos cidadãos, além de expor segredos comerciais e direitos de propriedade intelectual. O autor ressalta ainda que a simples disponibilização do código-fonte não garante explicabilidade e que a transparência, quando mal calibrada, pode beneficiar apenas grupos especializados, ampliando desigualdades sociais e informacionais.

Essas limitações também são reconhecidas por Dignum (2020, p. 227-229), que afirma que a transparência não pode ser entendida como absoluta. A abertura irrestrita de informações pode colocar em risco segredos industriais e até comprometer a segurança de determinados sistemas. Por essa razão, a autora defende que a transparência deve ser pensada de forma contextualizada, calibrada de acordo com o tipo de tecnologia em questão e com o público ao qual as informações se destinam, de modo a equilibrar o dever de abertura com a proteção de interesses legítimos.

Ananny e Crawford (2018, p. 977-982) sistematizam dez limites que demonstram por que a transparência, embora relevante, não pode ser tomada como solução suficiente para a *accountability*. Primeiro, apontam a desconexão do poder, pois tornar algo visível não assegura mudança efetiva, já que práticas como a corrupção podem persistir mesmo quando expostas. Segundo, indicam que a transparência pode causar danos, sobretudo ao ameaçar a privacidade e a segurança de grupos vulneráveis. Terceiro, pode também ocultar intencionalmente, quando a divulgação excessiva de dados serve mais para confundir do que para esclarecer. Quarto, cria falsos binarismos, ao reduzir o debate à oposição entre segredo absoluto

e abertura total. Quinto, reforça um modelo neoliberal de agência, que transfere ao indivíduo a responsabilidade de monitorar sistemas complexos. Sexto, não gera necessariamente confiança, já que a abertura pode inclusive alimentar cinismo social. Sétimo, está sujeita a limites profissionais, pois especialistas acabam por filtrar e controlar o que pode ser visto. Oitavo, evidencia que ver não é compreender, dado que abrir o código não significa captar seus efeitos sociais e políticos. Nono, sofre de limitações técnicas, porque nem mesmo engenheiros compreendem integralmente sistemas avançados como redes neurais profundas. E, por fim, apresenta limitações temporais, já que a dinâmica mutável dos sistemas impede que uma fotografia estática represente sua operação ao longo do tempo.

O reconhecimento dessas limitações demonstra que a transparência, embora imprescindível, não pode ser tratada como resposta única aos desafios da opacidade algorítmica. Ao contrário, ela precisa ser acompanhada de instrumentos capazes de transformar a abertura de informações em efetiva compreensão e em condições reais de controle social. Nesse sentido, a literatura mais recente tem se concentrado em desenvolver estratégias técnicas e normativas que buscam superar os riscos da transparência meramente declaratória, oferecendo mecanismos concretos de explicabilidade, auditoria e responsabilização.

Cheong (2024, p. 7-8), em conclusão ao seu estudo, sustenta que a transparência não pode ser entendida de forma isolada, mas como um verdadeiro ecossistema multifacetado. Em sua dimensão técnica, envolve instrumentos como técnicas de inteligência artificial explicável, auditorias independentes e registros de sistemas que possibilitam rastreabilidade. No campo jurídico, conecta-se ao direito à explicação previsto no GDPR, às obrigações específicas do AI Act e a mecanismos de responsabilização legal. Já na dimensão social, depende da oferta de explicações claras e acessíveis, da participação democrática nos processos de governança e da construção de confiança pública. A principal mensagem do autor é que a transparência, por si só, não é suficiente: precisa ser traduzida em mecanismos concretos, ajustada às necessidades do público destinatário e acompanhada de formas efetivas de *accountability* (CHEONG, 2024, p. 7-8).

Ananny e Crawford (2018, p. 984-985) sustentam que a transparência não deve ser compreendida como uma panaceia capaz de, sozinha, assegurar *accountability* em sistemas complexos. Para os autores, o ideal de “ver para saber” é insuficiente diante da natureza dinâmica, técnica e política dos algoritmos, pois a visibilidade por

si só não garante compreensão nem mudança social. A efetiva responsabilização exige ampliar a visão de transparência para enxergar não só o interior dos sistemas, mas também seus efeitos, contextos e práticas de uso. Nesse sentido, a transparência deve ser vista como um instrumento entre outros, que precisa ser complementado por auditorias independentes, regulação institucional, engajamento democrático e mecanismos de correção que superem a mera exposição de informações. Assim, a transparência deve ser integrada a outras formas de governança que considerem tais dimensões, sob pena de permanecer como um ideal limitado e, muitas vezes, ineficaz.

Burrell (2016, p. 10-12), em seu estudo sobre a opacidade algorítmica, conclui que a superação dessa não pode ser buscada em soluções únicas, mas sim em estratégias múltiplas e complementares. Entre elas, a autora aponta a necessidade de regulação e auditorias externas, de políticas de educação e literacia em programação, do incentivo ao uso de modelos mais transparentes sempre que viável e da inclusão das vozes sociais diretamente afetadas para identificar discriminações. Ainda que a transparência total seja inatingível, tais medidas são capazes de reduzir riscos de injustiça e ampliar a *accountability* no uso de sistemas baseados em aprendizado de máquina.

O exame da literatura e da legislação evidencia que a transparência percorreu uma trajetória histórica que vai do ideal iluminista de visibilidade e racionalidade, passando pela sua positivação em legislações de proteção de dados, até chegar ao debate contemporâneo sobre a governança da inteligência artificial. Os autores trazidos acima apontam que ela se manifesta em múltiplas dimensões, técnica, jurídica e social, e cumpre funções essenciais como fortalecer a confiança, viabilizar o escrutínio público e reduzir assimetrias informacionais. Ao mesmo tempo, destacam-se limites significativos, seja pela complexidade técnica dos modelos de aprendizado profundo, pela sobrecarga de informações que não geram efetiva compreensão, ou pelo risco de violação de segredos industriais e da privacidade.

À luz do percurso analisado, é possível afirmar que a transparência não deve ser compreendida como um valor absoluto ou uma solução universal, mas como um fundamento instrumental da governança algorítmica. Sua relevância não reside na mera abertura de códigos ou na divulgação de relatórios técnicos, mas na capacidade de tornar as decisões compreensíveis, contestáveis e corrigíveis, permitindo a efetiva atribuição de responsabilidades. Ao final, mais do que satisfazer um ideal de visibilidade, o princípio deve servir como alicerce para a construção de confiança

social e para o fortalecimento da legitimidade democrática frente ao poder crescente da inteligência artificial.

O conjunto dessas contribuições demonstra que a transparência se consolidou como princípio normativo central, mas que sua efetividade depende da articulação com outros instrumentos de governança e com a exigência de *accountability*.

Portanto, está claro que a transparência não deve ser entendida como um fim em si mesma, mas como um fundamento para a *accountability*. Ao tornar visíveis os processos de coleta, tratamento e decisão, a transparência abre caminho para a identificação de responsabilidades e para a criação de mecanismos de correção. Nesse sentido, ela deve ser compreendida como o fundamento que sustenta a *accountability*, como dever de prestação de contas e de responsabilização efetiva dos sistemas algorítmicos, pois só a partir da abertura e inteligibilidade dos sistemas é possível cobrar efetivamente quem responde por seus impactos (COVARRUBIAS, 2021, p. 1267-1275).

Assim, a transparência deve ser entendida não como ponto de chegada, mas como condição prévia para a efetividade da responsabilização. É justamente nesse ponto que se insere o debate sobre a *accountability*, que representa a dimensão prática e normativa da prestação de contas e da atribuição de responsabilidades no uso da inteligência artificial que será discutido no item 4.2. Se a transparência constitui o fundamento indispensável para a *accountability*, sua efetividade depende de instrumentos capazes de operacionalizá-la, tema que será tratado no item 4.3

3.3.2 *Accountability* Algorítmica

A noção de *accountability* possui raízes históricas que remontam ao período da conquista normanda da Inglaterra. Inicialmente vinculada à ideia de contabilidade, o termo surgiu como prática administrativa destinada ao controle patrimonial do soberano, notadamente por meio dos Domesday Books de 1085, nos quais os detentores de propriedades eram obrigados a declarar e justificar a extensão de seus bens ao rei. Essa prática de auditoria periódica, que se associava a um juramento de fidelidade, representava um mecanismo embrionário de prestação de contas na esfera estatal. Com o passar dos séculos, o conceito desprendeu-se de sua conotação

estritamente contábil e financeira, assumindo progressivamente um caráter político, marcado pela inversão da relação originária: não mais os súditos prestando contas ao monarca, mas as autoridades públicas sendo responsabilizadas perante os cidadãos (BOVENS, 2007, p. 447).

Esse processo de transformação intensificou-se a partir do final do século XX, quando, sobretudo no contexto anglo-americano, a *accountability* foi ampliada como princípio estruturante da gestão pública. Sob a influência da New Public Management no Reino Unido e das reformas administrativas promovidas nos Estados Unidos, o termo consolidou-se como categoria normativa de governança, passando a exprimir não apenas mecanismos de controle financeiro, mas também exigências de legitimidade democrática, integridade administrativa e confiança institucional. Nesse sentido, a *accountability* deixou de ser concebida apenas como técnica de auditoria para se afirmar como verdadeiro ícone da boa governança, articulando-se com valores como transparência, justiça e responsabilidade na condução da política pública (BOVENS, 2007, p. 448).

Porém, a *accountability* apresenta múltiplas tipologias, que revelam a amplitude do conceito para além da esfera estritamente governamental. É possível identificá-la em sua dimensão política, quando autoridades eleitas prestam contas perante cidadãos e parlamentos; em sua dimensão jurídica, quando indivíduos e organizações respondem perante tribunais; em sua vertente administrativa, marcada pela necessidade de gestores justificarem suas escolhas diante de órgãos de controle e auditoria; e em sua dimensão social, na qual empresas, associações e demais instituições privadas passam a ser cobradas pela opinião pública e pela sociedade civil quanto à legalidade, à ética e aos impactos de suas atividades. Essa classificação evidencia que a *accountability*, enquanto princípio normativo, não se restringe à gestão pública, mas alcança também atores privados e corporativos, cuja atuação está cada vez mais inserida em contextos de forte assimetria informacional e de riscos coletivos, demandando mecanismos de responsabilização transparentes e efetivos (BOVENS, 1998, p. 24-28).

No contexto desta pesquisa, o enfoque recai sobre a *accountability* aplicada aos algoritmos, categoria que herda essa multiplicidade de dimensões e a projeta para o cenário contemporâneo da inteligência artificial. A responsabilização algorítmica manifesta-se tanto na esfera pública, quando governos utilizam sistemas automatizados para governar, fiscalizar ou controlar populações, quanto na esfera

corporativa, em que empresas desenvolvem e exploram tecnologias capazes de impactar diretamente direitos fundamentais, relações de consumo e dinâmicas sociais mais amplas. Trata-se de um campo no qual a exigência de prestação de contas ultrapassa a tradicional fronteira da administração pública, alcançando também os agentes privados que operam com estruturas tecnológicas de alto poder decisório e informacional.

Compreendida sua evolução histórica, é importante delinear qual é o conceito de *accountability*. De acordo com Richard Mulgan, a concepção clássica de *accountability* deve ser compreendida como uma relação de prestação de contas externa, composta por três elementos fundamentais: (i) a existência de uma autoridade externa perante a qual se deve responder; (ii) a configuração de uma interação social, que envolve questionamentos, exigências de justificações e possibilidade de resposta; e (iii) a presença de uma instância dotada de autoridade, capaz de avaliar a conduta e impor consequências (MULGAN, 2000, p. 555-556).

À luz dessa concepção, evidencia-se que a efetividade da *accountability* algorítmica depende da existência de uma autoridade competente, capaz de exercer, de forma legítima e técnica, o papel de fórum de responsabilização. Sem um ente institucional que detenha poderes de fiscalização, avaliação e aplicação de sanções, a exigência de prestação de contas tende a se esvaziar em mera retórica.

A literatura contemporânea evidencia, contudo, que o conceito de *accountability* sofreu uma expansão significativa, extrapolando sua concepção clássica de prestação de contas perante uma autoridade externa. Como aponta Mulgan (2000, p. 556-557), o termo passou a ser empregado em sentidos mais amplos, incluindo a ideia de ética pessoal, de responsividade às demandas sociais e até mesmo de deliberação democrática. Esse processo de dilatação semântica demonstra a centralidade normativa da *accountability* no discurso jurídico e político, mas também revela o risco de perda de precisão analítica, caso todas essas dimensões sejam confundidas sob um único rótulo.

Neste movimento de ampliação conceitual e de aplicação da *accountability*, essa foi incorporada como princípio estruturante no campo da proteção de dados pessoais. A sua primeira aparição em instrumentos normativos se deu a partir das Diretrizes da OCDE de 1980, que já previam a responsabilidade dos controladores não apenas pelo cumprimento das normas, mas também pela necessidade de demonstrar conformidade sempre que instados a fazê-lo. Esse movimento normativo

consolidou-se em instrumentos posteriores, como a Diretiva 95/46/CE da União Europeia e, de forma ainda mais robusta, no Regulamento Geral de Proteção de Dados (GDPR), que impôs obrigações de documentação, relatórios de impacto e governança organizacional como expressões concretas do dever de responsabilização. A *accountability* transformou-se em padrão jurídico internacional, marcando a passagem de uma concepção meramente formal de obediência à lei para um modelo de responsabilidade proativa e contínua, no qual cabe ao agente de tratamento, ou seja, os governos e corporações, internalizar a proteção de dados como prática institucional (BYGRAVE, 2014, p. 164-169).

No Brasil, esse princípio encontrou expressão normativa na Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), que o consagra de forma explícita em seu art. 6º, inciso X, sob a denominação de “responsabilização e prestação de contas”. A LGPD estabelece que os agentes de tratamento devem demonstrar a adoção de medidas eficazes capazes de comprovar a observância das normas de proteção de dados e a efetividade de seus programas de governança. Essa formulação confere caráter proativo à *accountability*, deslocando a ênfase do mero cumprimento formal da lei para a necessidade de evidenciar práticas de conformidade e de internalizar a cultura de proteção de dados na rotina institucional (Brasil, 2018).

Nesse movimento de expansão conceitual, a *accountability* também foi transposta para o campo da inteligência artificial, em razão da crescente delegação de tarefas decisórias a sistemas automatizados. Como observam Novelli, Taddeo e Floridi (2024, p. 1872), assegurar a legitimidade do uso da IA implica garantir que esses sistemas atuem de modo justo, conforme valores fundamentais que não podem ser comprometidos, e que sejam acompanhados de processos institucionais adequados de responsabilização. A *accountability* deixa de ser apenas um princípio normativo abstrato e passa a constituir requisito essencial como um pilar para a governança da IA orientando tanto o desenho das tecnologias quanto os mecanismos de supervisão, fiscalização e reparação de danos.

Observa-se, portanto, que a trajetória da *accountability* revela um movimento contínuo de expansão e sofisticação normativa, partindo de registros contábeis medievais até se consolidar como princípio estruturante da boa governança contemporânea. Ao longo desse percurso, o conceito deixou de se restringir ao âmbito estatal para alcançar também instituições privadas, passando a figurar em instrumentos jurídicos internacionais e nacionais como elemento indispensável para a

proteção de direitos fundamentais na era digital. Esse processo evidencia que a *accountability*, em sua dimensão moderna, não pode mais ser entendida apenas como uma técnica de auditoria ou mecanismo burocrático, mas sim como um dever jurídico e ético transversal, aplicável a diferentes esferas sociais e capaz de orientar tanto a atuação do poder público quanto a das corporações. É nesse horizonte ampliado que se insere a sua transposição para o domínio da inteligência artificial, onde assume contornos ainda mais complexos em razão da opacidade algorítmica e da multiplicidade de atores envolvidos.

Com este movimento, a *accountability* se tornou princípio transversal em importantes marcos normativos internacionais. A *Recommendation of the Council on Artificial Intelligence*, publicada pela Organização para Cooperação e Desenvolvimento Econômico em 2019, representou a primeira diretriz intergovernamental a tratar especificamente do tema, adotada por 42 países, e incluiu expressamente a responsabilização como requisito essencial para uma IA confiável e centrada no ser humano, com mecanismos de avaliação de conformidade e reparação de danos (OECD, 2019). Poucos anos depois, a *Recommendation on the Ethics of Artificial Intelligence*, aprovada pela UNESCO em 2021, reforçou esse caminho ao ser ratificada por 193 Estados-membros, destacando a necessidade de responsabilização de todos os agentes do ciclo de vida da IA por eventuais violações de direitos humanos, bem como a criação de instâncias de supervisão e avaliação de impacto ético (UNESCO, 2021). No âmbito europeu, as *Ethics Guidelines for Trustworthy AI*, elaboradas pela Comissão Europeia em 2019, detalharam sete requisitos para a confiabilidade da IA, colocando a *accountability* como elemento central ao lado da transparência, robustez técnica e não discriminação, e propondo ferramentas práticas de auditoria, documentação e identificação de responsáveis (Comissão Europeia, 2019). Em conjunto, esses documentos sinalizam uma convergência normativa internacional: a *accountability* emerge não apenas como valor ético, mas como critério operativo indispensável para orientar políticas públicas, guiar empresas e assegurar confiança social na utilização da inteligência artificial.

Seguindo essas diretrizes, na regulação da inteligência artificial no âmbito da União Européia, denominada *Artificial Intelligence Act*, incorporou a *accountability* como um de seus eixos estruturantes. O regulamento estabelece que os sistemas de IA classificados como de alto risco devem atender a requisitos específicos de governança, que incluem a implementação de sistemas de gestão de riscos,

documentação técnica detalhada, manutenção de registros (logs), mecanismos de supervisão humana, auditorias e medidas de transparência voltadas aos usuários. Esses dispositivos não apenas impõem obrigações de conformidade aos desenvolvedores e operadores, mas também exigem que possam demonstrar, a qualquer momento, que suas práticas estão em conformidade com a legislação. Assim, a *accountability* assume caráter normativo concreto: em vez de permanecer como valor ético abstrato, converte-se em dever jurídico vinculante, cujo descumprimento acarreta sanções administrativas e financeiras relevantes. Essa abordagem reflete uma transição da responsabilidade meramente declaratória para um modelo de responsabilização proativa e verificável, em que o ônus da prova recai sobre aqueles que projetam, comercializam e utilizam sistemas de IA (UNIÃO EUROPEIA, 2024).

No mesmo caminho, o Projeto de Lei sobre Inteligência Artificial no Brasil (PL 2.338/2023) incorporou expressamente a *accountability* como princípio estruturante. O texto estabelece que o desenvolvimento e o uso da IA devem observar a prestação de contas, a responsabilização e a reparação integral de danos (art. 3º, X), além da rastreabilidade das decisões durante todo o ciclo de vida dos sistemas como meio de atribuição de responsabilidades (art. 3º, IX). O projeto prevê ainda a obrigatoriedade de avaliações preliminares e relatórios de impacto algorítmico para sistemas de alto risco, bem como a criação de estruturas de governança capazes de assegurar transparência, mitigação de vieses e supervisão humana efetiva. No campo da responsabilidade civil, inova ao adotar um regime de responsabilidade objetiva para fornecedores e operadores de sistemas de alto risco ou de risco excessivo, deslocando o ônus da prova em favor da vítima. Dessa forma, a proposta aproxima-se de modelos internacionais como o AI Act, mas adapta o conceito de *accountability* ao contexto brasileiro.

Diante desse panorama, torna-se evidente que a *accountability* assume papel central como princípio jurídico e instrumento regulatório indispensável para a governança da inteligência artificial. Sua incorporação em marcos normativos internacionais, no AI Act europeu, no projeto de regulação da IA no Brasil e na própria LGPD e GDPR revela uma convergência em torno da necessidade de mecanismos que tornem visíveis e auditáveis os processos decisórios algorítmicos, impondo aos agentes responsáveis o dever de justificar, documentar e demonstrar conformidade. Ao articular responsabilidade proativa, supervisão contínua e possibilidade de sanção,

a *accountability* se apresenta como resposta estruturante aos riscos da opacidade algorítmica, que obscurece critérios e racionalidades de sistemas automatizados, e da assimetria informacional, que coloca cidadãos e consumidores em posição de vulnerabilidade frente a governos e corporações. Mais do que um valor ético, trata-se de um dever normativo transversal, essencial para assegurar transparência, reforçar a confiança pública e garantir a efetividade da proteção de direitos fundamentais no ecossistema digital.

Para garantir sua efetividade, a *accountability* não pode ser reduzida a um dever exclusivamente estatal, uma vez que os principais riscos e impactos dessa tecnologia decorrem de sua produção, implementação e exploração por atores privados. Katyal (2019, p. 57-61) sustenta que a velocidade do desenvolvimento tecnológico, aliada à escala global e à complexidade dos sistemas algorítmicos, limita a capacidade de resposta do Estado, tornando imprescindível que as corporações assumam uma parcela ativa de responsabilidade. Essa responsabilização deve ocorrer por meio da adoção de mecanismos internos e externos de controle, como códigos de conduta corporativos, comitês de ética em IA, relatórios periódicos de impacto e transparência, canais de denúncia com proteção a denunciantes e auditorias independentes que permitam a verificação de conformidade por terceiros. Tais instrumentos constituem formas de *private accountability* que complementam a ação do poder público, mas que só alcançam efetividade quando acompanhados de supervisão estatal e de pressão social, sob pena de se reduzirem a práticas meramente simbólicas ou estratégias de autopreservação reputacional. A autora adverte, nesse sentido, para o risco de captura regulatória, quando empresas moldam regras em benefício próprio, e reforça que a legitimidade da *accountability* privada depende de um equilíbrio dinâmico entre autorregulação empresarial, fiscalização governamental e participação da sociedade civil.

Assim, a evolução da *accountability*, desde sua origem vinculada à contabilidade medieval até sua consolidação como princípio estruturante da proteção de dados e da governança da inteligência artificial, revela um processo de contínua ampliação de escopo e sofisticação normativa. Contudo, a transposição desse valor para o universo algorítmico coloca desafios inéditos: a opacidade estrutural dos sistemas de IA, marcada pela complexidade técnica e pela impossibilidade de plena explicação de seus processos internos, e a assimetria informacional, que concentra poder decisório e conhecimento nas mãos de poucos agentes, fragilizando a posição

de cidadãos e consumidores.

Garantir a efetividade da *accountability* nesse contexto exige ir além da mera previsão formal em textos normativos, demandando a criação de mecanismos técnicos, regulatórios e institucionais que permitam rastrear decisões, atribuir responsabilidades e reparar danos de forma tempestiva. Trata-se de deslocar a *accountability* de um ideal jurídico-político abstrato para uma prática operacionalizada em ferramentas concretas de governança algorítmica, capazes de equilibrar inovação tecnológica com a tutela efetiva dos direitos fundamentais. É nesse ponto que se impõe aprofundar a análise dos instrumentos jurídicos e técnicos de efetivação da *accountability* na inteligência artificial, tema que será examinado a seguir.

4 TRANSPARÊNCIA E *ACCOUNTABILITY* NA MITIGAÇÃO DOS DESAFIOS INFORMACIONAIS

A análise realizada no Capítulo 3 demonstrou que transparência e *accountability* constituem os princípios estruturantes necessários à reconstrução da tutela jurídica da proteção de dados na sociedade informacional. Entretanto, o reconhecimento teórico desses princípios, por si só, não é suficiente para enfrentar os riscos estruturais da opacidade algorítmica e da assimetria informacional. A consolidação de uma governança capaz de equilibrar poder informacional exige que tais princípios sejam traduzidos em mecanismos concretos, aptos a orientar práticas institucionais, processos técnicos e padrões corporativos. Assim, o desafio que se impõe não é apenas afirmar o valor normativo desses princípios, mas identificar como eles podem se materializar no funcionamento real dos sistemas de inteligência artificial.

Nesse sentido, este capítulo tem por objetivo examinar os instrumentos técnicos, jurídicos, institucionais, corporativos e sociais que operacionalizam a transparência e a *accountability* na prática. A sociedade informacional é marcada por fluxos massivos de dados, modelos de alto grau de complexidade e cadeias decisórias distribuídas entre diversos atores públicos e privados; por isso, a efetividade dos princípios depende da articulação de tecnologias explicáveis, registros e documentação contínua, governança organizacional robusta, fiscalização estatal e participação democrática. A governança da IA, portanto, não pode ser compreendida como tarefa isolada de um único nível regulatório: exige uma arquitetura multinível que traduza princípios normativos em procedimentos verificáveis.

A partir dessa perspectiva, o capítulo avança na sistematização de um conjunto integrado de ferramentas, técnicas e de governança, destinadas a viabilizar a aplicação dos princípios discutidos no capítulo anterior. O foco desloca-se, agora, da fundamentação teórica para os meios concretos de implementação, buscando demonstrar que apenas a articulação entre instrumentos internos ao design dos sistemas (como explicabilidade, documentação, logs e rastreabilidade) e mecanismos externos de regulação, auditoria, responsabilidade corporativa e controle social é capaz de mitigar, de forma consistente, os riscos gerados pela inteligência artificial. Trata-se, portanto, de analisar como transparência e *accountability*, uma vez reconhecidas como fundamentos normativos, se convertem em práticas operacionais

que permitem atribuir responsabilidades, reduzir assimetrias e fortalecer a proteção de dados na sociedade informacional.

4.1 Da Teoria à Prática: Instrumentos De Efetivação

A trajetória histórica e normativa da *accountability* e da transparência demonstra que, embora ambos os conceitos tenham se consolidado como princípios jurídicos e valores éticos fundamentais, sua efetividade depende da capacidade de serem traduzidos em práticas concretas. Reconhecer a necessidade de responsabilização e de abertura informacional em textos legais e diretrizes internacionais representa apenas o primeiro passo; a etapa seguinte exige identificar instrumentos que possam tornar as decisões algorítmicas não apenas auditáveis, rastreáveis e justificáveis, mas também compreensíveis e verificáveis para diferentes públicos. Em outras palavras, *accountability* e transparência precisam deixar de figurar apenas como categorias de discurso normativo para assumirem contornos de exigibilidade prática, funcionando como parâmetros operacionais da governança algorítmica.

Nesse sentido, a efetivação desses princípios demanda a identificação de instrumentos operacionais, aptos a transpor a retórica normativa para exigências práticas e verificáveis. Esses instrumentos podem ser agrupados em duas dimensões complementares: de um lado, os instrumentos técnicos, que incorporam no próprio design e funcionamento da IA soluções de inteligibilidade, rastreabilidade e documentação; de outro, os instrumentos de governança, que abrangem arranjos institucionais, mecanismos corporativos, participação social e marcos regulatórios capazes de supervisionar, fiscalizar e impor consequências normativas.

Essa sistematização permitirá demonstrar como soluções tecnológicas internas e estruturas de governança externas se articulam para mitigar a opacidade algorítmica e reduzir as assimetrias informacionais. Mais do que categorias discursivas, a transparência e a *accountability* assumem, assim, contornos de exigibilidade prática, constituindo parâmetros concretos para a regulação e para a legitimidade democrática da inteligência artificial na sociedade informacional contemporânea.

4.1.1 Instrumentos Técnicos

Os instrumentos técnicos constituem a primeira camada de operacionalização da transparência e da *accountability* em sistemas de inteligência artificial. Diferenciam-se por estarem embutidos diretamente no design, na arquitetura e no funcionamento dos algoritmos, funcionando como mecanismos internos de rastreabilidade, inteligibilidade e auditabilidade das decisões automatizadas. Seu propósito central é reduzir a opacidade algorítmica, aproximando a lógica interna dos sistemas de padrões minimamente compreensíveis para usuários, reguladores e demais atores sociais, sem depender exclusivamente de instâncias externas de fiscalização. Trata-se, portanto, de soluções que integram a responsabilização ao próprio ciclo de vida da tecnologia, desde a coleta e tratamento de dados até a geração de resultados, incorporando exigências de explicabilidade, registro, monitoramento e controle automatizado.

4.1.1.1 Explicabilidade ou Inteligência Artificial Explicável

Nesse contexto, o mais citado instrumento técnico dentro dos estudos objetos desta pesquisa bibliográfica voltado à efetivação da transparência e da *accountability* é a explicabilidade. A explicabilidade em inteligência artificial, usualmente designada pela sigla em inglês XAI (explainable artificial intelligence), consolidou-se como um dos instrumentos técnicos centrais para a efetivação da transparência e da *accountability* no ecossistema digital contemporâneo. Sua origem remonta ao programa XAI da *Defense Advanced Research Projects Agency* (DARPA), lançado em 2017, que buscou fomentar uma agenda de pesquisa orientada a reduzir a opacidade dos sistemas de aprendizado profundo utilizados em contextos de segurança e defesa. O projeto tinha como propósito central viabilizar explicações compreensíveis sobre decisões automatizadas de alto impacto, tanto para engenheiros quanto para operadores, reguladores e cidadãos com uma pergunta prática: “como tornar inteligíveis sistemas de aprendizado profundo altamente complexos, sem comprometer seu desempenho e utilidade em cenários críticos, como defesa e segurança nacional?” (GUNNING *et al.,* 2021, p. 2-4).

O programa XAI da DARPA procurou tornar as decisões da inteligência artificial mais claras de três formas principais: primeiro, criando modelos que já fornecessem explicações junto com os resultados; segundo, desenvolvendo técnicas para traduzir,

depois da decisão, a lógica interna de sistemas complexos em justificativas compreensíveis; e, por fim, testando essas explicações com usuários reais, para garantir que fossem de fato úteis, claras e capazes de gerar confiança no uso da tecnologia (GUNNING *et al.,* 2021, p. 5-9).

Em seu documento final, o programa estipulou algumas lições aprendidas ao longo de sua execução: demonstrou que não existe explicação universal, pois diferentes públicos, como engenheiros, reguladores e cidadãos, demandam diferentes níveis de compreensão; revelou que a explicabilidade não implica necessariamente perda de desempenho, podendo inclusive melhorar a eficácia dos sistemas quando integrada ao design; e evidenciou a importância de calibrar as explicações ao nível de complexidade da tarefa e à carga cognitiva dos usuários (Gunning *et al.,* 2021, p. 7-9). Ao articular ciência da computação, psicologia cognitiva e interação humano-máquina, o programa consolidou a ideia de que a explicabilidade não é apenas uma questão técnica, mas também social e normativa.

Hansen e Rieger (2019, p. 43-46) ressaltam que a chamada “interpretabilidade” ou “explicabilidade” talvez não seja um conceito totalmente novo, mas a reatualização de preocupações já antigas. Os autores mostram que já nos anos 1990 havia tentativas de visualizar redes neurais, mas o crescimento da complexidade dos modelos atuais intensificou a demanda. Assim, ele propõe que para enfrentar esse desafio a explicabilidade deve ser avaliada a partir de critérios como consistência, estabilidade e utilidade prática das explicações.

No campo acadêmico, diversos autores passaram a reforçar a conexão entre explicabilidade, transparência e *accountability* algorítmica. Weller (2019, p. 23-25) sustenta que a explicabilidade está intrinsecamente ligada à responsabilidade, pois apenas sistemas de inteligência artificial capazes de oferecer justificativas inteligíveis para suas decisões podem ser considerados confiáveis e socialmente aceitáveis. Na visão de Dignum, a transparência e *accountability* só atinge sua finalidade prática quando acompanhada da explicabilidade (Dignum, 2020, p. 224). Cheong (2024, p. 2-4) destaca que a transparência em sistemas de inteligência artificial não pode ser dissociada das técnicas de Inteligência Artificial Explicável (XAI), configurando um meio de mitigar a opacidade algorítmica. Diakopoulos (2020, p. 210-213) diz que a proposta da XAI não é simplesmente abrir o código-fonte, mas oferecer explicações adaptadas ao nível de compreensão do usuário, capazes de mostrar de forma acessível os critérios ou fatores que levaram a determinada decisão algorítmica.

Dessa forma, a XAI busca reduzir a opacidade da “caixa-preta” dos modelos complexos, criando condições mínimas para que cidadãos, reguladores e especialistas possam avaliar, contestar ou confiar nas decisões automatizadas.

A lição herdada do programa DARPA de que não existe explicabilidade universal, pois diferentes públicos, como engenheiros, reguladores e cidadãos, demandam diferentes níveis de compreensão foi reforçada ao longo do tempo por estudos acadêmicos. Arya et al. (2019, p. 2-4) aprofundam a ideia ao sustentar que a explicabilidade deve ser concebida como um fenômeno relacional, variando conforme o que é explicado (dados, modelo ou decisão), como é explicado (métodos intrínsecos ou pós-hoc, locais ou globais) e para quem é explicado (engenheiros, reguladores, usuários finais). Em outras palavras, a explicação só cumpre seu papel quando é ajustada ao destinatário, tornando-se realmente útil e compreensível.

Isso significa que informações técnicas, por si sós, não são suficientes, é preciso traduzi-las em explicações acessíveis e adaptadas a diferentes públicos, como usuários comuns, autoridades regulatórias ou a sociedade em geral. Essa mediação entre complexidade técnica e compreensão social é apontada como requisito indispensável para a construção de confiança e para a aceitação ética dos sistemas de inteligência artificial (Dignum, 2020, p. 225-226). Cheong (2024, p. 4-5) alerta no mesmo sentido que a explicabilidade enfrenta limites, sobretudo no equilíbrio entre a complexidade técnica dos modelos e a necessidade de clareza para usuários, reguladores e sociedade. Dessa forma, os autores concordam que a explicabilidade é um instrumento legitimo para efetivação da transparência e *accountability* mas requer calibração cuidadosa e adaptação ao contexto em que é aplicada.

A explicabilidade, portanto, deve ser entendida como prática contextualizada, que ajusta a forma da explicação à capacidade cognitiva e às necessidades normativas de cada público. Essa ideia se conecta diretamente ao debate jurídico, pois demonstra que a transparência normativa, para ser efetiva, precisa ser acompanhada de inteligibilidade e acessibilidade.

A ideia de Inteligência Artificial Explicável (XAI) é genérica, sendo que essa pode ser efetivada por diferentes técnicas. Quanto as técnicas de explicação, a forma como essa é construída também faz diferença: ela pode ser intrínseca, quando o próprio modelo é desenhado para ser transparente e compreensível desde o início, ou pós-hoc, quando técnicas externas são aplicadas para interpretar decisões já tomadas por modelos complexos. (Arya *et al.,* 2019, p. 4). Trabalhos de caráter

panorâmico ajudaram a organizar o campo. Guidotti et al. (2018, p. 92-95), por exemplo, elaboraram uma tipologia que diferencia explicações “globais”, voltadas a descrever o funcionamento integral de um sistema, das “locais”, que se concentram em justificar uma decisão específica.

Dentre os esforços de sistematização de seus métodos e categorias, Molnar (2019, p. 11-15) apresenta um guia bastante citado que distingue dois grandes caminhos: de um lado, os modelos chamados “intrínsecos”, como regressões lineares e árvores de decisão, que já nascem mais transparentes e fáceis de interpretar; de outro, as chamadas técnicas “pós-hoc”, desenvolvidas para explicar sistemas mais complexos depois de prontos.

De acordo com Guidotti *et al*. (2018, p. 95-97), os modelos intrínsecos, também chamados de *transparent box models*, possuem estrutura naturalmente compreensível para os usuários. Nessa categoria enquadram-se técnicas como árvores de decisão, regressões lineares ou logísticas e modelos baseados em regras de associação, os quais permitem identificar de maneira direta como os dados de entrada se relacionam com as saídas produzidas. Tais modelos dispensam ferramentas adicionais de explicação, pois sua lógica interna já é acessível e verificável.

As árvores de decisão são modelos que funcionam como um fluxograma de perguntas e respostas. Cada decisão é representada por um nó da árvore, que direciona o caminho de acordo com as características do caso. No final, chega-se a uma conclusão clara e verificável. Já as regressões lineares ou logísticas explicam as decisões do sistema atribuindo pesos numéricos a cada variável. Em termos práticos, é como se cada fator contribuísse com uma pontuação específica para o resultado. Isso facilita compreender quais elementos foram mais determinantes em cada decisão, pois a relação entre entrada e saída é expressa de forma direta por meio de equações lineares (GUIDOTTI *et al.,* 2018, p. 96).

As chamadas árvores de decisão funcionam de maneira muito parecida com a lógica de uma sentença judicial estruturada em premissas. Imagine um juiz que segue um roteiro de perguntas: “o contrato foi assinado?”, “houve inadimplemento?”, “existem provas de má-fé?”. Cada resposta conduz a um novo ponto, até que, ao final, chega-se a uma decisão clara. Já as regressões lineares e logísticas lembram a fixação de critérios objetivos para calcular uma indenização: cada fator relevante (como o grau de culpa, a extensão do dano ou o tempo de duração) recebe um peso

numérico, que se soma aos demais até compor o resultado final.

Embora as técnicas intrínsecas de explicabilidade, como árvores de decisão e regressões, representem modelos naturalmente transparentes, elas não costumam ser empregadas nos sistemas de inteligência artificial mais avançados. Isso porque como explicado no item 3 deste trabalho, os modelos contemporâneos, baseados em aprendizado profundo e redes neurais, atingem patamares de desempenho que só são possíveis à custa de maior complexidade e, consequentemente, de maior opacidade, aproximando-se do paradigma das ‘caixas-pretas’.

É justamente por essa limitação dos modelos intrínsecos que surgem as técnicas pós-hoc. Elas são uma resposta prática ao dilema: como explicar modelos de alto desempenho que funcionam como “caixas-pretas”?

Entre essas técnicas pós-hoc, duas se tornaram paradigmáticas: o LIME e o SHAP. O primeiro, criado por Ribeiro, Singh e Guestrin (2016, p. 1137-1139), gera explicações aproximadas a partir da análise de pequenas variações nos dados de entrada, permitindo que um humano compreenda de forma simplificada como o modelo chegou à determinada decisão. Já o SHAP, desenvolvido por Lundberg e Lee (2017, p. 4765-4767), utiliza conceitos da teoria dos jogos para medir quanto cada variável contribuiu para o resultado. Apesar de partirem de fundamentos distintos, ambos cumprem a mesma função: aproximar a lógica interna dos sistemas algorítmicos dos seus usuários e reguladores, fornecendo justificativas auditáveis e verificáveis.

De forma mais didática e metafórica, imagine uma sentença proferida por uma inteligência artificial que decide negar um benefício previdenciário. Para explicar essa decisão, o LIME simula pequenas alterações em pontos específicos do caso, como a idade do requerente, o tempo de contribuição ou o tipo de documento apresentado, e observa se o resultado mudaria. Assim, pode-se perceber quais fatores foram mais determinantes para o indeferimento, quase como se o juiz mostrasse “se tivesse um ano a mais de contribuição, o resultado seria diferente”.

Por sua vez, SHAP analisa a mesma sentença de forma global e sistemática, como se distribuísse uma pontuação a cada elemento considerado na decisão. Ele calcula quanto cada variável, idade, tempo de contribuição, histórico médico, renda, entre outros, contribuiu positiva ou negativamente para o resultado final. É como se a IA dissesse: “a idade pesou 30%, o tempo de contribuição 50% e a renda 20% para chegar a este veredito”.

Assim, diversas são as técnicas e formas de explicabilidade das inteligências artificiais, não existindo um modelo único de Inteligência Artificial Explicável (XAI). Adadi e Berrada (2018, p. 523-526) revisaram criticamente os métodos existentes, ressaltando que nenhum deles é capaz de oferecer explicabilidade plena em todos os contextos, o que reforça a necessidade de pluralidade metodológica. Em razão da diversidade de aplicações e da complexidade crescente dos sistemas de inteligência artificial, a explicabilidade não pode ser reduzida a um método único ou a uma fórmula centralizada, mas deve ser pensada como um conjunto de estratégias complementares que se adaptam ao risco, ao público destinatário e à natureza da decisão automatizada.

Neste ponto, a integração entre dimensões técnicas e jurídicas da XAI vem sendo cada vez mais destacada, pois surge o problema de como definir quais são as técnicas e mecanismos mais adequados e proporcionais para cada caso. McDermid et al. (2021, p. 7-9) alertam que explicações superficiais ou meramente cosméticas podem comprometer a finalidade ética da explicabilidade, convertendo-a em instrumento de *compliance* simbólico.

Para evitar esse risco, consideram que a definição de quais técnicas de explicabilidade devem ser aplicadas em cada caso não pode ser orientada por critérios puramente tecnológicos, mas deve considerar o público destinatário, a finalidade da explicação e o nível de risco envolvido. McDermid *et al.* (2021, p. 13-15) observam que explicações locais são indispensáveis para proteger direitos individuais, permitindo que cada pessoa compreenda e conteste decisões automatizadas que a afetam diretamente, como previsto no direito à explicação do GDPR. Já as explicações globais são essenciais para fins de supervisão regulatória e de *accountability* organizacional, pois permitem avaliar padrões sistêmicos de funcionamento, conformidade com normas e eventuais vieses estruturais. O autor defende, ainda, que a explicabilidade não deve ser tratada como um recurso isolado, mas como parte de um "*assurance case*", isto é, um dossiê integrado de conformidade e segurança, semelhante ao utilizado em setores de alto risco como a aviação e a saúde. Nesse modelo, os mecanismos de XAI deixam de ser apenas instrumentos técnicos e passam a compor uma arquitetura de governança, capaz de assegurar que a prestação de contas e a transparência sejam proporcionais ao impacto social e jurídico de cada aplicação algorítmica (MCDERMID *et al.,* 2021, p. 16-18).

A literatura jurídica reforça que a explicabilidade opera como ponte entre a

opacidade técnica e a responsabilização legal. Hildebrandt e Gutwirth (2019, p. 82-85) mostram que a exigência de explicações está diretamente vinculada ao dever de prestação de contas, constituindo condição para a efetividade do princípio da *accountability*. No mesmo sentido, o estudo *Accountability of AI under the Law: The Role of Explanation* (Edwards; Veale, 2018, p. 45-48) argumenta que, sem mecanismos de explicação, a responsabilização jurídica tende a se tornar inviável, pois não há como identificar a lógica das decisões nem atribuir responsabilidades.

Esse elo entre técnica e norma aparece consolidado em legislações recentes. O GDPR (União Europeia, 2016), ao prever no artigo 15[1] o dever de fornecer "informações significativas sobre a lógica envolvida", inaugurou o debate normativo sobre o direito à explicação. A LGPD (Brasil, 2018), em seu art. 6º, VI, incluiu a transparência como princípio, reforçando a exigência de informações claras e acessíveis, e em seu artigo 20, §1º, incluiu a obrigação do controlador de fornecer informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada. Já o AI Act europeu (União Europeia, 2024) estabelece obrigações específicas de explicabilidade para sistemas de alto risco, impondo requisitos de documentação técnica, divulgação de dados de treinamento e garantias de interpretabilidade dos resultados. No projeto de lei n° 2338, de 2023, para regulação da IA no Brasil, a explicabilidade aparece como princípio ao lado da transparência, inteligibilidade e auditabilidade no art. 3º, inciso VI e prevê o direito a explicação (BRASIL, 2023).

Nesse contexto, a explicabilidade emerge como requisito transversal que articula técnica, ética e Direito. Se por um lado fornece instrumentos concretos para reduzir a opacidade algorítmica, por outro conecta-se diretamente a direitos fundamentais como a autodeterminação informativa, a não discriminação e o princípio da dignidade da pessoa humana. Sua função, portanto, não é apenas descritiva, mas normativa: ao tornar compreensíveis as decisões algorítmicas, a XAI viabiliza a efetividade do princípio da transparência e sustenta a *accountability* como dever de

[1] Artigo 15 do RGPD
Direito de acesso do titular dos dados
1. O titular dos dados tem o direito de obter do controlador a confirmação de que os seus dados pessoais estão ou não a ser tratados e, se for esse o caso, o direito de aceder aos dados pessoais e (...)
(h) a existência de tomada de decisão automatizada, incluindo a definição de perfis, referida no artigo 22.º (1) e (4) e, pelo menos nesses casos, informações significativas sobre a lógica envolvida, bem como a importância e as consequências previstas desse tratamento para o titular dos dados.

prestação de contas. O desafio atual reside em calibrar essas exigências de modo proporcional, evitando tanto a opacidade absoluta quanto a ilusão de explicações superficiais.

Conclui-se que a explicabilidade em inteligência artificial não pode ser reduzida a uma questão de engenharia de sistemas. Trata-se de um verdadeiro requisito jurídico e ético, indispensável para a proteção de direitos fundamentais na sociedade informacional. A XAI, ao traduzir decisões algorítmicas complexas em explicações inteligíveis, cria as condições para que o dever de transparência se converta em prática efetiva e para que a *accountability* deixe de ser mera retórica normativa, transformando-se em responsabilização concreta. Em um cenário em que algoritmos mediam cada vez mais dimensões da vida social, a explicabilidade se firma como pilar de legitimidade democrática e como instrumento de tutela da autonomia e da confiança dos indivíduos frente ao poder tecnológico das corporações e instituições públicas.

4.1.1.2 Logging e soluções em *blockchain*

No campo da auditoria algorítmica, uma das ferramentas mais tradicionais é o chamado *logging*, que consiste no registro sistemático das ações realizadas por um programa, geralmente em arquivos criados imediatamente antes ou depois de sua execução. Esses registros funcionam como um histórico de funcionamento, permitindo identificar falhas, verificar padrões de comportamento e até atribuir responsabilidades em casos de mau uso. Embora sejam instrumentos amplamente utilizados em sistemas computacionais, sua eficácia depende da correta seleção dos eventos a serem documentados e da garantia de que os logs não sejam adulterados, o que muitas vezes exige proteção por controles de acesso ou armazenamento seguro em sistemas externos (KROLL *et al*., 2017, p. 661-664).

Embora parecidos, o *logging* não se confunde com a explicabilidade se concentre na tarefa de traduzir decisões algorítmicas em termos acessíveis e compreensíveis para diferentes públicos, ela não se confunde com os mecanismos de logs, registros e auditorias. Estes últimos não têm como objetivo principal simplificar a linguagem técnica, mas sim assegurar a documentação fiel e verificável de todo o processo decisório automatizado, funcionando como um “rastro de evidências” que

permite reconstituir o caminho percorrido pelo sistema desde a entrada de dados até a produção do resultado. Trata-se, portanto, de instrumentos voltados à rastreabilidade e à verificabilidade externa, que oferecem a base técnica necessária para auditorias independentes, análises regulatórias e, em última instância, para a atribuição de responsabilidades em caso de falhas ou violações de direitos. Se a explicabilidade aproxima os algoritmos dos cidadãos e reguladores por meio da inteligibilidade, os registros e auditorias cumprem a função complementar de conferir materialidade probatória e suporte institucional à *accountability* algorítmica.

Oluwagbade (2025, p. 3-5) também argumenta que essa é uma das estratégias mais promissoras para efetivar a transparência e a *accountability* na inteligência artificial com a adoção das chamadas “caixas-pretas éticas” (*Ethical Black Boxes*), concebidas como dispositivos capazes de registrar entradas (*logging)*, processos e decisões dos sistemas algorítmicos, em analogia aos gravadores de voo da aviação. Esses registros, ao documentarem o funcionamento interno da IA, permitiriam auditorias posteriores, análises de conformidade regulatória e, sobretudo, a atribuição de responsabilidades em casos de falhas ou danos. O *logging*, portanto, na mesma ideia trazida por McDermid *et al*. (2021, p. 16-18) deve fazer parte de “*assurance case*”, isto é, um dossiê integrado de conformidade e segurança, semelhante ao utilizado em setores de alto risco como a aviação e a saúde.

De acordo com Kroll *et al.* (2017, p. 652-669), os logs constituem um instrumento central de auditabilidade, pois permitem rastrear o funcionamento interno de sistemas algorítmicos e viabilizam a identificação de responsabilidades em caso de falhas. Contudo, os autores ressaltam que a simples existência de registros não é suficiente para assegurar transparência plena, já que sua efetividade depende da integração com mecanismos técnicos de explicabilidade e com estruturas jurídicas capazes de atribuir consequências normativas.

De acordo com Reisman *et al*. (2018, p. 16-18), os registros internos de funcionamento, como logs técnicos e documentação de processos, constituem condição indispensável para que as auditorias e fiscalizações sejam efetivas. Sem esse material verificável, essas se limitariam a declarações formais sem valor prático, pois não seria possível auditar decisões, identificar falhas ou atribuir responsabilidades de forma transparente. Neste ponto, o AI Act europeu desponta como marco regulatório pioneiro, ao prever a exigência de mecanismos de logging em sistemas de alto risco, enquanto nos Estados Unidos ainda prevalece uma abordagem

fragmentada, apoiada em iniciativas como o *Algorithmic Accountability Act* (OLUWAGBADE, 2025, p. 5-6)

Contudo, o *logging* envolve desafios técnicos significativos, tais como a necessidade de capturar dados em tempo real sem comprometer o desempenho, a garantia de integridade e imutabilidade dos logs, para o que se sugerem soluções baseadas em *blockchain e hashing* (Oluwagbade, 2025, p. 6-9). As soluções baseadas em blockchain também são apontadas como uma ferramenta eficaz para efetivação da transparência por Akther et al. (2024). Os autores defendem que a tecnologia pode funcionar como uma infraestrutura de confiança para a inteligência artificial, uma vez que sua natureza descentralizada e imutável permite registrar de forma auditável os dados, os processos e as decisões dos sistemas algorítmicos. Essa rastreabilidade reforça a possibilidade de auditorias independentes, facilita a conformidade regulatória com legislações como o GDPR e o AI Act e aumenta a confiança social no uso da IA, sobretudo em setores sensíveis. Ainda que enfrente desafios de escalabilidade, integração técnica e compatibilização com direitos fundamentais como o direito ao esquecimento, a blockchain surge como uma estratégia complementar às caixas-pretas éticas para mitigar a opacidade algorítmica e fortalecer a *accountability* dos sistemas (AKTHER *et al*., 2024, p. 5-7).

A tecnologia blockchain foi inicialmente apresentada por Satoshi Nakamoto em 2008, no artigo técnico que descreveu o funcionamento do Bitcoin. Nesse trabalho, o autor propôs uma infraestrutura digital capaz de registrar transações de forma descentralizada, sem depender de uma autoridade central de confiança. Para isso, utilizou-se a ideia de uma cadeia de blocos (*blockchain*), em que cada bloco contém um conjunto de transações validadas e é conectado ao anterior por meio de mecanismos criptográficos, formando um registro único, público e imutável. Essa estrutura garante que todas as operações fiquem permanentemente acessíveis e verificáveis, criando uma espécie de livro contábil distribuído que é compartilhado entre todos os participantes da rede (NAKAMOTO, 2008, p. 1-2).

Embora inicialmente concebida para viabilizar a moeda digital Bitcoin, a proposta rapidamente demonstrou potencial muito mais amplo, servindo como base para aplicações diversas que exigem segurança, rastreabilidade e transparência, desde contratos inteligentes até sistemas de governança digital. Tapscott e Tapscott (2016) apresentam a blockchain como uma “tecnologia que registra informações de forma descentralizada, transparente e imutável”. Os autores destacam que a grande

inovação está na desintermediação, onde elimina-se a necessidade de uma autoridade central, substituída por um sistema descentralizado de verificação coletiva. Por isso, blockchain é vista como uma plataforma de confiança digital, capaz de garantir transparência, rastreabilidade e imutabilidade das informações.

Aplicada ao campo da transparência algorítmica, a lógica da blockchain poderia funcionar como um registro confiável das operações internas de sistemas de inteligência artificial. Cada dado de entrada, parâmetro de processamento ou decisão produzida poderia ser gravado de forma imutável, impedindo que os controladores adulterassem retrospectivamente essas informações. Isso permitiria que reguladores, auditores independentes e até os próprios usuários tivessem acesso a trilhas verificáveis do funcionamento do sistema, tornando possível compreender como as decisões foram tomadas e atribuir responsabilidades em caso de falhas ou discriminações. Assim, a blockchain não apenas reforçaria a confiabilidade dos registros, como também criaria uma camada adicional de prestação de contas, essencial para mitigar a opacidade algorítmica.

Gao e Zhang (2024) defendem que a tecnologia blockchain pode assumir papel estratégico na governança da inteligência artificial, sobretudo no enfrentamento da opacidade algorítmica. Por ser estruturada como um sistema descentralizado e imutável de registros, a blockchain permite documentar de forma transparente e auditável todo o ciclo de vida da IA, desde a coleta e uso de dados de treinamento até a atualização de modelos e a geração de decisões automatizadas. Essa rastreabilidade cria uma base confiável de informações que não depende da confiança em intermediários, uma vez que qualquer alteração ou manipulação indevida nos registros seria imediatamente detectada, reforçando a segurança e a integridade dos processos. Os autores destacam que, nesse sentido, a blockchain funciona como um mecanismo de *accountability* embutido, viabilizando auditorias independentes e aumentando a confiança social no uso da IA em setores sensíveis. Reconhecem, contudo, que sua aplicação enfrenta desafios relevantes, como os problemas de escalabilidade técnica, o custo computacional elevado, a dificuldade de conciliar a imutabilidade dos registros com direitos fundamentais, a exemplo do direito ao esquecimento, e a necessidade de integração normativa com legislações já vigentes, como o GDPR e o AI Act. Apesar desses obstáculos, a blockchain é apresentada como instrumento promissor para o fortalecimento da transparência e da *accountability* na regulação da inteligência artificial, especialmente quando combinada

a outros mecanismos de governança tecnológica e jurídica (GAO; ZHANG, 2024, p. 55-57).

Assim, o *logging* e as caixas-pretas éticas, buscam registrar em tempo real entradas e decisões para permitir reconstruções posteriores e atribuição de responsabilidades, enquanto a blockchain desponta como infraestrutura tecnológica para assegurar a imutabilidade e a confiabilidade desses registros.

A análise dos instrumentos técnicos evidência que a efetividade da transparência e da *accountability* em inteligência artificial depende de soluções integradas ao próprio funcionamento da tecnologia. Não se trata aqui de esgotar todos os mecanismos possíveis, mas de destacar aqueles mais recorrentes e relevantes na literatura consultada, que têm orientado a pesquisa científica e o debate regulatório. Entre eles, a explicabilidade (XAI) permite traduzir decisões complexas em justificativas compreensíveis e os mecanismos de *logging* e registros auditáveis asseguram rastreabilidade e atribuição de responsabilidades, com tecnologias como blockchain reforçam a integridade e a confiabilidade desses dados. Em conjunto, tais estratégias demonstram que a redução da opacidade algorítmica exige não apenas abertura informacional, mas também ferramentas concretas de inteligibilidade e documentação, capazes de viabilizar a responsabilização jurídica e social de quem projeta e utiliza esses sistemas.

Contudo, ainda que os instrumentos técnicos sejam fundamentais para conferir inteligibilidade e rastreabilidade às decisões algorítmicas, eles não bastam por si mesmos. Sua efetividade depende de estruturas externas capazes de supervisionar, fiscalizar e impor consequências normativas, de modo a evitar que explicações técnicas se reduzam a mero formalismo. É nesse ponto que se inserem os instrumentos de governança que serão explorados no próximo item.

4.1.2 Instrumentos de Governança Algorítmica

Ao lado dos mecanismos técnicos, a efetividade da transparência e *accountability* na inteligência artificial depende também da existência de instrumentos institucionais, normativos e regulatórios robustas, capazes de supervisionar, monitorar e, quando necessário, intervir nos processos decisórios mediados por algoritmos. Esses instrumentos compreendem tanto os arranjos formais de regulação estatal

quanto mecanismos corporativos internos e formas de participação social. Esses instrumentos devem assegurar que os princípios jurídicos de transparência e de *accountability* não permaneçam apenas no técnico, mas sejam traduzidos em práticas concretas de governança, inseridas em organizações públicas, privadas e coletivas.

A inteligência artificial está no contexto da sociedade informacional, a qual foi explorada no segundo capítulo desta pesquisa. A sociedade informacional, como vimos, não é apenas tecnológica, mas global e em rede. A lógica da interconexão, das comunidades virtuais e da inteligência coletiva (Lévy, 1999), somada à emergência de uma economia informacional global em rede (Castells, 1999), revela que os fluxos de dados, capitais e decisões não respeitam fronteiras nacionais. Nesse contexto, os algoritmos, hoje centrais à organização da produção, do consumo e da comunicação, também passam a operar em uma escala transnacional, moldando comportamentos sociais e econômicos para além do controle exclusivo de cada Estado. A dimensão global da sociedade em rede exige, portanto, que a discussão sobre transparência e *accountability* na inteligência artificial se afaste de perspectivas isoladas ou meramente locais, reconhecendo que os riscos e impactos da IA se propagam em uma arquitetura global.

Isso se coaduna com o observado por Luigi Ferrajoli (2002, p. 41-52) que trata da crise do Estado Nação, expondo que o velho Estado se preocupou em atender suas demandas coletivas internas e se tornou fraco demais em relação as demandas decorrentes do processo de globalização econômica e de interdependências existentes. Ferrajoli recorre a integração pelo Direito Internacional para superação desta crise: “[...] através da superação da própria forma do Estado nacional e através da reconstrução do direito internacional, fundamentado não mais sobre a soberania dos Estados, mas desta vez sobre a autonomia dos povos” (Ferrajoli, 2002, p. 52).

Essa percepção da insuficiência do Estado-nacional isolado é compartilhada também no campo da regulação da IA por pesquisadores da área. Luciano Floridi, Huw Roberts, Emmie Hine e Mariarosaria Tadde destacam sobre os riscos da IA:

> Esses riscos transcendem as fronteiras nacionais e renovaram os apelos por uma governança global mais forte da IA, entendida aqui como o processo pelo qual interesses diversos que transcendem fronteiras são acomodados, sem uma única autoridade soberana, para que ações cooperativas possam ser tomadas para maximizar os benefícios e mitigar os riscos da IA. Secretário-Geral das Nações Unidas, António Guterres, o Primeiro-Ministro britânico, Rishi Sunak, e o CEO da OpenAI, Sam Altman, argumentaram pela criação de um novo órgão internacional de IA modelado em instituições

> existentes, como o Painel Intergovernamental sobre Mudanças Climáticas (IPCC) e a Agência Internacional de Energia Atômica (IAEA) (FLORIDI *et al.,* 2024, p. 1).

É nesse sentido que a governança dos algoritmos deve ser compreendida como uma governança global e em rede. Tal governança não pode restringir-se à atuação estatal isolada, mas precisa articular diferentes atores, órgãos públicos nacionais e internacionais, empresas privadas e sociedade civil, em processos cooperativos de definição de padrões, monitoramento e fiscalização.

Essa conclusão também está alinhada ao conceito de pluralismo jurídico transnacional de Teubner (2023, p. 9-31), onde defende que o Direito precisa ser reformulado para incluir novas fontes de direito que surgem de processos sociais independentes das estruturas estatais tradicionais. A teoria do pluralismo jurídico, que anteriormente se concentrava nas formas jurídicas dentro dos estados-nação, agora deve considerar os processos informais de criação do direito na sociedade global. Essa nova abordagem permite uma compreensão mais clara de como normas legais podem emergir espontaneamente em diferentes setores da sociedade mundial. Um exemplo disso é a *Lex Mercatoria*, que estabelece normas jurídicas para mercados globais sem a intervenção estatal.

Essa visão de Teubner, integrada com a conclusão trazida pelos demais autores, nos leva a crer que, em razão dessa nova realidade global, um dos caminhos que parece mais adequado é o fortalecimento tanto das iniciativas globais, órgãos internacionais públicos, órgãos internacionais privados, regulações nacionais e sua efetiva conexão e comunicação entre esses institutos, de forma jurídica pluralista, com a participação dos diferentes atores internacionais.

Isso pode ser relacionado também a ideia de José Eduardo Faria (2010) em seu texto sobre a globalização econômica, onde discute a intensificação das restrições estruturais do direito positivo como uma consequência da globalização, que desestabiliza a relação entre Estado nacional, economia nacional e cidadania. Isso compromete a soberania do Estado e limita o alcance das normas jurídicas, afetando a eficácia dos direitos sociais e econômicos. Além disso, Faria observa que surgem normatividades paralelas em níveis infraestatais e supraestatais, como mecanismos de mediação, arbitragem e normas autoimpostas por corporações e instituições financeiras. Essas novas normatividades desafiam a efetividade das normas jurídicas tradicionais.

De forma semelhante, Sabino Cassese (2005), ao tratar do direito administrativo global, destaca que esse não atua de forma isolada, mas por meio de um sistema complexo de autorregulação e cooperação entre Estados e entidades privadas. Observa que o pluralismo do sistema legal global, caracterizado pela sobreposição de normas de várias organizações internacionais, pode fornecer um modelo relevante para o desenvolvimento e a implementação de normas regulatórias. Além disso, Cassese afirma que a eficácia dessas normas depende de processos decisórios conjuntos e da implementação cooperativa entre níveis nacionais e globais.

Este conceito de pluralismo jurídico e direito global é especialmente pertinente para a governança dos algoritmos, que atravessa fronteiras e setores, exigindo uma abordagem colaborativa e integrada para sua regulação.

No campo jurídico e político, a governança global é frequentemente concebida como o conjunto de mecanismos, processos e instituições, formais e informais, por meio dos quais atores públicos, privados e sociais interagem para produzir normas, coordenar ações e administrar problemas comuns em escala transnacional. Diferentemente do modelo clássico centrado exclusivamente nos Estados, a governança global pressupõe uma arquitetura policêntrica, em que organizações internacionais, empresas, sociedade civil e comunidades epistêmicas compartilham responsabilidades e exercem diferentes graus de influência sobre a formação de regras e práticas. Esse conceito, desenvolvido no âmbito das Relações Internacionais e progressivamente incorporado ao Direito, traduz a percepção de que os desafios contemporâneos da regulação da inteligência artificial, extrapolam fronteiras nacionais e demandam respostas articuladas em um sistema normativo plural e interdependente.

Sobre a governança global, Alcindo Gonçalves (2006, p.1) traz que:

> O relatório da Comissão sobre Governança Global, publicado em 1996, definiu governança como “a totalidade das diversas maneiras pelas quais os indivíduos, as instituições, públicas e privadas, administram seus problemas comuns”. Essa definição aponta claramente que a governança, entendida como os meios e processos pelos quais uma organização ou sociedade se dirigem, é construída simultaneamente pelo Estado e pelos atores não governamentais.

Por sua vez, para o contexto desta pesquisa, surge a governança de algoritmos, trazida por Doneda e Almeida (2016, p. 142-148), os quais destacam que, diante da crescente opacidade e autonomia dos sistemas de decisão automatizada, a

governança de algoritmos deve ser compreendida como um processo que envolve a articulação de instrumentos técnicos, jurídicos e institucionais voltados à transparência, à prestação de contas e à mitigação de riscos sociais e éticos. Não se trata apenas de regular o código em si, mas também de incidir sobre o ambiente em que os algoritmos operam, especialmente os dados que os alimentam, os arranjos organizacionais que os utilizam e os marcos normativos que delimitam sua atuação. Nesse contexto, a governança algorítmica demanda a participação coordenada de diferentes atores: o Estado, como garantidor de direitos e regulador; as empresas privadas, como principais desenvolvedoras e usuárias de sistemas algorítmicos; e a sociedade civil, como destinatária e fiscalizadora desses processos. A interação entre esses atores é fundamental para reequilibrar os poderes assimétricos produzidos pela tecnologia, permitindo maximizar benefícios e reduzir os potenciais efeitos nocivos da opacidade algorítmica.

Apesar da crescente ênfase na necessidade de uma governança global da inteligência artificial, não se pode desconsiderar o papel fundamental dos estados nacionais e suas regulações. É por meio delas que se estabelecem padrões jurídicos concretos de proteção de direitos fundamentais, como demonstram o GDPR e AI Act europeu, a LGPD e proposta de regulação da IA brasileira e outras. Contudo, embora imprescindíveis, tais marcos são limitados diante da lógica transnacional da sociedade informacional, na qual dados, capitais e decisões atravessam fronteiras em tempo real. Nesse sentido, as normas nacionais oferecem a base mínima de proteção, mas somente adquirem plena efetividade quando integradas a arranjos cooperativos mais amplos, capazes de lidar com riscos e responsabilidades que se manifestam em escala global.

Como explorado, a governança algorítmica, ao exigir articulação entre diferentes níveis e atores, concretiza-se por meio de instrumentos institucionais variados. Esses instrumentos vão além da dimensão técnica dos algoritmos e alcançam as estruturas organizacionais, os mecanismos corporativos, os espaços de participação social e os marcos normativos formais, compondo um ecossistema de transparência e *accountability* multinível. Cada um desses eixos cumpre função específica: enquanto órgãos de supervisão e estruturas institucionais asseguram monitoramento e *enforcement*, os mecanismos corporativos voltados à *private accountability* buscam alinhar práticas empresariais a padrões éticos e jurídicos; a participação social garante legitimidade democrática às decisões tecnológicas; e os

marcos normativos, tanto nacionais quanto internacionais, oferecem o enquadramento jurídico vinculante que consolida as obrigações e responsabilidades. A análise desses subitens que será feita a seguir tem como objetivo compreender como tais dimensões se articulam em um modelo de governança global e em rede, indispensável para enfrentar os riscos da inteligência artificial na sociedade informacional.

4.1.2.1 Instrumentos normativos e regulatórios nacionais

Os instrumentos normativos e regulatórios constituem a base estruturante da governança algorítmica, pois são eles que conferem juridicidade e exigibilidade aos princípios de transparência e *accountability*. Ainda que soluções técnicas e institucionais possam oferecer meios de rastreabilidade e de supervisão, é o direito positivo que consolida tais exigências em padrões vinculantes, transformando valores éticos e diretrizes de boas práticas em deveres normativos dotados de força coercitiva. Sem esse enquadramento jurídico, os compromissos com a transparência e a responsabilização correm o risco de permanecer no plano da autorregulação simbólica, sem mecanismos efetivos de *enforcement.*

A análise das regulações locais de inteligência artificial neste trabalho toma como referência três estudos recentes que oferecem perspectivas complementares sobre o tema. O primeiro é o relatório intitulado Panorama da Regulação da Inteligência Artificial no Brasil e no Mundo (ANBIMA, 2025), que apresenta um levantamento atualizado das principais iniciativas regulatórias em curso, com ênfase no Brasil, União Europeia, Estados Unidos e China. O segundo é o Relatório de Pesquisa – Regulação da IA no Mundo, elaborado pelo TechLaw/USP (Marques et al., 2024), que amplia o escopo da análise ao mapear a situação de 173 países e destacar experiências inovadoras em Estados do Sul Global. O terceiro é o estudo Regulação da IA – Benchmarking de Países Selecionados (ENAP, 2022), que compara de forma sistemática modelos nacionais e regionais, evidenciando as convergências e divergências entre paradigmas regulatórios. Em conjunto, esses documentos oferecem uma base sólida para compreender os diferentes modelos de regulação da IA no mundo.

Esses estudos convergem em identificar três polos regulatórios centrais na governança da inteligência artificial: União Europeia, Estados Unidos e China. Esses

blocos normativos exercem influência normativa e econômica considerável, projetando padrões regulatórios para além de suas fronteiras e condicionando o comportamento de empresas e governos (Marques et al., 2024, p. 4; ANBIMA, 2025, p. 1).

No âmbito europeu, o *Artificial Intelligence Act*, aprovado em 2024, consolidou-se como a primeira legislação geral de IA de alcance supranacional. Inspirado no modelo de proteção de dados do GDPR, o *AI Act* adota uma lógica de regulação baseada no risco, entendido como o grau de ameaça que a tecnologia pode representar para a proteção de direitos fundamentais, a segurança dos indivíduos, a saúde pública e a ordem democrática. Assim, categoriza usos proibidos, considerados inaceitáveis por violarem direitos essenciais, aplicações de alto risco, submetidas a exigências estritas de transparência, explicabilidade, auditoria e supervisão humana, e sistemas de risco limitado, com obrigações mais brandas de informação ao usuário. O diploma é ainda complementado por instrumentos de soft law, como o Código de Prática para IA de Propósito Geral, configurando um modelo de elevada densidade normativa, reconhecido como referência internacional, mas também criticado pelos elevados custos de conformidade que impõe aos agentes regulados (ANBIMA, 2025, p. 1; ENAP, 2022, p. 8-9; Marques *et al.,* 2024, p. 10-11).

Nos Estados Unidos, a regulação da inteligência artificial caracteriza-se por uma abordagem fragmentada e setorial, distribuída entre ordens executivas, legislações estaduais e diretrizes de agências reguladoras. O marco mais relevante em nível federal é o *National Artificial Intelligence Initiative Act* de 2020, que instituiu o *National Artificial Intelligence Initiative* (NAII) como comitê multidisciplinar de coordenação das políticas nacionais. Para assessorá-lo, foi criado em 2022 o *National AI Advisory Committee* (NAIAC), composto por representantes do setor privado, sociedade civil e academia, sob coordenação do Departamento de Comércio (ENAP, p. 16, 2022). A estratégia regulatória tem avançado por meio de ordens executivas, como a *Executive Order 14179 – Removing Barriers to American Leadership in Artificial Intelligence* (2025), que reorganizou a política de IA com foco na competitividade global. Em paralelo, regulações estaduais vêm surgindo em áreas específicas, como a disciplina de *deepfakes* (ANBIMA, p.2, 2025).

Em 2025 foi publicado o plano *Winning the Race: America's AI Action Plan* estruturado em três eixos: acelerar a inovação, com o objetivo de remover barreiras regulatórias e estimular inovação liderada pelo setor privado; construir infraestrutura,

com o objetivo de criar a base material para sustentar a liderança tecnológica, com expansão de *data centers*, fabricas de semicondutores e outras tecnológicas; e liderança internacional, com objetivo de projetar o poder tecnológico dos EUA no cenário global. O plano parte da premissa de que a IA é uma corrida estratégica comparável à corrida espacial, em que quem dominar a IA dominará o futuro econômico e militar. Trata-se, portanto, de um modelo pragmático e orientado ao mercado e à geopolítica global, que busca manter a liderança tecnológica norte-americana (USA, 2025).

A China, por sua vez, estrutura um modelo centralizado e setorial, implementando normas específicas para algoritmos de recomendação, reconhecimento facial, mídias sintéticas e IA generativa (ANBIMA, p. 2, 2025). Em 2017 lançou o Plano de Desenvolvimento da Próxima Geração de Inteligência Artificial, que estabelece metas estratégicas em três etapas. A primeira, até 2020, visava consolidar a IA como motor de crescimento econômico e instrumento para a melhoria da qualidade de vida, impulsionando o país pela inovação. A segunda, com horizonte em 2025, prevê avanços teóricos significativos, a consolidação da IA como força motriz da reestruturação econômica e a construção de uma sociedade inteligente. Já a terceira etapa, projetada para 2030, busca posicionar a China como centro global de inovação em IA, com economia e sociedade inteligentes plenamente consolidadas, estabelecendo a base para sua afirmação como potência econômica e tecnológica (Marques et al., 2024, p. 9-10). Trata-se de uma abordagem marcada pelo protagonismo estatal, também orientada ao mercado e à geopolítica global.

Em paralelo, a China vem buscando projetar-se como defensora da governança global da IA propondo, em 2025, a criação de uma organização internacional específica para coordenar a regulação dessa tecnologia, inspirada em experiências como o IPCC e a IAEA (ANBIMA, 2025, p. 1).

Os estudos também evidenciam o papel crescente do Sul Global, no qual se destacam abordagens inovadoras e pragmáticas. A Política Nacional de IA de Ruanda, por exemplo, enfatiza inclusão, capacitação e implementação escalonada; a Arábia Saudita, por meio da SDAIA, aposta em guias práticos para IA generativa; e a Índia prioriza a aplicação da tecnologia em setores estratégicos, como saúde e agricultura. Essas experiências demonstram que países emergentes não são meros receptores de padrões normativos estrangeiros, mas atuam como laboratórios regulatórios, elaborando soluções flexíveis e inclusivas voltadas às suas

necessidades de desenvolvimento econômico (Marques et al., 2024, p. 11–12).

No Brasil, o debate regulatório concentra-se no Projeto de Lei nº 2.338/2023, já aprovado no Senado e em tramitação na Câmara dos Deputados. A proposta adota a lógica da regulação baseada em risco e cria categorias específicas para sistemas de risco proibido, alto risco e risco ordinário. Inspirado declaradamente no AI Act, a aproximação ao modelo europeu evidencia a tentativa de alinhar o país às melhores práticas internacionais (ANBIMA, 2025, p. 1)

À luz desse panorama, uma crítica recorrente à estratégia brasileira é a tendência à mimetização integral do padrão europeu. Os relatórios analisados sugerem que nenhum dos casos estrangeiros deve ser transposto de modo automático, sob pena de reproduzir custos e arranjos institucionais alheios às capacidades nacionais. O desafio brasileiro consiste em formular uma regulação híbrida, pragmática e contextualizada, capaz de dialogar com padrões internacionais sem abdicar das especificidades socioeconômicas e institucionais do país (ENAP, 2022, p. 12-13; Marques *et al.,* 2024, p. 29-30).

Como conclusão comum, os estudos registram a ausência de consenso e fragmentação regulatória. Há um reconhecimento unânime de que não existe consenso internacional sobre como regular a IA. Essa fragmentação cria riscos de assimetria regulatória, competição normativa e barreiras comerciais. Os três relatórios concordam que, devido à natureza transnacional da IA, regulações nacionais não bastam. Surge a necessidade de cooperação internacional ou pelo menos de convergência normativa mínima, ressaltando a importância de uma arquitetura de governança global ou arranjos multilaterais, neste ponto, destaca-se a liderança da China sobre o assunto com sua proposta organização internacional para IA (ANBIMA, 2025, p.2; ENAP, 2022, p. 12-13; Marques *et al*., 2024, p. 29-30).

Em síntese, observa-se que os marcos nacionais de regulação da inteligência artificial, ainda que fundamentais para conferir juridicidade às exigências de transparência e *accountability*, revelam-se insuficientes quando analisados isoladamente. União Europeia, Estados Unidos e China delineiam paradigmas distintos, enquanto países emergentes ensaiam modelos adaptativos voltados a suas realidades socioeconômicas. No entanto, a ausência de consenso global e a fragmentação normativa reforçam a percepção de que apenas normas domésticas não são capazes de enfrentar os riscos transnacionais associados à IA. Esse cenário evidencia a necessidade de articulação com instrumentos normativos e regulatórios

internacionais, que busquem promover convergência mínima, reduzir assimetrias e estruturar uma governança global capaz de mitigar os efeitos da opacidade algorítmica e assimetria informacional em escala global.

4.1.2.2 Instrumentos internacionais

Entre os instrumentos internacionais já examinados anteriormente, destacam-se três documentos de referência que marcaram a consolidação de princípios comuns sobre a inteligência artificial. A *Recommendation of the Council on Artificial Intelligence*, aprovada pela OCDE em 2019, foi o primeiro padrão intergovernamental sobre o tema, propondo valores e recomendações de políticas públicas para promover uma IA confiável, centrada nos direitos humanos e no desenvolvimento sustentável (OECD, 2019). Dois anos depois, a *Recommendation on the Ethics of Artificial Intelligence*, da UNESCO, ratificada por 193 Estados-membros, ampliou esse horizonte ao enfatizar a proteção de direitos fundamentais, a diversidade cultural e a necessidade de mecanismos de supervisão ética (UNESCO, 2021). No âmbito europeu, as *Ethics Guidelines for Trustworthy AI*, publicadas pela Comissão Europeia em 2019, estabeleceram critérios de confiabilidade estruturados em sete requisitos, incluindo transparência, robustez técnica, não discriminação e mecanismos de governança responsáveis (COMISSÃO EUROPEIA, 2019).

Embora não tenham caráter juridicamente vinculante, esses documentos se enquadram na categoria de *soft law*, constituindo diretrizes e recomendações de natureza ética e política. Ainda assim, sua influência normativa é significativa pois orientam legislações nacionais, servem de referência para o setor privado e moldam a atuação de organismos internacionais na construção de padrões regulatórios. A ampla adesão de Estados-membros, aliada ao prestígio institucional da OCDE, da UNESCO e da Comissão Europeia, confere a esses textos um efeito normativo indireto, mas de grande relevância prática, funcionando como etapa preparatória para marcos jurídicos mais robustos e como instrumentos de convergência mínima em um cenário de fragmentação regulatória.

Nesse cenário, observa-se que, ao lado das recomendações da OCDE, da UNESCO e da Comissão Europeia, outras iniciativas internacionais têm buscado preencher as lacunas de coordenação global e responder ao caráter transnacional da

inteligência artificial. Trata-se de fóruns e arranjos multilaterais que, embora em grande medida não vinculantes, exercem papel relevante na construção de consensos mínimos, no fomento à cooperação internacional e na difusão de boas práticas. Entre eles, destacam-se a Parceria Global em Inteligência Artificial (GPAI), lançada em 2020; o Processo de Hiroshima do G7, iniciado em 2023; e a atuação da Organização das Nações Unidas, que em 2024 aprovou uma resolução histórica sobre o tema.

Em junho de 2020, a Parceria Global em Inteligência Artificial (GPAI) foi lançada, com apoio inicial de países como Canadá e França. Esta iniciativa foi criada para preencher a lacuna de um fórum internacional que pudesse promover a cooperação na IA, assegurando que seu desenvolvimento e uso sejam centrados no ser humano e éticos. O GPAI é uma iniciativa multissetorial que reúne especialistas da comunidade científica, indústria, sociedade civil, governos e organizações internacionais. A missão do GPAI é unir a teoria e a prática da IA de ponta com a implementação concreta em questões-chave, promovendo o uso responsável e ético da tecnologia (GPAI, 2024).

Os objetivos específicos do GPAI são apoiar o desenvolvimento e o uso da IA de maneira centrada no ser humano, orientada pelos direitos humanos, inclusiva, diversificada, inovadora e sustentável, a fim de fortalecer o desenvolvimento econômico. Além disso, o GPAI visa facilitar a cooperação internacional multissetorial, monitorando e articulando o trabalho realizado nacional e internacionalmente para preencher lacunas de conhecimento, melhorar a coordenação e fortalecer a cooperação internacional em IA. A organização também serve como um centro de referência para questões específicas de IA, assegurando a implementação de práticas confiáveis no ecossistema de IA (GPAI, 2024).

Porém, apesar de possuir força na comunidade internacional, o GPAI por si só, é insuficiente, pois importantes atores internacionais não estão vinculados, a exemplo da China que não é um membro e nenhuma de suas empresas (FLORIDI *et al.,* 2024, p. 7).

Lançada em 2023, na reunião do G7 realizada em Hiroshima, Japão, a chamada "G7 Hiroshima Process" nasceu em resposta a crescentes apelos por governança global da inteligência artificial. Os líderes do G7 reconheceram a urgência de diretrizes globais coordenadas para fomentar um uso e desenvolvimento responsáveis e éticos da inteligência artificial, especialmente com o rápido avanço de tecnologias emergentes como modelos fundacionais e IA generativa. Os principais

objetivos do Processo de Hiroshima incluem o fortalecimento da cooperação internacional, facilitando a colaboração entre as principais economias para lidar com desafios e oportunidades emergentes em IA; estabelecer padrões comuns, orientações de transparência, responsabilidade e ética para a implementação de IA; minimizar os riscos, criando abordagens para reduzir os riscos das tecnologias de IA de alto impacto; e promover a ética em IA, garantindo que o desenvolvimento de IA respeite os direitos humanos e valores democráticos, com a humanidade no centro (COMISSÃO EUROPEIA, 2023).

Para este fim, desenvolveram importantes documentos como o Código de Conduta Internacional Voluntário para Sistemas Avançados de IA e os Princípios Orientadores Internacionais para Sistemas Avançados de IA. O Código de Conduta pretende estabelecer diretrizes voluntárias para as práticas de negócios das organizações envolvidas no desenvolvimento dos sistemas de IA mais avançados, identificando e minimizando riscos, incentivando o compartilhamento responsável de dados e desenvolvendo políticas de governança de riscos. Os Princípios Orientadores são concebidos para garantir IA segura, confiável e transparente em todo o mundo, implementando salvaguardas de segurança robustas e priorizando a pesquisa para mitigar riscos sociais e de segurança (COMISSÃO EUROPEIA, 2023).

Essa iniciativa tem como base o trabalho da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), o que traz problemas de percepção quanto a sua falta de legitimidade por ser uma organização predominantemente europeia. Uma das possibilidades seria que essa iniciativa fosse apoiada por organizações internacionais mais plurais, como o G20 (FLORIDI *et al.,* 2024, p. 11).

Em março de 2024, a Assembleia Geral das Nações Unidas adotou uma resolução histórica sobre inteligência artificial (IA), refletindo a necessidade urgente de regular e orientar o desenvolvimento e uso da IA em um contexto global. Proposta pelos Estados Unidos e co-patrocinada por mais de 120 Estados Membros, a resolução sublinha a importância de promover sistemas de IA que sejam seguros e confiáveis, assegurando que estas tecnologias respeitem os direitos humanos e reforcem a segurança. Entre os principais objetivos da resolução estão a aceleração do progresso rumo aos Objetivos de Desenvolvimento Sustentável (ODS) utilizando a IA para enfrentar desafios globais, como mudanças climáticas, saúde pública e educação, além de desenvolver um marco regulatório internacional que aborde questões de ética, transparência e responsabilidade (ONU News, 2024).

A resolução estabelece diretrizes específicas para garantir a segurança e confiabilidade dos sistemas de IA ao longo de todo o seu ciclo de vida. Isso inclui a implementação de práticas rigorosas de desenvolvimento, como testes de segurança, auditorias independentes e documentação detalhada para assegurar transparência e responsabilidade. Além disso, a resolução enfatiza a necessidade de que a IA seja utilizada de maneira que respeite e promova os direitos humanos e os valores democráticos, evitando aplicações que possam violar esses direitos, como vigilância em massa ou discriminação algorítmica. A inclusão e a equidade são promovidas como princípios fundamentais na implementação das tecnologias de IA, garantindo que seu desenvolvimento beneficie todas as partes da sociedade de forma justa e ética (ONU News, 2024).

Essa, dentre as iniciativas citadas até o momento, é a que possui mais legitimidade e pluralidade de atores internacionais, porém, ao mesmo tempo, ainda habita somente o campo teórico das intenções, com poucos conteúdos relevantes ou vinculantes sobre o tema.

No âmbito do G20, os *AI Principles*, adotados em 2019 com base nos princípios da OCDE, estabeleceram um marco de soft law que vem orientando a governança internacional da inteligência artificial. O documento enfatiza que sistemas de IA devem ser desenvolvidos e utilizados de maneira confiável (*trustworthy AI*), assegurando não apenas crescimento inclusivo e respeito a direitos fundamentais, mas também transparência, explicabilidade e *accountability* como condições indispensáveis para a legitimidade social da tecnologia. A transparência é tratada como requisito de inteligibilidade, permitindo que usuários compreendam os resultados e possam contestá-los; a explicabilidade surge como instrumento para tornar claros os critérios e dados utilizados nas decisões automatizadas; e a *accountability* é afirmada como dever jurídico e ético de responsabilização dos agentes envolvidos em todo o ciclo de vida da IA. Embora não vinculante, o documento reforça a necessidade de políticas públicas nacionais articuladas com a cooperação internacional, projetando padrões mínimos de governança em escala global (G20, 2019).

A partir desse conjunto de iniciativas internacionais, nota-se a consolidação de um campo normativo global em torno da inteligência artificial, ainda fortemente marcado por instrumentos de *soft law*. Documentos da OCDE, da UNESCO e da Comissão Europeia estabeleceram princípios fundacionais; fóruns como a GPAI, o Processo de Hiroshima do G7 e a recente resolução da Assembleia Geral da ONU

reforçaram a necessidade de cooperação multilateral; e o G20 difundiu padrões mínimos ao redor da ideia de uma IA confiável, transparente e responsável. Em comum, tais arranjos buscam preencher a lacuna regulatória deixada pelos marcos nacionais, oferecendo diretrizes éticas e políticas de alcance transnacional, ainda que não vinculantes. Contudo, a ausência de mecanismos efetivos de *enforcement* e a exclusão de atores centrais, como a China, evidenciam os limites dessas iniciativas.

Nesse contexto que surge a estratégia chinesa de assumir papel central na definição da arquitetura normativa internacional da IA com o seu Plano de Ação para a Governança Global da Inteligência Artificial. O documento estrutura-se em treze eixos, que abrangem desde a criação de padrões técnicos globais de segurança e sustentabilidade até o fortalecimento da interoperabilidade regulatória em escala planetária. A proposta enfatiza a centralidade da ONU como fórum legítimo para coordenar a governança global e destaca a necessidade de incluir o Sul Global no processo decisório, mediante transferência de tecnologia, capacitação e apoio a infraestrutura digital. Ao conjugar discurso de cooperação multilateral com a projeção de sua própria liderança, a China busca moldar os rumos da governança global da inteligência artificial, transformando-a em instrumento estratégico de política internacional e de afirmação geopolítica (CHINA, 2025).

Em síntese, observa-se que a governança internacional da inteligência artificial ainda se encontra em estágio incipiente, apoiada majoritariamente em instrumentos de *soft law* que, embora dotados de grande relevância normativa e simbólica, carecem de mecanismos efetivos de aplicação. A coexistência de múltiplas iniciativas demonstra uma clara busca por convergência mínima em torno de valores comuns, como transparência e *accountability*. Contudo, a fragmentação entre diferentes polos de poder e a ausência de adesão universal, notadamente da China em alguns desses espaços, revelam os limites desse arranjo normativo. Assim, a governança global da IA permanece em construção, caracterizada por tensões entre cooperação e competição, bem como pela disputa de protagonismo entre atores estatais e organizações internacionais.

4.1.2.3 Instrumentos corporativos e *private accountability*

Se, por um lado, os instrumentos internacionais de *soft law* têm contribuído para

delinear princípios e valores mínimos da governança da inteligência artificial, por outro, a ausência de mecanismos de *enforcement* desloca parte significativa da responsabilidade para o setor privado. Nesse contexto, emergem os chamados mecanismos corporativos de private *accountability*, por meio dos quais empresas de tecnologia e organizações desenvolvedoras de IA instituem práticas internas de governança voltadas à transparência, ao gerenciamento de riscos e à responsabilização perante a sociedade e os reguladores.

Essa tendência decorre do fato de que os marcos tradicionais de responsabilidade pública se mostram insuficientes para lidar com os riscos específicos da inteligência artificial. Conforme teoria da opacidade algorítmica de Jenna Burrell (2016) compreendida anteriormente nesta pesquisa, muitos dos danos produzidos por sistemas algorítmicos escapam da regulação estatal por opacidade intencional e estrutural, na medida que resultam de decisões privadas, descentralizadas e opacas, tomadas no interior de grandes empresas de tecnologia. Tal cenário produz um vazio normativo no qual os cidadãos permanecem sujeitos a impactos relevantes sem garantias adequadas de reparação, reforçando a necessidade de que práticas de *private accountability* sejam concebidas como complemento indispensável à regulação pública (KATYAL, 2019, p. 52-53).

Katyal (2019, p. 60–62) desenvolve o conceito de *private accountability*, concebido como um conjunto de deveres e práticas a serem assumidos diretamente pelas empresas que desenvolvem ou utilizam inteligência artificial. Entre essas práticas, destaca a *due diligence* algorítmica para identificar e mitigar riscos, a realização de auditorias internas e externas, a adoção de códigos de conduta corporativos e a criação de mecanismos que permitam a contestação por parte dos usuários. A autora sustenta que, mesmo na ausência de imposições estatais específicas, tais organizações possuem uma obrigação ética e regulatória de responder publicamente pelos impactos sociais decorrentes de seus sistemas de IA.

Dialogando com esta ideia, Lu (2020, p. 102-108) sustenta que a *accountability* privada em matéria de inteligência artificial deve ser compreendida como extensão da responsabilidade social corporativa, com a sigla em inglês CSR, e da política Ambiental, Social e Governança - ESG. Sustenta que tal como ocorre com a divulgação de impactos ambientais, diversidade e práticas de governança, também seria necessário que as empresas tornassem públicos os riscos, critérios e incidentes vinculados ao uso de algoritmos, uma vez que a IA não é apenas questão de inovação

tecnológica, mas envolve responsabilidades públicas das corporações.

A *accountability* privada decorre da necessidade de enxergar como solução apenas as regulações estatais e normas internacionais de *soft law*, trazendo em complementação modelos como a autorregulação corporativa e a proteção a denunciantes (*whistleblowers*), para que a indústria privada crie, de forma proativa, mecanismos de responsabilidade e mitigação de riscos associados à inteligência artificial (KATYAL, 2019, p. 59-61).

Nenhum destes textos reconhece o conceito de *private accountability* como uma solução única ou global capaz de sanar, por si só, os dilemas da opacidade algorítmica e da assimetria informacional. Por isso é adequado compreender esse como um dos instrumentos possíveis dentro de uma arquitetura de governança algorítmica que se constrói de maneira global e em rede. A regulação efetiva da inteligência artificial exige a interação entre diferentes camadas normativas, desde legislações nacionais até marcos internacionais de *soft law*, passando por práticas empresariais voluntárias e mecanismos de cooperação multissetorial. Nessa perspectiva, a *private accountability* atua como peça complementar de um mosaico mais amplo de governança, em que a conjugação de esforços públicos e privados é condição essencial para conferir efetividade para transparência e *accountability* algorítmica.

Entre as principais ferramentas de *private accountability* destacadas na literatura, Novelli, Taddeo e Floridi (2024, p. 1875–1881) ressaltam a importância de códigos de conduta, relatórios de impacto algorítmico, conhecidos como AIA, auditorias independentes e mecanismos de contestabilidade como instrumentos centrais. Katyal (2019, p. 60–62) também enfatiza a necessidade de autorregulação com instrumentos como códigos de conduta, auditorias interas/exeternas, e também práticas complementares, como a criação de canais de contestabilidade para usuários e a proteção de denunciantes - *whistleblowers*, de modo a assegurar mecanismos de fiscalização endógena capazes de enfrentar a opacidade algorítmica.

Lu (2020, p. 128-153), por sua vez, com o viés de *enforcement* mercadológico, sugere a implementação regras de *disclosure* como da SEC - *Securities and Exchange Comission,* que se aplica para empresas listadas em bolsa, para divulgação pública de riscos, métricas e critérios utilizados nos sistemas de IA como extensão natural das obrigações já assumidas pelas empresas em matéria ambiental e de governança. A ideia é que um regime semelhante de disclosure regulatório poderia obrigar empresas

a revelar quais algoritmos utilizam, quais riscos de viés, discriminação ou falha foram identificados e como funcionam os mecanismos de mitigação e auditoria.

Compreendido o conceito de *private accountability* e os principais instrumentos pelos quais pode ser operacionalizada, cabe analisar de que maneira essa lógica tem sido incorporada em regulações nacionais e em marcos internacionais de *soft law*. Diversas iniciativas recentes evidenciam a tendência de articular responsabilidades privadas no interior de arranjos regulatórios mais amplos, seja por meio de obrigações explícitas em legislações, seja pela difusão de padrões voluntários em fóruns multilaterais. Assim, observa-se que a *private accountability* não se limita ao campo da autorregulação corporativa, mas vem sendo gradualmente institucionalizada como elemento integrante da governança algorítmica global.

É possível perceber que tanto o AI Act europeu, quanto o Projeto de Lei nº 2.338/2023, para regulação nacional da IA (Brasil, 2023), incorporam instrumentos concretos nesse sentido. O AI Act exige que provedores de sistemas de alto risco implementem programas internos de gestão de risco, mantenham documentação técnica detalhada, adotem registros de operação (*logging*), elaborem relatórios de impacto e se submetam a auditorias periódicas, além de incentivar códigos de conduta voluntários como forma de autorregulação setorial (União Europeia, 2024). Já o PL brasileiro, inspirado nesse modelo, prevê programas de governança em IA baseados em boas práticas e códigos de conduta, impõe a elaboração de relatórios de impacto algorítmico (RIA) e estabelece a obrigação de canais de contestação e revisão humana para decisões automatizadas (BRASILI, 2023).

Nos principais instrumentos internacionais de soft law voltados à inteligência artificial, observa-se a presença recorrente de mecanismos de private *accountability*. O G20 AI Principles (2019) estabeleceu a necessidade de transparência, explicabilidade, rastreabilidade e gestão contínua de riscos, impondo às empresas a responsabilidade pelo funcionamento confiável da IA (G20, 2019). A Recommendation of the Council on AI da OCDE reforçou esse quadro ao recomendar práticas de responsible disclosure, contestabilidade e condutas empresariais responsáveis ao longo de todo o ciclo de vida dos sistemas (OECD, 2019). De forma convergente, a Recommendation on the Ethics of Artificial Intelligence da UNESCO destacou a importância de relatórios de impacto ético, auditorias independentes e mecanismos de reparação (UNESCO, 2021). Já as Ethics Guidelines for Trustworthy AI da Comissão Europeia sistematizaram tais exigências em sete requisitos de

confiabilidade, incluindo transparência, auditabilidade, mecanismos de contestação e responsabilização corporativa (Comissão Europeia, 2019). Em conjunto, esses documentos projetam a *accountability* privada como elemento indispensável da governança algorítmica global, funcionando como complemento à regulação estatal.

Destaca-se que dentre os modelos nacionais de governança algorítmica analisados, há países que que apostam predominantemente na *private accountability* e na autorregulação corporativa como eixos centrais. É o caso dos Estados Unidos, cuja estratégia regulatória mantém forte orientação ao mercado, transferindo para as empresas a responsabilidade de adotar códigos de conduta, frameworks de governança e instrumentos voluntários de transparência e mitigação de riscos (Marques et al., 2024, p. 9; ANBIMA, 2025). No mesmo sentido, o Reino Unido adota uma abordagem “pro-inovação”, marcada por diretrizes setoriais não vinculantes e pela ênfase em práticas empresariais voluntárias (Marques et al., 2024, p. 13). Singapura constitui outro exemplo emblemático, ao estruturar o *Model AI Governance Framework* e guias regulatórios aplicados ao setor financeiro, concebidos como instrumentos de *soft law* que induzem a autorregulação (ANBIMA, 2025,). Japão e Austrália também se inserem nesse grupo, ao priorizarem a elaboração de guias e códigos éticos voluntários para orientar as empresas no uso responsável da inteligência artificial, em vez da criação de um marco legislativo rígido (ENAP, 2022). Em comum, esses países assumem que a regulação estatal deve ser leve e flexível, cabendo ao setor privado a principal responsabilidade de operacionalizar a governança algorítmica por meio de mecanismos de *accountability* internos e autorregulatórios.

Assim, relatórios de impacto, códigos de conduta empresariais, sistemas de auditoria interna, comitês de ética em IA e adesão a padrões internacionais voluntários constituem exemplos de instrumentos de *private accountability*, que buscam compensar as lacunas normativas estatais e internacionais. Ainda que frequentemente criticados por seu caráter voluntário e pela assimetria de poder entre grandes corporações e demais atores sociais, esses mecanismos têm se consolidado como peças centrais na construção de um ecossistema regulatório híbrido, no qual a autorregulação corporativa interage com padrões jurídicos e éticos globais.

4.1.2.4 Participação social e deliberação pública

Foi compreendido que a governança, de forma conceitual, demanda a participação de todos os atores envolvidos sobre o seu objeto. No caso da governança algorítmica, pela natureza difusa e transversal da inteligência artificial, essa também não pode restringir-se às esferas estatais, internacionais ou corporativas. Se até aqui foram analisados o papel do Estado, dos organismos internacionais e das empresas privadas, é igualmente imprescindível considerar a sociedade civil, enquanto destinatária direta dos impactos da IA e, ao mesmo tempo, agente fiscalizador e participante ativo desses processos.

A base para este tema foi explorada anteriormente através dos tópicos da transparência, *accountability* e especialmente a explicabilidade, ao entender que esses princípios só cumprem seu papel quando são ajustados ao destinatário, ou seja, a sociedade civil, tornando-se realmente útil e compreensível. O combate a opacidade algorítmica e a assimetria informacional se configura como uma premissa para a participação da sociedade civil, para que essa seja munida de informações amplas, verídicas e inteligíveis sobre o assunto.

Importante destacar também que ao tratar de participação social e deliberação pública estamos falando, majoritariamente, sobre *accountability*, que dentro da sua definição abarca a prestação de contas. Mas também estamos falando sobre transparência e explicabilidade, conceitos estes que também são fundamentais para os atores da sociedade civil.

Jonathan Fox (2015, p. 346) desenvolve o conceito de *social accountability*, que se refere as práticas e mecanismos pelos quais cidadãos e organizações da sociedade civil exercem monitoramento, avaliação e pressão sobre governos e instituições privadas, buscando assegurar maior transparência, responsividade e equidade nas decisões que afetam a coletividade. Diferentemente da prestação de contas vertical tradicional, em que o Estado responde formalmente às instituições de controle, a *accountability* social enfatiza a criação de canais horizontais de participação, nos quais a sociedade se torna coprodutora do processo de governança. Trata-se, portanto, de um instrumento que fortalece a democracia ao ampliar o espaço para que os cidadãos não apenas recebam informações, mas também possam influenciar ativamente políticas públicas e decisões institucionais.

Este conceito deve ser levado em consideração na governança da inteligência

artificial. Ao enfatizar o papel ativo da sociedade civil na fiscalização de políticas públicas e na exigência de transparência dos agentes de poder, essa abordagem reforça que a construção de mecanismos de responsabilização não pode se restringir ao Estado ou ao setor privado. No campo da IA, em que riscos de opacidade algorítmica e assimetrias informacionais afetam diretamente os cidadãos, a incorporação de instrumentos de participação social amplia a legitimidade e a efetividade da governança. A deliberação pública, a atuação de organizações da sociedade civil e o fortalecimento de canais institucionais de contestação funcionam como contrapesos essenciais para assegurar que a tecnologia seja desenvolvida e utilizada de forma ética, equitativa e alinhada aos direitos fundamentais.

A participação da sociedade civil aparece como um eixo estruturante em diferentes instrumentos de soft law sobre inteligência artificial. A Recomendação da UNESCO sobre a Ética da IA (2021) enfatiza a inclusão social e a pluralidade de vozes, defendendo a criação de instâncias de supervisão ética abertas e participativas, o que legitima democraticamente a governança da IA (UNESCO, 2021). A Recomendação da OCDE sobre IA (2019) também reconhece o papel da sociedade civil e da academia, ao lado de governos e empresas, no processo de formulação e fiscalização de políticas públicas relacionadas à IA, reforçando a necessidade de um debate público informado e multissetorial (OECD, 2019). Já as Diretrizes Éticas da Comissão Europeia para uma IA Confiável (2019) colocam a participação de stakeholders como requisito essencial, incluindo ONGs, associações de consumidores e cidadãos diretamente afetados, para garantir a confiabilidade e legitimidade social da tecnologia (Comissão Europeia, 2019). Por fim, a Resolução da Assembleia Geral da ONU sobre IA (2024) destaca inclusão e equidade como princípios centrais, vinculando a participação da sociedade civil aos Objetivos de Desenvolvimento Sustentável, de modo a assegurar que os benefícios da IA sejam distribuídos de forma justa e não restritos a grupos privilegiados (ONU, 2024).

Nos estudos comparativos sobre regulação de IA, observa-se que a participação da sociedade civil constitui um eixo relevante de legitimação democrática. A pesquisa da ENAP (2022) indica que em democracias consolidadas, como EU, EUA, Canadá e Reino Unido, a construção de marcos normativos de inteligência artificial tende a envolver consultas públicas e engajamento de organizações sociais. No Brasil, o panorama descrito pela ANBIMA (2025) revela a incorporação de audiências públicas e diálogos com setores da sociedade no processo legislativo, fato que

prolonga a sua aprovação desde 2023. Já o relatório do TechLab/USP (Marques et al., 2024) chama atenção para o Sul Global, destacando países como Ruanda e Índia que, diferentemente das potências tradicionais, têm apostado em arranjos participativos e inclusivos como forma de legitimar suas estratégias nacionais de IA. Em comum, essas experiências apontam que a deliberação pública e a inclusão cidadã não são acessórios, mas elementos centrais para assegurar que a governança algorítmica reflita valores democráticos e sociais.

Assim, a participação da sociedade civil na governança algorítmica pode se materializar diversas formas, uma delas são as consultas públicas, em que qualquer cidadão ou organização pode enviar contribuições antes da aprovação de leis, regulações e diretrizes. Outra forma são os fóruns e grupos multissetoriais, que reúnem governo, empresas, academia e organizações da sociedade para debater riscos e propor soluções. Também existem conselhos consultivos e comitês de ética que reservam espaço para ONGs e associações de consumidores, permitindo que a voz da sociedade esteja presente nas decisões. Além disso, instrumentos como canais de denúncia, plataformas de contestação e organizações de defesa de direitos digitais ajudam a fiscalizar práticas empresariais e cobrar maior transparência. Essas diferentes modalidades de participação, que já aparecem em recomendações internacionais, reforçam a ideia de que a governança da IA só será legítima e eficaz se incluir ativamente os cidadãos que são diretamente afetados por seus impactos.

No Brasil, por exemplo, o Projeto de Lei nº 2338/2023 sobre IA passou por consulta pública no Senado e, em apenas duas semanas, recebeu mais de 2 mil páginas de contribuições, demonstrando a mobilização da sociedade brasileira em torno do tema. Organizações civis requereram, inclusive, transparência na publicação dessas contribuições para assegurar um debate verdadeiramente aberto e informado, prática alinhada à tradição do país em consultas legislativas abertas (como ocorreu no Marco Civil da Internet e na Lei de Dados Pessoais) (ITS Rio,

Organizações têm pleiteado assentos fixos em comitês e grupos de elaboração de políticas de IA, em pé de igualdade com governo e setor privado. Um exemplo foi a carta aberta de mais de 40 entidades pedindo inclusão efetiva de representantes civis na cúpula global AI Action Summit em 2025 que vai reuniu 80 chefes de estado e muitos líderes empresariais do setor de inteligência artificial (IA), enfatizando que representantes da sociedade civil devem ter assento permanente à mesa para uma governança legítima (DATA PRIVACY BRASIL, 2025).

Assim, a sociedade civil vem tomando consciência de seu papel dentro da governança algorítmica e a participação social e deliberação pública vem assumindo papel fundamental dentro dos instrumentos de efetivação da transparência e *accountability.* Conclui-se que a participação social e a deliberação pública constituem elementos centrais para legitimar decisões regulatórias, ampliar a transparência e assegurar que os arranjos normativos reflitam valores democráticos, direitos fundamentais e expectativas coletivas. Em um campo marcado por opacidade técnica e assimetria informacional, a inclusão da sociedade civil amplia o escrutínio público e fortalece os mecanismos de transparência e *accountability*, funcionando como contrapeso às assimetrias de poder entre governos, grandes corporações e cidadãos.

Em síntese, embora os instrumentos técnicos, institucionais, corporativos e sociais analisados até aqui revelem um avanço concreto na direção de uma governança algorítmica mais transparente e responsável, eles ainda operam de forma dispersa e com graus distintos de maturidade e efetividade. A diversidade de mecanismos existentes demonstra a amplitude das tentativas de enfrentamento dos riscos informacionais, mas também expõe lacunas relevantes, sejam técnicas, jurídicas ou institucionais, que impedem a formação de um modelo coeso e capaz de responder, de maneira integrada, à opacidade algorítmica e à assimetria informacional que caracterizam a sociedade informacional. Essas insuficiências tornam evidente a necessidade de sistematizar essas iniciativas em uma arquitetura mais abrangente, alinhada à lógica informacional, global e em rede da contemporaneidade, movimento que será desenvolvido no item seguinte.

4.2 Sistematização e Avanços Necessários Para o Enfrentamento dos Desafios Informacionais

Para enfrentar os desafios informacionais, inicialmente, deve ser resgatada a ideia e o modo de funcionamento da sociedade informacional e da sociedade em rede. Conforme Lévy (1999, p. 125-129), a sociedade informacional é estruturada por três palavras de ordem, sendo elas interconexão, comunidades virtuais e inteligência coletiva, que definem o próprio surgimento do ciberespaço e moldam a forma como informações são produzidas, distribuídas e apropriadas. Castells reforça essa dinâmica ao caracterizar a sociedade em rede como informacional, global e

interconectada: informacional porque os dados se tornaram o núcleo da economia; global porque fluxos informacionais atravessam fronteiras em alta velocidade e sem restrições; e em rede porque não existe mais um núcleo de poder, mas uma rede interconectada de poderes (Castells, 1999). Esse novo ambiente, marcado por interdependência tecnológica, circulação transnacional de dados e reconfiguração das formas de sociabilidade, gera desafios inéditos que não podem ser enfrentados pelos institutos tradicionais do Direito. As categorias jurídicas clássicas, concebidas para realidades estáveis, territorializadas e analógicas, revelam-se insuficientes diante de sistemas algorítmicos de alta complexidade, cuja atuação ocorre em escalas globais e em ambientes regulatórios assimétricos. Assim, qualquer estratégia voltada à mitigação dos riscos informacionais, especialmente aqueles associados à opacidade algorítmica e à assimetria informacional, deve ser, necessariamente, informacional, global e em rede, refletindo a própria lógica estruturante da sociedade contemporânea.

Assim, a superação dos riscos informacionais característicos da sociedade informacional, demanda uma abordagem integrada, multissetorial e ancorada em diferentes níveis de governança. A análise desenvolvida no item anterior demonstrou que os instrumentos de efetivação da transparência e da *accountability* se distribuem em diferentes camadas: técnica, normativa, institucional, corporativa e social. O presente item sistematiza essa integração e identifica avanços necessários para o enfrentamento dos desafios informacionais.

A primeira dimensão é a camada técnica. Essa fornece os meios materiais que tornam possível a transparência e a *accountability*, constituindo o núcleo operativo que condiciona a efetividade dos instrumentos jurídicos, normativos, corporativos e sociais que dependem, inevitavelmente, da existência de uma base tecnológica para serem efetivados. Nesse âmbito, os instrumentos de explicabilidade, sejam eles intrínsecos ou pós-hoc, oferecem meios distintos e complementares para reconstruir a lógica decisória de modelos avançados, permitindo tanto a compreensão global de seu funcionamento quanto a justificação localizada de decisões específicas. A esses mecanismos somam-se práticas sistemáticas de registro, que documentam entradas, saídas e eventos críticos ao longo do ciclo de vida algorítmico, funcionando como uma forma de “memória” técnica. Tais instrumentos, embora distintos, convergem para um mesmo objetivo: fornecer infraestrutura técnica capaz de sustentar práticas robustas de transparência e *accountability.*

Essa camada técnica, todavia, só desempenha sua função quando inserida em uma camada normativa ou regulatória que transforma capacidades técnicas em deveres jurídicos. Esse nível compreende legislações ou regulações nacionais, regionais e internacionais, regulações setoriais, padrões internacionais de governança, diretrizes multilaterais e instrumentos de *soft law* elaborados por organismos intergovernamentais, diretrizes infralegais e interpretações administrativas. A regulação atua como um filtro que seleciona quais requisitos técnicos devem ser implementados obrigatoriamente, estipulando padrões mínimos de transparência e *accountability*. Por meio dessa conversão normativa, os resultados produzidos pelos mecanismos técnicos deixam de depender da boa vontade das organizações e passam a constituir elementos de conformidade jurídica. É essa camada que determina, por exemplo, quando um sistema de IA deve ser submetido a uma auditoria externa, quando deve produzir relatórios de impacto, ou quando deve fornecer explicações a titulares e autoridades, bem como quais instrumentos técnicos devem ser utilizados em cada caso. Sem esse enquadramento normativo, a técnica permanece dispersa, sem força obrigatória ou padronização.

A camada institucional representa um terceiro nível indispensável, ao assegurar que as obrigações legais sejam efetivamente fiscalizadas. Aqui se inserem autoridades de proteção de dados, agências reguladoras setoriais, instâncias certificadoras e organismos estatais responsáveis por realizar auditorias, aplicar sanções e emitir diretrizes. Essa camada funciona como ponte entre a norma e a prática, pois confere autoridade, coercibilidade e capacidade de supervisão aos padrões regulatórios. Sua atuação depende, em grande medida, da existência de evidências técnicas suficientes para permitir controle, razão pela qual se funda diretamente sobre os instrumentos da camada técnica. Ao mesmo tempo, fornece orientações que influenciam o desenvolvimento de regras jurídicas e modelos corporativos de conformidade, o que revela seu caráter duplo: fiscalizador e normativo.

A quarta dimensão desse ecossistema é a camada corporativa, responsável pela implementação cotidiana da governança algorítmica no âmbito das organizações. Trata-se do nível em que se desenvolvem relatórios de impacto, comitês internos de ética, auditorias internas, códigos de conduta algoritmos, canais de denúncia e outros. Essa camada constitui o espaço privilegiado da chamada *private accountability*, que desempenha papel fundamental em contextos de rápida inovação tecnológica e

regulação incompleta. Na prática, as empresas são responsáveis por transformar padrões normativos em rotinas operacionais, definindo fluxos, controles, procedimentos e estruturas que viabilizam a governança efetiva dos sistemas de IA. A interação entre a camada corporativa e a camada técnica é especialmente relevante: a empresa depende de instrumentos técnicos para produzir documentação, explicações e rastreabilidade, mas depende também de padrões normativos para definir o grau de rigor necessário para cada tipo de sistema.

Por fim, a camada social integra mecanismos de participação pública, deliberação coletiva, controle democrático e escrutínio da sociedade civil. A literatura recente tem ressaltado o papel da social accountability na governança algorítmica, destacando a importância de envolver organizações civis, especialistas, associações de consumidores e cidadãos na fiscalização do uso de tecnologias de alto impacto. Consultas públicas, fóruns multissetoriais, comitês participativos, plataformas de contestação e organizações de defesa de direitos digitais são instrumentos que reforçam a legitimidade da regulação e funcionam como contrapesos às assimetrias técnicas e econômicas que caracterizam o setor. A participação social também retroalimenta as camadas normativas e institucionais, ao pressionar por aprimoramentos regulatórios, maior transparência e maior responsabilidade por parte de agentes privados e públicos.

Sistematizados dessa forma, os instrumentos técnicos e de governança revelam não apenas sua função individual, mas sobretudo sua interdependência estrutural. A técnica, com a compreensão de como o sistema funciona, produz inteligibilidade, com informação técnica estruturada, com as condições materiais para e efetivação da transparência e *accountability*; a norma estabelece como esses sistema deve funcionar, definindo obrigações e padrões, transformando essas condições em obrigações jurídicas; as instituições asseguram a fiscalização, transformando obrigações em controle; as empresas implementam procedimentos concretos de controle, através da *private accountability*; e a sociedade, através da *social accountability,* legitima, monitora e pressiona por melhorias. Trata-se de um ciclo contínuo de governança algorítmica, no qual cada nível funciona como condição para o funcionamento dos demais. A ausência de qualquer uma dessas camadas compromete a eficácia de todo o sistema: sem instrumentos técnicos, não há como auditar; sem normas, não há padronização; sem instituições, não há *enforcement*; sem empresas comprometidas, não há implementação; sem sociedade civil, não há

legitimidade ou controle democrático.

Em resumo, a técnica alimenta a norma e fornece evidências, a norma obriga a empresa, a instituição fiscaliza, a empresa implementa e presta contas à sociedade, e a sociedade monitora e pressiona o Estado para aprimorar a regulação e as empresas ao seu cumprimento, gerando um ciclo contínuo para enfrentamento dos desafios informacionais.

Embora todos os elementos que compõem essa arquitetura já existam, sua aplicação costuma ocorrer de maneira fragmentada, com cada camada operando de forma relativamente autônoma e dissociada das demais. A ausência de um arranjo integrado limita a capacidade desses instrumentos de enfrentar, de modo efetivo, os riscos informacionais característicos da inteligência artificial. A consolidação de uma visão sistêmica, que compreenda a interdependência estrutural entre mecanismos técnicos, normas jurídicas, arranjos institucionais, práticas corporativas e participação social, constitui um dos avanços teóricos e práticos indispensáveis para mitigar a opacidade algorítmica e a assimetria informacional. Essa integração, contudo, ainda se encontra em estágio incipiente, o que reforça a necessidade de uma reflexão sistematizadora capaz de explicar como essas camadas podem convergir, quais lacunas persistem e quais transformações são necessárias para que produzam efeitos concretos. Em última análise, a eficácia dos princípios de transparência e *accountability* depende justamente dessa articulação coerente, multinível e em rede, apta a formar uma arquitetura de governança capaz de operar simultaneamente em diferentes escalas, técnica, normativa, institucional, corporativa e social, e de responder de maneira coordenada aos desafios que caracterizam a sociedade informacional.

Concluída a sistematização das camadas técnicas, normativas, institucionais, corporativas e sociais, bem como demonstrada a necessidade de compreendê-las de forma integrada e não fragmentada, torna-se possível identificar um conjunto de avanços indispensáveis para que essa arquitetura alcance efetividade prática.

O primeiro avanço consiste na urgência de mecanismos de cooperação internacional que transcendam modelos regionais fragmentados. A assimetria entre polos regulatórios, União Europeia, Estados Unidos, China e Sul Global, evidencia fragmentação normativa que compromete o enfrentamento dos desafios informacionais. Considerando que, conforme Lévy (1999, p. 125-129), a sociedade informacional opera por interconexão e inteligência coletiva, e que, segundo Castells

(1999), a economia é global, informacional e em rede, a governança da IA não pode permanecer restrita a estruturas domésticas, pois a atuação isolada dos Estados-nação é insuficiente para regular tecnologias cuja natureza é transnacional desde a origem. Apesar da crescente proliferação de instrumentos nacionais e regionais, o cenário global permanece marcado por assimetrias estruturais de poder tecnológico e por regimes regulatórios desconectados entre si, o que compromete a eficácia desses instrumentos.

A ausência de padrões compartilhados intensifica fenômenos como a arbitragem regulatória, permitindo que empresas de grande porte selecionem jurisdições mais permissivas ou apostem na inconsistência entre ordenamentos para evitar mecanismos de responsabilização. Nesse contexto, a coordenação internacional não é uma escolha política secundária, mas uma condição estruturante para equilibrar assimetrias informacionais e assegurar que instrumentos técnicos, normativos e institucionais produzam efeitos proporcionais aos riscos envolvidos

Assim, a governança da inteligência artificial precisa avançar na direção de mecanismos genuinamente informacionais, globais e em rede, em consonância com a lógica estrutural da própria sociedade contemporânea. Sem esse movimento de coordenação global, mesmo os arranjos nacionais mais sofisticados continuarão estruturalmente limitados, pois enfrentarão sistemas algorítmicos que não reconhecem fronteiras, operam em ecossistemas transnacionais e concentram recursos computacionais, informacionais e estratégicos muito além do alcance das regulações domésticas.

Essa conclusão constitui uma das contribuições centrais deste trabalho: sem coordenação internacional, os instrumentos técnicos e jurídicos permanecem impotentes diante de agentes privados cuja atuação transcende fronteiras.

O segundo avanço necessário consiste na superação do distanciamento estrutural entre conhecimento técnico e prática jurídica, que ainda caracteriza a atuação de operadores do direito, autoridades públicas e instituições responsáveis pela governança da inteligência artificial. A literatura especializada tem reiterado que a abordagem exclusivamente dogmática ou normativista é insuficiente para lidar com sistemas algorítmicos de alta complexidade. Estudos como os de Edwards e Veale (2017, p. 40-43) e de Wachter, Mittelstadt e Floridi (2017, p. 845-847) demonstram que direitos previstos em legislações como o GDPR, especialmente o suposto direito à explicação, permanecem inefetivos diante da opacidade estrutural dos modelos de

aprendizado de máquina. Isso ocorre porque o direito, quando operado sem compreensão técnica adequada, converte-se em mero enunciado abstrato, dissociado das condições materiais necessárias para sua aplicação.

Esse cenário revela um problema mais profundo: operadores jurídicos, reguladores e membros da sociedade civil ainda não dominam, nem são estimulados a dominar, os instrumentos técnicos indispensáveis para compreender, fiscalizar e contestar decisões algorítmicas. A consequência é a formação de uma assimetria cognitiva que favorece agentes privados detentores de conhecimento especializado e limita a capacidade de escrutínio democrático. Princípios como transparência e *accountability* tornam-se, assim, incapazes de produzir efeitos concretos quando aplicados por atores que não compreendem a linguagem, a lógica e as limitações técnicas dos sistemas que pretendem regular.

Para enfrentar esse déficit estrutural, torna-se necessário que operadores jurídicos, reguladores e membros da sociedade civil sejam expostos, de forma sistemática, a conhecimentos técnicos fundamentais sobre explicabilidade, *logging*, arquitetura algorítmica, modelos de aprendizado de máquina e riscos informacionais. Sem esse movimento de capacitação e atualização, a governança algorítmica permanecerá presa a um juridicismo abstrato, incapaz de produzir efeitos materiais que mitiguem a opacidade e a assimetria informacional.

Desse modo, o avanço necessário consiste na integração entre saber técnico, prática jurídica e compreensão social, de modo que a regulação, a fiscalização e o controle democrático deixem de operar com base em uma racionalidade meramente formal e passem a incorporar os elementos técnicos indispensáveis para enfrentar os riscos informacionais. Trata-se de reconhecer que o conhecimento jurídico, isolado, não é suficiente para lidar com tecnologias cuja lógica interna ultrapassa radicalmente as categorias clássicas do Direito.

Essa reflexão constitui também um dos objetivos deste estudo: contribuir para aproximar o campo jurídico dos instrumentos técnicos que condicionam a efetividade da transparência e da *accountability*, oferecendo subsídios para que operadores do direito, instituições públicas e a sociedade compreendam que a tutela dos direitos na sociedade informacional depende, necessariamente, da superação da distância entre técnica e normatividade.

O terceiro avanço necessário diz respeito ao reconhecimento de que não existe, e não pode existir, uma solução técnica única capaz de enfrentar de maneira

abrangente a opacidade algorítmica. A literatura especializada tem reiterado que instrumentos como explicabilidade, logging, documentação técnica estruturada, técnicas pós-hoc, modelos intrinsicamente interpretáveis e soluções de blockchain cumprem funções distintas e complementares, sendo incapazes, isoladamente, de oferecer uma resposta suficiente aos desafios informacionais contemporâneos. A crença em uma técnica universal, seja ela XAI, auditorias externas, modelos interpretáveis ou qualquer outro mecanismo isolado, constitui uma ilusão metodológica que compromete a própria efetividade da governança algorítmica. Cada técnica possui limitações intrínsecas e atua apenas sobre parte dos problemas que compõem a opacidade estrutural dos sistemas de inteligência artificial.

Por essa razão, a literatura tem convergido para a necessidade de uma abordagem plural, capaz de combinar diferentes métodos em arranjos proporcionais ao risco, ao contexto, ao tipo de decisão e ao público destinatário. Portanto, o avanço necessário consiste na superação explícita de qualquer expectativa de solução única ou “receita de bolo” para enfrentar a opacidade algorítmica. A governança da inteligência artificial exige ecossistemas metodológicos complexos, compostos por diferentes técnicas que se reforçam mutuamente e que precisam ser mobilizadas de forma estratégica, contextual e proporcional ao risco. Somente uma abordagem plural e integrada é capaz de responder à diversidade de aplicações algorítmicas e aos desafios informacionais que caracterizam a sociedade contemporânea.

Por fim, além dessas contribuições centrais, destacam-se contribuições secundárias que complementam a arquitetura proposta. A primeira refere-se ao reconhecimento dos limites estruturais da explicabilidade como ideal absoluto. Modelos de aprendizado profundo, dada sua complexidade interna, caráter não linear e natureza adaptativa, não permitem explicações completas ou perfeitamente transparentes. Exigir transparência integral pode gerar distorções, criar expectativas jurídicas irreais e comprometer até mesmo a eficiência técnica desses sistemas. A explicabilidade deve, portanto, ser tratada como instrumento proporcional ao risco, adequado ao contexto e calibrado ao público destinatário, e não como valor absoluto ou solução universal.

A segunda refere-se ao potencial ainda subexplorado da *blockchain* como instrumento para enfrentamento dos desafios trazidos pela IA. A imutabilidade, a rastreabilidade e a descentralização dos registros criam condições privilegiadas para auditorias independentes, documentação confiável e responsabilização efetiva.

Apesar desse potencial, trata-se de um campo ainda incipiente na literatura, marcado por escassez de estudos. Essa lacuna teórica revela a necessidade de pesquisas mais aprofundadas que examinem, de forma rigorosa, como essa tecnologia pode contribuir para a mitigação da opacidade algorítmica e o fortalecimento da transparência e *accountability*.

Por fim, a última contribuição secundária tem destaque para o Brasil e diz respeito à inadequação de importar integralmente modelos regulatórios estrangeiros, especialmente o modelo europeu, para o contexto brasileiro. A experiência recente mostra que transposições normativas descontextualizadas tendem a produzir arranjos institucionais frágeis e incapazes de gerar efeitos concretos. O descompasso entre o texto da LGPD e sua efetivação prática exemplifica essa dificuldade: embora a legislação brasileira seja robusta em termos formais, sua aplicação permanece limitada, marcada por baixa fiscalização, pouca cultura institucional de proteção de dados e práticas de "*compliance* cosmético". Assim, a construção de um modelo próprio, ajustado às capacidades e necessidades nacionais, constitui elemento essencial para a governança algorítmica no país.

Em conjunto, esses avanços e contribuições apontam para a necessidade de um modelo de governança algorítmica que seja, simultaneamente, informacional, global, interdependente e metodologicamente plural. Apenas um arranjo integrado, capaz de combinar técnica, norma, instituições, práticas corporativas e participação social, poderá enfrentar adequadamente os desafios impostos pela opacidade algorítmica e pela assimetria informacional na sociedade contemporânea.

5 CONCLUSÃO

A presente dissertação partiu de um problema concreto e contemporâneo: estruturar a efetiva tutela jurídica da proteção de dados pessoais no contexto da inteligência artificial (IA) diante da opacidade algorítmica e da assimetria informacional. A investigação não se limitou a mapear os desafios informacionais da IA como também buscou compreender quais princípios jurídicos e instrumentos técnicos, jurídicos, corporativos e sociais podem efetivamente reequilibrar relações moldadas pela sociedade informacional.

Partiu-se da hipótese de que a proteção de dados só se tornará eficaz se os princípios de transparência e *accountability* forem tomados como fundamentos normativos estruturantes da regulação. O exame crítico permitiu aferir que a hipótese se confirmou: a tutela jurídica de dados pessoais diante da IA exige reconhecer a transparência e a *accountability* como princípios estruturantes. Esses princípios não são meros slogans, mas referencial normativo que orienta o desenho de regras concretas e a construção de instrumentos técnicos e de governança.

Do ponto de vista teórico, a investigação compreendeu que a IA não inaugura uma ruptura total, mas intensifica tendências já identificadas na evolução dos direitos de liberdade, privacidade e proteção de dados. A retomada dos fundamentos filosóficos da liberdade negativa, positiva e como conceito de não dominação, em conjunto com a análise histórica de que privacidade, que decorre do conceito de liberdade, evoluiu da ideia clássica de “direito de estar só” para o ideal de autodeterminação informativa, e, posteriormente, para a proteção de dados como direito fundamental, evidenciou que a proteção de dados se conecta ao direito de não sofrer interferências arbitrárias e de controlar o próprio destino.

Essa reflexão mostrou que a liberdade, na era digital, não se resume à ausência de coerção, mas exige instituições capazes de impedir o poder capilar de vigilância e manipulação que deriva de sistemas algorítmicos. Quando a informação se torna elemento estruturante da vida econômica e social, o controle sobre dados pessoais deixa de ser mera escolha individual e passa a integrar a própria definição de liberdade.

Essa evolução expressa uma transformação profunda: o surgimento da sociedade informacional, onde o poder deriva de fluxos contínuos de dados, estruturados por interconexão, inteligência coletiva e pela lógica global em rede, onde

a informação pessoal se tornou recurso estratégico, reorganizando sociabilidades, economias e formas de autoridade. Nesse contexto, a proteção de dados se converteu no principal mecanismo jurídico de contenção da vigilância e da exploração econômica dos indivíduos.

A partir desta conclusão teórica, a pesquisa examinou a evolução tecnológica e o surgimento da inteligência artificial como fenômeno técnico, econômico e social, inserido na sociedade informacional. Nesta análise, foi possível observar uma combinação sinérgica entre a evolução tecnológica identificada a partir da década de 1970, a massificação dos dados promovida pelo fenômeno do Big Data, e o avanço nas inovações em inteligência artificial, especialmente com o advento dos modelos conexionistas, do aprendizado profundo, da IA generativa e do modelo de linguagem de larga escala, a partir da década de 1990, com a consolidação amplificada da sociedade informacional.

Foi identificado que a IA aprofunda desafios da sociedade informacional como a opacidade algorítmica e a assimetria informacional. A opacidade algorítmica, foi analisada a partir das três dimensões, intencional, técnica e estrutural, e verificou-se que os algoritmos tornam-se caixas-pretas por razões econômicas (proteção de segredos e concentração de poder), pela complexidade intrínseca dos modelos e pela distância cognitiva entre especialistas e o público e produzem um ambiente de decisões automatizadas que escapam ao escrutínio democrático e ao controle institucional. Simultaneamente, a assimetria informacional reforçada pela concentração de infraestrutura computacional, demonstrou que poucos conglomerados tecnológicos detêm poder suficiente para influenciar mercados, regular comportamentos e moldar escolhas individuais.

Diante destes desafios, o estudo permitiu compreender que a transparência e a *accountability*, entendidas não como valores abstratos, mas como instrumentos estruturantes, são as bases normativas capazes de reequilibrar as relações informacionais. Transparência, neste caso, compreendida como condição pró-ética, e que implica fornecer informações verdadeiras, significativas e contextualizadas que permitam aos indivíduos compreenderem e contestarem decisões algorítmicas. Ao mesmo tempo, a *accountability*, traduz-se na existência de mecanismos de responsabilização e prestação de contas, com autoridade externa capaz de exigir justificativas, questionar procedimentos e aplicar sanções. Somente a conjugação desses princípios, aplicada de forma sistemática, pode reduzir a opacidade

algorítmica, redistribuir o conhecimento e assegurar que a autonomia dos titulares de dados não seja suprimida.

A partir da análise instrumentos técnicos e de governança aptos a operacionalizar transparência e *accountability* na IA, a dissertação propôs uma sistematização destes instrumentos necessários à efetivação desses princípios. A arquitetura construída demonstra que não há solução única para enfrentar a opacidade algorítmica e a assimetria informacional, e o seu combate depende da atuação conjunta de diferentes camadas interdependentes, e que essa solução deve ser, necessariamente, informacional, global e em rede, refletindo a própria lógica estruturante da sociedade informacional. Assim, essa sistematização buscou uma abordagem integrada, multissetorial e ancorada em cinco camadas: técnica, normativa, institucional, corporativa e social.

A camada técnica fornece os insumos materiais como explicabilidade , documentação e rastreabilidade, sem os quais nenhum mecanismo jurídico é operacionalizável. A camada normativa transforma essas capacidades técnicas em deveres jurídicos, estabelecendo padrões mínimos de transparência e *accountability*. A camada institucional, por sua vez, converte obrigações em instrumentos de fiscalização, sanção e interpretação, garantindo coerência e autoridade ao sistema. A camada corporativa coloca em prática esses comandos, estruturando mecanismos e processos internos de controle. Por fim, a camada social amplia a legitimidade do sistema, permitindo participação pública, contestação democrática, vigilância coletiva e atuação de organizações civis que pressionam tanto governos quanto empresas.

O principal achado da pesquisa reside na demonstração de que a eficácia da proteção de dados depende da articulação entre essas cinco camadas. Isoladamente, cada nível é insuficiente: a técnica não se converte automaticamente em direito; a norma, sem mecanismos técnicos e institucionais, torna-se letra morta; as instituições carecem de competência técnica sem insumos adequados; e as empresas tendem a adotar práticas cosméticas se não forem fiscalizadas e pressionadas por autoridades e pela sociedade. Assim, a conclusão reforça que a governança da IA exige uma arquitetura multinível e interconectada.

Esse modelo tem impactos relevantes no campo do Direito Negocial. Em ambientes negociais, a opacidade algorítmica e a assimetria informacional criam riscos concretos para a boa-fé objetiva, o equilíbrio contratual e a autodeterminação dos contratantes. A IA passou a ocupar funções centrais em processos negociais,

scoring de crédito, análise de risco, precificação dinâmica, marketing comportamental, mecanismos de compliance e tomada automatizada de decisões contratuais, e, nesse contexto, a falta de transparência amplia desigualdades informacionais e compromete a própria ideia de autonomia privada. A pesquisa concluiu que transparência e *accountability* são fundamentais para garantir que os contratantes tenham acesso à lógica decisória que influencia suas relações jurídicas e para assegurar que decisões automatizadas sejam contestáveis e auditáveis. Nesse ponto, o estudo oferece contribuições ao Direito Negocial ao indicar como deveres clássicos, dever de informação, dever de transparência e boa-fé objetiva, precisam ser reinterpretados em um ambiente de IA, a fim de limitar abusos de poder informacional e assegurar relações equilibradas.

A partir dessa sistematização, a dissertação identificou avanços necessários para consolidar uma governança informacional robusta. O primeiro é a coordenação internacional: a fragmentação regulatória entre União Europeia, Estados Unidos, China e Sul Global compromete qualquer tentativa de controle significativo e favorece arbitragens regulatórias. O segundo é a integração entre conhecimento técnico, jurídico e social: operadores do direito e reguladores precisam superar o juridicismo abstrato e desenvolver conhecimento técnico mínimo para compreender riscos, instrumentos e limitações. O terceiro é a pluralidade metodológica: a governança nunca poderá se apoiar em solução única, pois a opacidade resulta de múltiplas fontes e exige estratégias variadas e complementares, conectadas e em rede.

As contribuições secundárias destacaram ainda o potencial inexplorado da *blockchain* para reforçar rastreabilidade e documentação imutável, os limites intrínsecos da explicabilidade como ideal absoluto e, no caso do Brasil, a inadequação de importar integralmente modelos regulatórios estrangeiros, especialmente o modelo europeu, para o contexto brasileiro, exigindo a construção de um modelo próprio, ajustado às capacidades e necessidades nacionais.

As contribuições desta dissertação se estendem a diferentes dimensões. Em termos teóricos, oferece uma nova perspectiva para a articulação entre proteção de dados, IA, opacidade algorítmica, assimetria informacional, demonstrando como esses elementos são conectados na sociedade informacional e como podem ser enfrentados com transparência e *accountability*. Em termos normativos, apresenta uma síntese crítica de instrumentos capazes de orientar reformas legislativas e práticas regulatórias. Em termos institucionais, propõe uma arquitetura integrada

capaz de servir de referência para futuros estudos e iniciativas para regulação da IA. Em termos negociais, amplia a compreensão sobre os impactos da IA em relações contratuais e fornece parâmetros para reinterpretar deveres jurídicos clássicos. Em termos sociais, destaca o papel fundamental da participação democrática e da pressão pública na consolidação de mecanismos de controle.

Apesar dessas contribuições, a pesquisa reconhece limitações inerentes ao recorte metodológico adotado. A investigação concentrou-se em análise teórica e documental, sem testar empiricamente os instrumentos técnicos ou os modelos de governança propostos, o que restringe a verificação prática de sua eficácia.

Essas limitações abrem caminhos promissores para pesquisas futuras. São recomendáveis estudos empíricos que avaliem o funcionamento real de instrumentos técnicos. Também se mostram promissoras investigações sobre os efeitos da IA em desigualdades estruturais, vieses algorítmicos e discriminação indireta. No plano internacional, pesquisas sobre mecanismos multilaterais de governança e sobre o papel de organismos internacionais e autoridades regionais são fundamentais para enfrentar a natureza transnacional da IA. Por fim, dada a incipiência da literatura, recomenda-se aprofundamento sobre o uso de blockchain como instrumento de rastreabilidade e auditoria distribuída.

Em síntese, a dissertação confirmou que a tutela jurídica da proteção de dados pessoais diante da inteligência artificial só pode ser estruturada de modo eficaz se ancorada nos princípios de transparência e *accountability* e conclui que a tutela jurídica da proteção de dados na era da inteligência artificial não é um desafio meramente técnico ou jurídico, mas estrutural. A IA opera em um ecossistema global, informacional e em rede; portanto, sua governança deve refletir essa lógica. Apenas uma arquitetura multinível, técnica, normativa, institucional, corporativa e social, é capaz de enfrentar os riscos da opacidade algorítmica e da assimetria informacional, distribuindo poder, assegurando autonomia e promovendo justiça informacional. A proteção de dados pessoais, nesse contexto, torna-se não apenas um direito individual, mas um imperativo democrático: condição necessária para garantir que a inteligência artificial opere em favor da dignidade humana, da autodeterminação informativa e da liberdade nas relações contemporâneas.

REFERÊNCIAS

AI Now. **2023 LANDSCAPE: Confronting Tech Power.** Disponível em: https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf. Acesso em: 15 jun. 2024.

AKTHER, Afroja; AROBEE, Ayesha; ADNAN, Abdullah Al; AUYON, Omum; ISLAM, A. S. M. Johirul; AKTER, Farhad. **Blockchain as a platform for artificial intelligence (AI) transparency**. Emporia State University, Kansas, 2024.

ANANNY, Mike; CRAWFORD, Kate. **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic *accountability*.** New Media & Society, v. 20, n. 3, p. 973-989, 2018.

ANBIMA. **Panorama da regulação da inteligência artificial no Brasil e no mundo.** Rio de Janeiro: Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais, 2025.

ANTHONY, Lawrence F. W.; KANDING, Benjamin; SELVAN, Raghavendra. Carbontracker: **Tracking and Predicting the Carbon Footprint of Training Deep Learning Models.** arXiv preprint arXiv:2007.03051, 2020. Disponível em: https://arxiv.org/abs/2007.03051. Acesso em: 15 jun. 2024.

ARYA, V.; BELLAMY, R. K. E.; CHEN, P.-Y.; DHURANDHAR, A.; HIND, M.; HOOKER, S.; HOUDE, S.; LIAO, Q. V.; LUSENA, C.; MOHA, D.; MOURAD, S.; RAMAMURTHY, K. N.; ROSENFELD, C.; SAHA, D.; SATTIGERI, P.; SHAN, C.; SHI, Y.; SUN, Y.; TAN, S.; VASUDEVAN, V.; ZHANG, Y.; ZHANG, Y. **One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques**. arXiv preprint, arXiv:1909.03012, 2019.

BARRETO JUNIOR, Irineu Francisco; VENTURI JUNIOR, Gustavo. **Inteligência Artificial e seus efeitos na Sociedade da Informação.** IN: LISBOA, Roberto Senise (coord.). O Direito na Sociedade da Informação IV: movimentos sociais, tecnologia e atuação do Estado. São Paulo: Almedina, 2020.

BAUMAN, Zygmunt. **Modernidade e ambivalência.** Trad. de Marcus Antunes Penchel. Rio de Janeiro: Zahar, 1999.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. **On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?** *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* 2021. Disponível em: https://dl.acm.org/doi/10.1145/3442188.3445922. Acesso em: 15 jun. 2024.

BERLIN, Isaiah. Dois conceitos de liberdade. In.: BERLIN, Isaiah. **Estudos sobre a humanidade**: uma antologia de ensaios. Tradução Rosaura Eichenberg. São Paulo: Companhia das Letras, 2002.

BORENSTEIN, J.; HOWARD, A. **The Ethics of Autonomous Cars.** The Atlantic,

2020. Disponível em: https://www.theatlantic.com/technology/archive/2020/02/the-ethics-of-autonomous-cars/606769/. Acesso em: 15 jun. 2024.

BOVENS, Mark. **The quest for responsibility: *accountability* and citizenship in complex organisations.** Cambridge: Cambridge University Press, 1998.

BOVENS, Mark. **Analysing and assessing *accountability*: a conceptual framework.** European Law Journal, Oxford, v. 13, n. 4, p. 447-468, jul. 2007.

BRASIL. **Lei n° 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).** Brasília, DF: Presidência da República, 2018.

BRASIL, Senado Federal. **Projeto de Lei nº 2.**338 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: 10 out. 2023.

BRAUNEIS, Robert; GOODMAN, Ellen P. **Algorithmic Transparency for the Smart City.** Yale Journal of Law & Technology, vol. 20, nº 1, p. 103-175, 2018.

BROWN, Tom B. et al. **Language models are few-shot learners**. [S.l.]: OpenAI, 2020. Disponível em: https://arxiv.org/abs/2005.14165. Acesso em: 30 mar. 2025.

BROŻEK, Bartosz; KOŁODZIEJCZYK, Adam; WOLEŃSKI, Jan. **Ethical black boxes: Why we need to embed AI systems with audit logs for *accountability*.** AI & Society, v. 39, p. 1-13, 2024.

BRUNTON, Finn; NISSENBAUM, Helen. **Obfuscation: A User's Guide for Privacy and Protest.** Cambridge: MIT Press, 2015.

BURRELL, Jenna. **How the machine 'thinks': Understanding opacity in machine learning algorithms.** Big Data & Society, v. 3, n. 1, p. 1-12, 2016.

BURRELL, Jenna. **How the machine thinks: understanding opacity in machine learning algorithms.** Big Data & Society, v. 3, n. 1, 2016. Disponível em: https://journals.sagepub.com/doi/10.1177/2053951715622512. Acesso em: 14 jun. 2025.

BYGRAVE, Lee A. **Data privacy law: an international perspective.** Oxford: Oxford University Press, 2014.

CASSESE, Sabino. **Administrative Law Without the State?** The Challenge of Global Regulation. International Law and Politics, v. 37, n. 4, p. 663-691, 2005.

CASSESE, Sabino. **The globalization of law.** International Law and Politics, New York, v. 37, n. 4, p. 973-993, 2005.

CASTELLS, Manuel. **A sociedade em rede, 25.** ed. Rio de Janeiro: Paz e Terra, 2023.

CHEONG, Ben Chester. **Transparency and *accountability* in AI systems:**

safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, v. 6, 2024.

CHINA. **Action Plan for Global Governance of Artificial Intelligence.** Beijing: State Council Information Office, 2025. Tradução oficial em inglês publicada pelo China Aerospace Studies Institute (CASI). Disponível em: https://www.airuniversity.af.edu/Portals/10/CASI/documents/Translations/2025-07%20China%20Action%20Plan%20for%20AI.pdf . Acesso em: 29 set. 2025.

CITRON, Danielle Keats; PASQUALE, Frank. **The scored society: Due process for automated predictions.** Washington Law Review, v. 89, n. 1, p. 1-33, 2014. Disponível em: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/1. Acesso em: 14 jun. 2025.

COMISSÃO EUROPEIA. **'Declaração dos dirigentes do G7 sobre o processo de IA Hiroxima'.** Disponível em: https://digitalstrategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process. Acesso em: 10 jun. 2024.

DATA PRIVACY BRASIL. **Governança global de IA: agindo com a sociedade civil.** São Paulo: Data Privacy Brasil, 2025. Disponível em: https://www.dataprivacybr.org/governanca-global-de-ia-agindo-com-a-sociedade-civil/ . Acesso em: 29 set. 2025.

DIAKOPOULOS, N. **Transparency.** In: DUBBER, M., PASQUALE, F. and Das, S. The Oxford Handbook of Ethics of AI.OUP, 2020. p. 197 a 214.

DIAKOPOULOS, Nicholas. **Algorithmic *Accountability* Reporting: On the Investigation of Black Boxes.** Tow Center for Digital Journalism, 2013. Disponível em: https://academiccommons.columbia.edu/doi/10.7916/D8TT59TC. Acesso em: 14 jun. 2025.

DIGNUM, V. **Responsibility and Artificial Intelligence.** In: DUBBER, M., PASQUALE, F. and Das, S. The Oxford Handbook of Ethics of AI.OUP, 2020. p. 215 a 231.

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

DONEDA, Danilo. **ALMEIDA, V.** A. F. O que é a governança de algoritmos?. In: Tecnopolíticas da vigilância : perspectivas da margem / organização Fernanda Bruno … [et al.] ; [tradução Heloísa Cardoso Mourão … [et al.]]. - 1. ed. - São Paulo : Boitempo, 2018

DONEDA, Danilo César. **Da privacidade à proteção de dados pessoais.** 2. ed. São Paulo: Thomson Reuters Brasil, 2022.

DONEDA, Danilo Cesar Maganhoto. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados.** 2. ed. São Paulo: Thomson Reuters Brasil, 2020.

EDWARDS, Lilian; VEALE, Michael. ***Accountability* of AI under the law: the role of explanation.** Proceedings of the IEEE, [s. l.], v. 108, n. 3, p. 435-450, 2020. DOI: https://doi.org/10.1109/JPROC.2019.2940428.

EDWARDS, Lilian; VEALE, Michael. **Slave to the algorithm?** Why a right to an explanation is probably not the remedy you are looking for. Duke Law & Technology Review, v. 16, p. 18-84, 2017

ELIAS, Maria Ligia Ganacim Granado Rodrigues. **Liberdade como não interferência, liberdade como não dominação, liberdade construtivista: uma leitura do debate contemporâneo sobre a liberdade.** 2014. Tese (Doutorado em Ciência Política) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2014. doi:10.11606/T.8.2014.tde-16012015-152209. Acesso em: 2023-11-27.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA (ENAP). **Regulação da inteligência artificial: benchmarking de países selecionados.** Brasília: ENAP, 2022.

EUROPEAN COMMISSION. **Ethics guidelines for trustworthy AI.** Brussels: European Commission, 2019.

EUROPEAN COMMISSION. **White Paper on Artificial Intelligence: A European approach to excellence and trust.** European Commission, 2020. Disponível em: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Acesso em: 15 jun. 2024.

FACELI, Katti; LORENZETTI, Daniela; PARPINELLI, Rafael S.; ZANIN, Clarisse S. **Inteligência artificial: uma abordagem de aprendizado de máquina.** Rio de Janeiro: LTC, 2011.

FARIA, José Eduardo. **A Globalização Econômica e Sua Arquitetura Jurídica.** Revista Academia Judicial, 2010.

FERRAJOLI, Luigi. **A Soberania no Mundo Moderno.** São Paulo: Martins Fontes, 2002.

FLORIDI, Luciano. **The ethics of information.** Oxford: Oxford University Press, 2013.

FLORIDI, Luciano. **The fourth revolution: how the infosphere is reshaping human reality.** 1. ed. Oxford: Oxford University Press, 2014.

FOX, Jonathan. **Social *accountability*: what does the evidence really say?** World Development, [S. l.], v. 72, p. 346-361, 2015. DOI: https://doi.org/10.1016/j.worlddev.2015.03.011.

G20. **G20 AI Principles.** Tóquio: G20 Ministerial Meeting on Trade and Digital Economy, 2019. Disponível em: https://g7g20-documents.org/fileadmin/G7G20_documents/2019/G20/Japan/Leaders/2%20Leaders%27%20Annex/G20%20AI%20Principles_2019.pdf . Acesso em: 10 jun. 2024.

GANDY, Oscar. **The Panoptic Sort: A Political Economy of Personal Information.** Boulder: Westview, 1993.

GAO, Qiqi; ZHANG, Jiteng. **Artificial Intelligence Governance and the Blockchain.** In: ZHANG, J.; FLORIDI, L. (org.). Artificial Intelligence and the Rule of Law. Singapore: Springer, 2024.

GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE. **About GPAI.** Disponível em: https://gpai.ai/about/. Acesso em: 10 jun. 2024.

GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE. **Responsible AI Strategy for the Environment - Workshop Report.** 2023. Disponível em: https://gpai.ai/projects/responsibleai/Responsible%20AI%20Strategy%20for%20the%20Environment%20-%20Workshop%20Report.pdf. Acesso em: 15 jul. 2024.

GONÇALVES, Alcindo. **A legitimidade na governança global**. Trabalho apresentado no XV Congresso do CONPEDI – Conselho Nacional de Pesquisa e Pós-Graduação em Direito, Manaus, 2006.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning.** MIT Press, 2016.

GOOGLE. **'Frontier Model Forum: a new partnership to promote responsible AI'.** Disponível em: https://blog.google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum/. Acesso em: 10 jun. 2024.

GREENE, D.; HOFFMANN, A. L.; STARK, L. **Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning**. Proceedings of the 52nd Hawaii International Conference on System Sciences, p. 2122-2131, 2019. Disponível em: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/df8e4ef6-5c4d-40c8-8a6a-2cb5bc1ad571/content. Acesso em: 15 jun. 2024.

GUIDOTTI, R. et al. **A survey of methods for explaining black box models**. ACM Computing Surveys, v. 51, n. 5, p. 93–96, 2018. DOI: https://doi.org/10.1145/3236009

GUNNING, D., VORM, E., WANG, J.Y. **and TUREK, M.** DARPA's explainable AI (XAI) program: A retrospective. Applied AI Letters, 2: e61. https://doi.org/10.1002/ail2.61, 2021.

GÉRON, Aurélien. **Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras e TensorFlow: conceitos, ferramentas e técnicas para construir sistemas inteligentes.** Tradução de Pedro Jansen. Rio de Janeiro: Alta Books, 2019.

HANSEN, Lars Kai; RIEGER, Laura. **Interpretability in intelligent systems – a new concept?** In: SAMEK, Wojciech et al. (org.). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Cham: Springer, 2019. p. 41-49. DOI: https://doi.org/10.1007/978-3-030-28954-6_3

HARTMANN PEIXOTO, Fabiano; SILVA, Roberta Zumblick Martins da. **Inteligência Artificial e Direito.** Coleção Direito, Racionalidade e Inteligência Artificial. Curitiba: Alteridade, 2019.

HESTNESS, Joel et al. **Deep learning scaling is predictable, empirically.** arXiv preprint arXiv:1712.00409, 2017. Disponível em: https://arxiv.org/abs/1712.00409. Acesso em: 14 de junho de 2025.

HILDEBRANDT, Mireille; GUTWIRTH, Serge. **Privacy, due process and the computational turn: the philosophy of law meets the philosophy of technology.** Abingdon: Routledge, 2019.

HIROSHIMA PROCESS. **International Code of Conduct for Organizations Developing Advanced AI Systems.** 2023. Disponível em: https://digital-strategy.ec.europa.eu/pt/library/g7-leaders-statement-hiroshima-ai-process. Acesso em 15 jul. 2024.

HIROSHIMA PROCESS. **International Guiding Principles for Organizations Developing Advanced AI Systems.** 2023. Disponível em: https://digital-strategy.ec.europa.eu/pt/library/g7-leaders-statement-hiroshima-ai-process. Acesso em 15 jul. 2024. IPCC. Global Warming of 1.5°C. 2018. Disponível em: https://www.ipcc.ch/sr15/download/#full. Acesso em: 10 jun. 2024.

ICHOU, Rachel Pollack. **Opening the black box: in search of algorithmic transparency.** GigaNet 11th Annual Symposium, 2016. Disponível em: https://ssrn.com/abstract=3837723. Acesso em: 14 jun. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE.Recenseamento do Brazil em 1872. **Rio de Janeiro: Typographia Nacional, 1872.** Disponível em: https://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS%20-%20RJ/Recenseamento_Brazil_1872.pdf. Acesso em: 8 mar. 2025.

INSTITUTO DE TECNOLOGIA E SOCIEDADE DO RIO (ITS Rio). **Requerimento de transparência na consulta pública do PL 2338/2023 sobre inteligência artificial.** Rio de Janeiro: ITS Rio, 2024. Disponível em: https://itsrio.org/pt/comunicados/requerimento-transparencia-consulta-publica-pl2338-inteligencia-artificial/#:~:text=de%202024%20e%20encerrou%20o,em%20torno%20desse%20importante%20tema . Acesso em: 29 set. 2025.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. **The global landscape of AI ethics guidelines.** Nature Machine Intelligence, v. 1, n. 9, p. 389-399, 2019. Disponível em: https://www.nature.com/articles/s42256-019-0088-2. Acesso em: 15 jun. 2024.

Kaminski, Margot E., Understanding Transparency in Algorithmic *Accountability* (June 8, 2020). **Forthcoming in Cambridge Handbook of the Law of Algorithms, ed.** Woodrow Barfield, Cambridge University Press (2020)., U of Colorado Law Legal Studies Research Paper No. 20-34, Available at SSRN: https://ssrn.com/abstract=3622657

KAPLAN, Jerry. **Generative Artificial Intelligence: What Everyone Needs to Know.** New York: Oxford University Press, 2023.

KATYAL, Sonia K. **Private *accountability* in the age of artificial intelligence.** UCLA Law Review, Los Angeles, v. 66, n. 1, p. 54-103, 2019.

KROLL, Joshua A.; HUEY, Joanna; BAROCAS, Solon; FELTEN, Edward W.; REIDENBERG, Joel R.; ROBINSON, David G.; YU, Harlan. **Accountable algorithms.** University of Pennsylvania Law Review, vol. 165, nº 3, p. 633-705, 2017.

LU, Sylvia. **Data Privacy, Human Rights, and Algorithmic Opacity.** California Law Review, vol. 110, nº 6, p. 2087-2146, 2022.

LU, Sylvia. **Algorithmic Opacity, Private *Accountability*, and Corporate Social Disclosure in the Age of Artificial Intelligence, 23 Vand.** J. Ent. & Tech. L. 99 , 2020.

LUCCA, Newton de; MACIEL MADEIRA DEZEM, Renata Mota. **A proteção de dados pessoais no Brasil a partir da Lei n.** 13.709/2018: avanço ou retrocesso? In: STORÓPOLI, Eduardo; BASILIO, Nadir da Silva (org.). Direito Empresarial: estruturas e regulação. São Paulo: UNINOVE, 2018. p. 255-277.

LUCCA, Newton de; MACIEL, Renata Mota. **A Lei n. 13.709, de 14 de agosto de 2018: a disciplina normativa que faltava.** In: SIMÃO, José Fernando; PAVINATO, Tiago (coords.). Liber amicorum Teresa Ancona Lopez: estudos sobre responsabilidade civil. São Paulo: Almedina, 2021. p. 621-643.

LUNDBERG, S. M.; LEE, S. I. **A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS).** Red Hook: Curran Associates, 2017. p. 4765–4774.

LÉVY, Pierre. **Cibercultura.** Tradução: Carlos Irineu da Costa. São Paulo: Ed. 34, 1999.

LÉVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática.** Tradução: Carlos Irineu da Costa. São Paulo: Ed. 34, 1993.

LÉVY, Pierre. **O que é o virtual?**. Tradução: Carlos Irineu da Costa. São Paulo: Ed. 34, 1996.

MARQUES, Ana Beatriz et al. **Relatório de Pesquisa-Regulação da Inteligência Artificial ao Redor do Mundo (Research Report-Regulation of Artificial Intelligence Around the World).** 2024. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4803041. Acesso em: 29 mar. 2025.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, and Think.** Nova Iorque: Houghton Mifflin Harcourt, 2013.

McCarthy, John. **What is Artificial Intelligence?** Stanford University, 2007. Disponível em: http://jmc.stanford.edu/articles/whatisai/whatisai.pdf. Acesso em: 29 mar. 2025.

MCDERMID, John; WILLIAMSON, Philip; HABLI, Ibrahim. **Artificial intelligence explainability: The technical and ethical dimensions.** Philosophical Transactions of the Royal Society A, [s. l.], v. 379, n. 2207, p. 1-19, 2021. DOI: https://doi.org/10.1098/rsta.2020.0183.

MENDES, Laura Schertel Ferreira. **Autodeterminação informativa: a história de um conceito.**Pensar, Fortaleza, v. 25, n. 4, p. 1-18, out./dez. 2020. Disponível em: http://periodicos.unifor.br/rpen/article/view/10828. Acesso em: 8 mar. 2025.

MILL, John Stuart. **Sobre a liberdade.** Tradução de Álvaro Cabral. Rio de Janeiro: Nova Fronteira, 2011.

MITCHELL, Melanie. **Artificial intelligence: a guide for thinking humans.** New York: Farrar, Straus and Giroux, 2019.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.** Munich: [s.n.], 2019.

MORLEY, Jessica; FLORIDI, Luciano; KINSEY, Libby; ELHALAL, Anat. **From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices.** Science and Engineering Ethics, v. 26, p. 2141-2168, 2020. Disponível em: https://link.springer.com/article/10.1007/s11948-019-00165-5. Acesso em: 15 jun. 2024.

MULGAN, Richard. ***Accountability*: an ever-expanding concept?**. Public Administration, Oxford, v. 78, n. 3, p. 555-573, 2000.

NAKAMOTO, Satoshi. **Bitcoin: A Peer-to-Peer Electronic Cash System.** 2008. Disponível em: https://bitcoin.org/bitcoin.pdf. Acesso em: 30 ago. 2025.

NOVELLI, Claudio; TADDEO, Mariarosaria; FLORIDI, Luciano. ***Accountability* in artificial intelligence: what it is and how it works.** AI & Society, [s. l.], v. 39, p. 1871-1882, 2024.

OLUWAGBADE, Elizabeth. **Ethical Black Boxes in AI: Legal and Technical Challenges in Ensuring Transparency and *Accountability*.** [S. l.]: University of Cape Coast, 2025.

ONU. **Secretary-General's Roadmap for Digital Cooperation.** 2020. Disponível em: https://www.un.org/en/content/digital-cooperationroadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf. Acesso em: 15 jul. 2024.

ONU. **Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development.** 2024. Disponível em: https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=Ox48kS

JJK4LhveiqM1&fe=true. Acesso em: 15 jul. 2024.

ONU. **Transforming Our World: The 2030 Agenda for Sustainable Development.** 2015. Disponível em: https://sustainabledevelopment.un.org/post2015/transformingourworld. Acesso em: 15 jun. 2024.

ONU NEWS. **General Assembly adopts first-ever resolution on artificial intelligence.** 21 mar. 2024. Disponível em: https://news.un.org/en/story/2024/03/1147831 . Acesso em: 10 jun 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Recommendation of the Council on Artificial Intelligence.** Paris: OECD, 2019.

PARTNERSHIP ON AI. **About Us.** Disponível em: https://partnershiponai.org/about/. Acesso em: 10 jun. 2024.

PARTNERSHIP ON AI. **'About Us'.** Disponível em: https://partnershiponai.org/about/. Acesso em: 10 jun. 2024.

PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information.** Cambridge: Harvard University Press, 2015.

PASTORE, José. **Evolução tecnológica: repercussões nas relações de trabalho.** Disponível em: http://www.josepastore.com.br/artigos/rt/rt_246.htm. Acesso em: 20 jan. 2024.

PETTIT, Philip. **Republicanism: A Theory of Freedom and Government.** Oxford: Oxford University Press, 1997.

PINTO, Álvaro Vieira. **O conceito de tecnologia.** v.1. Rio de Janeiro: Contraponto, 2005.

QUEIROZ, Renata Capriolli Zocatelli Queiroz. **Encarregado de Proteção de Dados Pessoais – DPO: Regulamentação e Responsabilidade Civil.** São Paulo: Quartier Latin, 2022.

REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTINGTON, Meredith. **Algorithmic Impact Assessments: A Practical Framework for Public Agency *Accountability*.** Nova York: AI Now Institute, 2018. Disponível em

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. **"Why should I trust you?": Explaining the predictions of any classifier.** In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016. p. 1135–1144. DOI: https://doi.org/10.1145/2939672.2939778

RIBEIRO, Ana Carolina de Faria Silvestre. **Black box e o direito face à opacidade algorítmica.** Revista de Direito, Inovação e Tecnologia, vol. 3, nº 1, p. 18-38, 2021.

ROBERTS, Huw; HINE, Emmie; TADDEO, Mariarosaria; FLORIDI, Luciano. **Global**

AI governance: barriers and pathways forward. International Affairs, v. 100, n. 3, p. 1275-1286, 2024. Disponível em: https://academic.oup.com/ia/article/100/3/1275/7641064. Acesso em: 15 jun. 2024.

RODOTÁ, Stéfano. **A vida na sociedade da vigilância: a privacidade hoje.** Trad. Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

ROLNICK, David; DONTI, Priya L.; KAACK, Lynn H.; KOCHANSKI, Kelly; LACOSTE, Alexandre; SANKARAN, Karthik; BENGIO, Yoshua. **Tackling Climate Change with Machine Learning.** arXiv preprint arXiv:1906.05433, 2019. Disponível em: https://arxiv.org/abs/1906.05433. Acesso em: 15 jun. 2024.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach.** 4. ed. Hoboken, 2021.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach.** 3. ed. Prentice Hall, 2016.

SAMEK, Wojciech; MÜLLER, Klaus-Robert. **Towards explainable artificial intelligence.** In: SAMEK, Wojciech et al. (org.). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Cham: Springer, 2019. p. 5-22. DOI: https://doi.org/10.1007/978-3-030-28954-6_1

SCHUMPETER, Joseph. A**. Business Cycles. A Theoretical, Historical, and Statistical Analysis of the Capitalist Process**. Nova York: Mc Graw-Hill, 1939, 2v.

SCHWARTZ, Roy; DODGE, Jesse; SMITH, Noah A.; ETZIONI, Oren. **Green AI.** Communications of the ACM, v. 63, n. 12, p. 54-63, 2020. Disponível em: https://cacm.acm.org/magazines/2020/12/248829-green-ai/fulltext. Acesso em: 15 jun. 2024.

SILVA, Felipe Gomes; COSTA, Patricia Ayub da; MUNIZ, Tania Lobo. **Liberdade na era digital: desafios da cibercultura e da sociedade de vigilância.** In: AYRES PINTO, Danielle Jacon; LANNES, Yuri Nathan da Costa; BRUNET, Laura Inés Nahabetián (Coord.). Anais do XIII Encontro Internacional do CONPEDI: Governo Digital, Direito e Novas Tecnologias I, Montevidéu, 2024. Florianópolis: CONPEDI, 2024. Disponível em: https://www.conpedi.org.br. Acesso em: 8 mar. 2025.

SKINNER, Quentin. **Liberty before liberalism.** Cambridge: Cambridge University Press, 1998.

SOLOVE, Daniel J. **Understanding Privacy.** Cambridge, Massachusetts: Harvard University Press, 2008. Disponível em: http://understanding-privacy.com. Acesso em: 8 mar. 2025.

STONE, Peter. **Artificial Intelligence and life in 2030: report of the 2015-2016.** Stanford University, 2016. Disponível em: https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf. Acesso em: 10 out. 2023.

STRUBELL, Emma; GANESH, Ananya; MCCALLUM, Andrew. **Energy and Policy Considerations for Deep Learning in NLP.** Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. Disponível em: https://www.aclweb.org/anthology/P19-1355/. Acesso em: 15 jun. 2024.

TAPSCOTT, Don; TAPSCOTT, Alex. **Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World.** New York: Penguin Publishing Group, 2016.

TEIXEIRA, Tarcísio; CHELIGA, Vinicius. **Inteligência Artificial: aspectos jurídicos.** 2. ed. Salvador: Ed. JusPODIVM, 2020.

TEUBNER, Gunther. **The Global Bukowina: On the Emergence of a Transnational Legal Pluralism.** In: TEBBE, Noah; TEUBNER, Gunther; (Orgs.). Global Law Without a State. Brookfield: Dartmouth Publishing Company, 1997, p. 3-28.

THE AI STANDARDS HUB. **Standards database.** Disponível em: https://aistandardshub.org/ai-standards-search/. Acesso em: 10 jun. 2024.

THE AI STANDARDS HUB. **'Standards database'.** Disponível em: https://aistandardshub.org/ai-standards-search/. Acesso em: 10 jun. 2024.

TURING, Alan. **Computing machinery and intelligence.** Mind: a quarterly review of psychology and philosophy. Oxford: Oxford University Press, v. 59, n. 236, p. 433-460, out. 1950. Disponível em: https://phil415.pbworks.com/f/TuringComputing.pdf. Acesso em: 15 jun. 2024.

UNDP. **Using AI to help achieve Sustainable Development Goals.** United Nations Development Programme, 2021. Disponível em: https://www.undp.org/publications/using-ai-help-achieve-sustainable-development-goals. Acesso em: 15 jun. 2024.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO). **Recommendation on the Ethics of Artificial Intelligence.** Paris: UNESCO, 2021.

UNITED STATES OF AMERICA. **America's AI Action Plan: winning the race.** Washington, D.C.: The White House, 2025.

UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016.**

UNIÃO EUROPEIA. **Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (AI Act).** Estrasburgo: Parlamento Europeu, 2024.

VINUESA, Ricardo; AZIZPOUR, Hossein; LEITE, Iolanda; BALAAM, Madeline; DIGNUM, Virginia; DOMISCH, Sami; … **NERINI, Francesco Fuso.** The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. Nature Communications, v. 11, n. 1, p. 233, 2020. Disponível em:

https://www.nature.com/articles/s41467-019-14108-y. Acesso em: 15 jun. 2024.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. **Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation.** International Data Privacy Law, v. 7, n. 2, p. 76-99, 2017.

WARREN, Samuel D.; BRANDEIS, Louis D. **The Right to Privacy, Harvard Law Review, Vol.** 4, No. 5., pp. 193-220, 1890.

WELLER, Adrian. **Transparency: motivations and challenges.** In: SAMEK, Wojciech et al. (org.). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Cham: Springer, 2019. p. 23-40. DOI: https://doi.org/10.1007/978-3-030-28954-6_2

WESTIN, Alan. **Privacy and freedom.** New York: Ig Publishing, 2015.